GaitViewer: Semantic Gait Data Analysis and Visualization Tool

Asan Agibetov[1], Karelia Elena Tecante Gutierrez[2], Chiara Eva Catalano[1], Giuseppe Patanè[1], Christof Hurschler[2], and Michela Spagnuolo[1]

[1] Italian National Research Council (CNR-IMATI), Genoa, Italy
{asan,chiara,giuseppe,michela}@ge.imati.cnr.it
[2] Hannover Medical School (LBB-MHH), Hannover, Germany
{tecantegutierrez.karelia,hurschler.christof}@mh-hannover.de

Abstract. Clinical gait analysis studies human locomotion by characterizing movement patterns with heterogeneous acquired *gait data* (e.g., spatio-temporal parameters, geometry of motion, measures of force). Lack of semantic integration of these heterogeneous data slows down collaborative studies among the laboratories that have different acquisition systems. In this work we propose a semantic integration methodology for gait data, and present GaitViewer - a prototype web application for semantic analysis and visualization of gait data. The proposed semantic integration methodology separates heterogeneous and mixed numerical and meta information in gait data. Ontology concepts represent the separated meta information, while numerical information is stored in a NoSQL database. Parallel coordinates visual analytics technique are used as an interface to the analytics tools proposed by the NoSQL database. We tailor GaitViewer for two common use-cases in clinical gait analysis: correlation of measured signals for different subjects, and follow-up analysis of the same subject. Finally, we discuss the potential of a large-scale adoption of frameworks such as GaitViewer for the next generation diagnosis systems for movement disorders.

Keywords: gait analysis, semantic interoperability, ontology, multivariate data visualization, visual analytics, information filtering, information retrieval

1 Introduction

The intricacy of human locomotion has captivated a great number of scientists over the years. At first glance, human motion or human gait may seem natural, graceful, rhythmic, and even effortless. However, it is such an extremely complex process that, despite more than a century making developments in this field, much remains to be uncovered. Indeed, the human gait is a distinct, unique and remarkably precise process, controlled by the central nervous system. It comprises three dimensional motions in a multiple linkage system involving the coordination of a great number of different muscles acting on different joints. As a result, the research community is still working on understanding and unfolding the underlying mechanisms that allow people to move.

Gait analysis [1] is one such effort from the research community, which involves instrumented measurement of the movement patterns by characterizing human motion

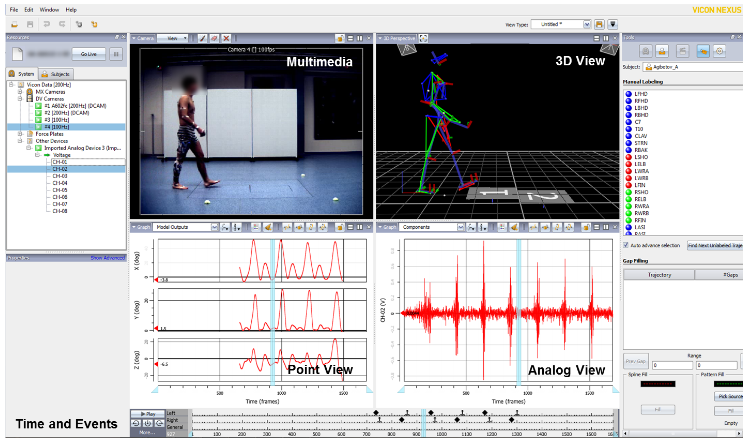


Fig. 1. Visualization of heterogeneous gait data in VICON (commercial software), from top to bottom in clockwise order: video sequences, motion capture skeleton animation, muscle activation timeseries, joint angles timeseries

(*e.g.,* walking, running, jumping, hopping). Human motion is usually characterized by spatio-temporal parameters (distance, speed, cadence, stride length), geometry of motion (kinematics), measures of force (ground reaction forces and joint moments), muscles activation and sometimes energy expenditure. Subsequent computational analyses of the aforementioned parameters contribute to diagnosis and treatment of various movement disorders [17]. As a result, gait analysis has expanded from a purely academic discipline into a useful tool to physicians and therapists which is distinctively referred as *clinical gait analysis*.

Undoubtedly, technological advancements have made acquisition, processing and analysis of gait data easier and faster. As described by Simon [14], nowadays there is a wide variety of new instrumentation and computer technology for observation and measurement of human motion, which include: motion capture, digital video camcorders, accelerometers, forceplates and electromyography (EMG). Well-established commercial gait data acquisition systems include Vicon (Motion Capture systems, Oxford-United Kingdom, Figure 1), Motion Analysis (Motion Analysis Corporation, Calfornia-United States), Qualysis (Qualysis AB, Gothenburg-Sweden), Codamotion (3D Motion Analysis systems, Leicestershire-United Kingdom), and zflo (zflomotion, Boston-United States). The usual scenario of gait data acquisition involves a patient undergoing an acquisition session, with a specific acquisition protocol, which produces large amount of heterogeneous data characterizing human motion. Each acquisition system follows its own ad-hoc data model for gait data analysis and visualization (Figure 1).

Need for semantic integration techniques. Commercial software accompanying acquisition systems provides basic means for gait data exportation as text files or spreadsheets, containing mixed meta and numerical information. The variety of exported data ranges from meta information (e.g., acquisition session number, acquisition protocol trial number, anatomical region of the marker) to the actual numerical information (e.g., timeseries). Such a disorganized and heterogeneous representation imposes problems on efficient consumption and interoperability of gait data. Before the numerical information can be fed into numerical packages for further analysis, several data preparation routines are usually performed manually. The data preparation routines involve separating signals corresponding to a specific acquisition session of a given subject, acquired with a desired anatomical marker. The results of the subsequent analysis are, in most cases, stored manually following an ad-hoc data model, if any, in text files and/or spreadsheets. The data sharing between the collaborating laboratories usually takes place via exchange of emails, with explanations on the structure of the gait data, which are prone to be ambiguous. The interoperability among the acquired data is crippled even more when different acquisition systems are used, which is a common case when different laboratories have different acquisition systems and they want to share data.

Gait data could benefit from the semantic integration techniques for a more efficient data consumption. For instance, an ontology describing gait data acquisition procedures could help organize acquired data sets, which relate to the same subject/acquisition session/anatomical marker, however have different encoding and/or representation. Principled and structured ways of storing raw numerical gait data could facilitate gait data analysis, visualization, and data exchange among the collaborating laboratories. A semantic integration methodology which combines conceptual classification of gait data with a more low-level (semi-) structured organization of numerical data, could likely open up new collaborative scenarios for the human motion researchers and clinicians alike, by facilitating data exchange, interpretation, and visualization.

In this work we propose a semantic integration methodology for gait data, and present GaitViewer - a prototype web application for semantic analysis and visualization of gait data. The proposed semantic integration methodology separates heterogeneous and mixed numerical and meta information in gait data. We use ontology concepts to represent the separated meta information, while numerical information is stored in a NoSQL database. GaitViewer leverages ontologies as a common semantic layer to access low-level numerical gait data stored in the database, and provides interactive visual interface, based on parallel coordinates plots, to the analytics tools proposed by the NoSQL database. We apply GaitViewer to two use-cases in the clinical gait analysis: correlation of measured signals for different subjects, and follow-up analysis of the same subject. This work has been a collaboration between biomedical engineers practicing clinical gait analysis, and computer scientists working on biomedical applications of semantic data analysis and visualization.

2 Related Work

Commercial gait data acquisition systems provide basic analysis tools that usually enables the user to create, edit and export data. These tools are mainly intended as a quick means of data quality inspection and report editing tool for gait laboratories; however they lack more complex visualization tools as well as analysis capabilities. More recently, free software, applications or toolboxes for Matlab have been developed to provide a wider range of analysis tools, hence increasing the potential to share data between laboratories that have different acquisition systems. For example, toolboxes for Matlab such as BTK and Mocap were created to facilitate gait data visualization and processing; nevertheless the user must have certain level of programming skills in order to work with the functions provided by the toolboxes [6, 2]. Barre and Armand [2] went a step further developing Mokka (Motion kinematic and kinetic analyzer), an open-source and cross-platform software that utilizes the BTK toolbox to analyze biomechanical data, but designed for scientists with non-programming skills. iGait, another Matlab application tool developed to derive numerous features uniquely from acceleration data, including spatio-temporal features, regularity, symmetry and spectral features [18].

In terms of gait data visualization, Federolf *et al.* [7] apply timeseries visualization and skeleton animation in order to study the temporal variability in gait. Spahić et al. [15] provide gait visualization system based on Paraview (free available software). And Manal and Stanhope [10] suggested an alternate method of reporting movement pattern deviations relative to normative data by color-coding the magnitude and the direction of the difference (color coded deviations).

Open-source initiatives and Matlab toolboxes provide a solid toolchain for complex gait analysis and visualization, which mainly focus on gait data of one subject at time. Although they do address, to some extent, the problem of gait data sharing among the laboratories that have different acquisition systems, such solutions do not go as far as to support collaboration scenarios in which disparate repositories of gait data are mined and searched through a common interface.

In other bioinformatics domains conceptual semantic integration techniques have already contributed to collaborative analysis and data interoperability. In particular, Semantic Web based methods have been introduced, which are designed to add meaning to the raw data by using formal description of the concepts, terms, properties and relationships encoded within the data [3]. Ontology Based Data Access Approach (OBDA [9]) demonstrates an integration of high-level conceptual layer (ontologies) and low-level layer for storing raw data in relational database management systems (RDBMS).

3 Semantic integration of gait data

In this section we present the developed methodology for the semantic integration of gait data. Our methodology consists in: i) semantic characterization of the acquisition session workflow for the gait analysis, encoded in an ontology, ii) semantic characterization of signals subject's movement via critical points computations per distinct phase of a gait cycle, iii) human movement signal persistence to a document-based (NoSQL) database [13] to support subjects measurement comparison via querying and aggregation framework, iv) web-based visualization of multivariate subject's movement measurement using parallel coordinates [8].

Ontology of gait data collection. A common scenario of gait data acquisition involves a patient undergoing an acquisition session, where specific acquisition protocols are employed for a specific study. As the result heterogeneous data characterizing human motion are acquired. To increase the semantic interoperability of gait data we separate meta information from numerical information. We first model the extent of meta information one could extract from gait data, i.e., the medical background knowledge, consisting of patients/subjects undergoing acquisition sessions, (*e.g.,* pre total knee replacement, 3 months after total knee replacement). Such that, these acquisition sessions produce numerical measurements of human motion, registered by the sensors placed on patients anatomical markers (e.g., joints and muscles) in different body planes (e.g., flexion-extension, internal-external rotation). This medical background knowledge is captured by the *MultiScaleHuman Ontology* [11], developed within the EU FP7 "MultiScaleHuman" project, whose goal is to associate multi-scale biomedical data with anatomical entities, patient and acquisition session/protocol information to support CAD and Visualization systems for diagnosis of musculoskeletal diseases of a human knee. We have set up a Stardog triple store (Knowledge Base) with gait data at `http://stardog.plumdeq.xyz/mshOntology#!/browse/ontology\%3AKnee_angles` (username: anonymous, password: guest). Currently, we take into account two common types of numerical measurements: angle variation in joints, and muscle activation (Figure 2).

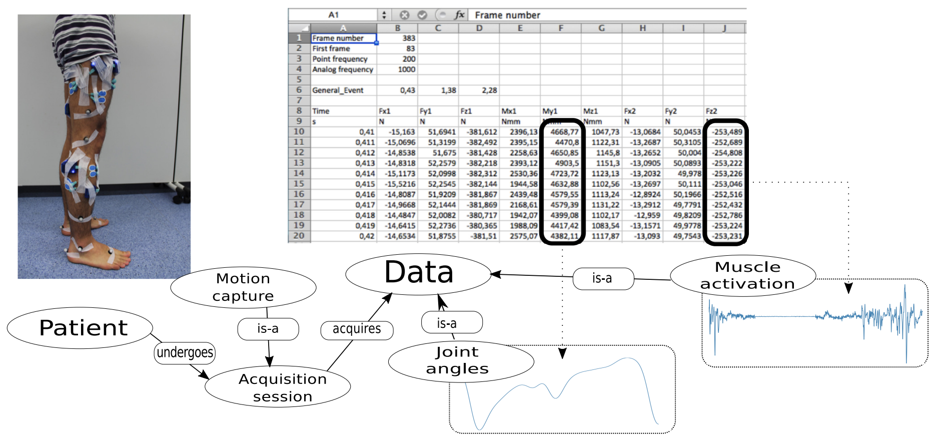


Fig. 2. Semantic interpretation of a typical gait data acquisition workflow, encoded in an ontology (excerpt of the ontology visualized). The acquisition workflow produces mixed meta and numerical data (visualized in a spreadsheet). Examples of muscle activation (EMG) and joint angles timeseries are provided.

Representation of gait data. Acquired numerical gait data may be represented as signals of different dimensions. In our work we consider 1D timeseries of two types: smooth and noisy. Joint angles (Figure 2, timeseries annotated with `Joint angles` ontology concept) and joint moments are smooth one-dimensional signals, while the muscle activation (EMG) (Figure 2, timeseries annotated with `Muscle activation` ontology concept) are noisy (contain measurement and calibration errors, resulting in many high frequencies) one-dimensional signals. We can represent 1D signals as sequences $x[n], n = 0, 1, \ldots, N-1$ or as vectors $\boldsymbol{x} = [x_0 x_1 \ldots x_{N-1}]^T$. All angles/moment signals, belonging to the same subject and acquired within the same acquisition session and trial, have the same number of sampled points and thus share the index set $\mathcal{I} \in \mathbb{Z}^+$. In the rest of the paper we refer to the acquired numerical gait data as signals and timeseries interchangeably.

Semantic characterization of signals. Because *walking* is such a fundamental skill associated with quality of life, it is the most common motion analyzed in health mobility centers. The walking pattern can be characterized by studying one full gait cycle, which is initiated by leaning the body above the legs, and continued with the contraction of leg muscles. The propulsive power of ankle muscles provoke controlled forward shift of the center of body mass, and the subsequent controlled fall of the body, which is only stopped by the initiation of the next gait cycle.

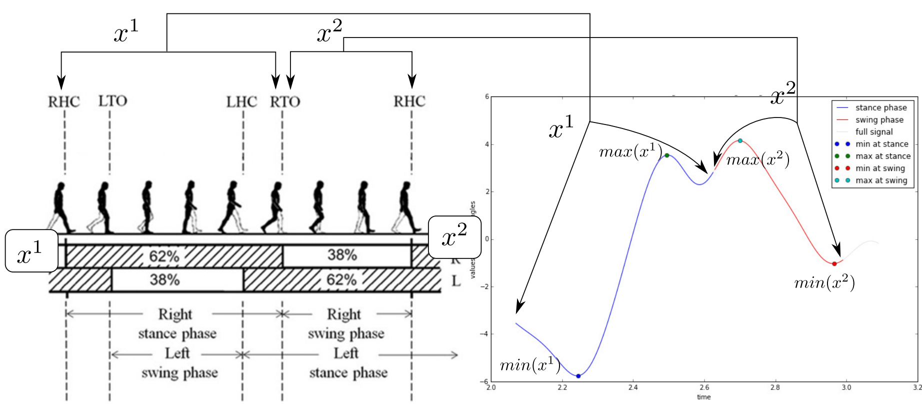


Fig. 3. Schematic (left) and digital (right) representations of the semantic characterization of one full gait cycle into the stance and swing phases (left part of the figure adopted from [12])

We use Mummolo's [12] semantic characterization of a gait cycle which divides every full gate cycle into two phases: stance and the swing phases, which roughly constitute 60% and 40% of the length of a signal (left part of the Figure 3). During one gait cycle, the stance phase is associated with two events: heel strike (HS) in the beginning, and toe off (TO) in the end. The swing phase of the gait cycle takes place between the toe off and the next heal strike (left part of Figure 3). We can thus represent each signal x as two subsignals (right part of the Figure 3) x^1, x^2, such that $x^1 := [x_0 \dots x_i], x^2 = [x_j \dots x_N]$, where $i = \lceil 0.6 \times |x| \rceil$ and $j = \lfloor 0.6 \times |x| \rfloor$ (*i.e.*, i - index of end of stance phase, j - start of swing phase). Subdivision of one signal into its *swing* and *stance* components allows us to perform refined analysis and comparison of signals. In particular, patient's gait data can be characterized by the critical points of the components, *i.e.*, $\min(x^1), \max(x^1), \min(x^2), \max(x^2)$. We also register the index of the critical points (i.e., $\arg\min x^1, x^2, \arg\max x^1, x^2$) to visualize them on the timeseries plot as bold points (right part of the Figure 3).

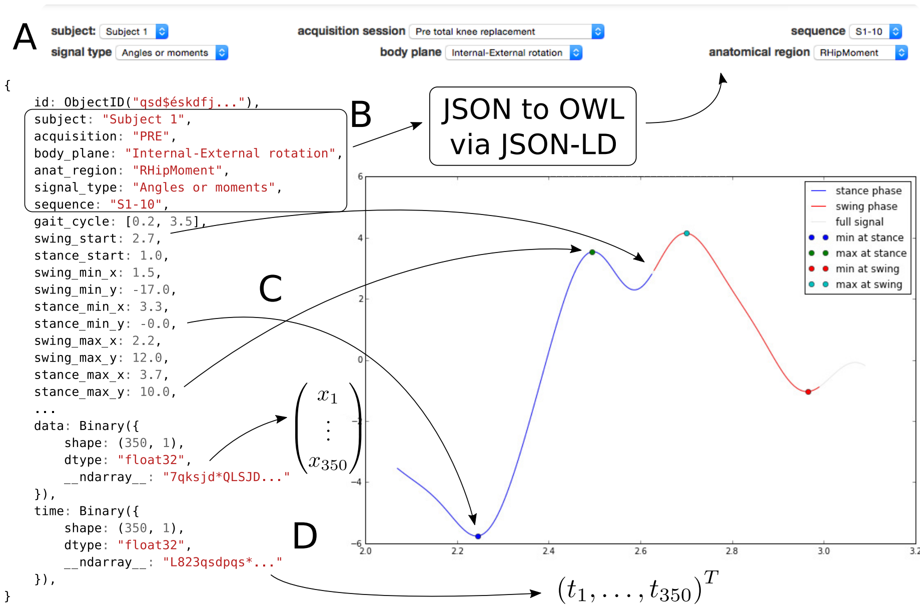


Fig. 4. Semantic characterization and JSON representation of gait signals. A: user interface (UI) for AJAX queries to remotely fetch gait signals (JSON objects). B: mapping between the JSON objects and instances of the ontology in the triple store via JSON-LD. C: semantic characterization of the signal. D: binary numerical data for the signal (time and signal values).

Persistence of gait data to a document-based database (NoSQL). The gait data which we need to store are not uniform data sets. We recurred to NoSQL database rather than a RDBMS, because NoSQL databases provide an agile management of heterogeneous data, as well as multiple analytical tools (aggregation pipeline) [5]. We represent each signal as one JSON object, which contains names of the concepts from the ontology (Figure 4, B), semantic characterization of the signal (Figure 4, C), and the raw numerical data for the signal (Figure 4, D). The raw numerical signal data are represented as two distinct vectors for the time and the value domains, and stored in binary format.

Storing gait data as JSON objects, among other things, enables network transmission of gait data, and thus allows service-oriented software architectures. We provide a simple user interface (UI) which consists of a set of dropbox lists (Figure 4, A) whose values are populated from the knowledge base. The values of these dropbox lists constitute a url query, through the AJAX interface for remote fetching of JSON objects. Example state of the dropbox list set to `Signal type: Angles or moments, Subject: 1, Acquisition session: Pre total knee replacement, Sequence: S4-05, Body plane: Flexion-Extension, Anatomical region: LHipMoment` will produce a url query `http://gaitviewer.plumdeq.xyz/gaitview/angles\_moments/1/PRE/S4-05/Flexion-Extension/LHipMoment/`. The backend replies with a JSON document containing the signal (navigating your browser to the above-mentioned link will fetch the JSON document). The user interface (UI) which consists of a set of dropbox lists (Figure 4, A) whose values are populated from the knowledge base is available at `http://gaitviewer.plumdeq.xyz/gaitview/subjects/`.

Mapping (Figure 4, B) between JSON objects and instances of the ontology in the triple store is realized via JSON linked data technology (JSON-LD), which converts JSON documents annotated with literal strings to RDF (Resource Description Framework) triples. To do so, we define JSON-LD context objects which contain rules of literal string to RDF triples conversion, i.e., transformation of literal strings `Left Knee` stored in JSON objects into respective set of RDF triples. This rule-based RDF triple generation is automatic, and can be modified (i.e., writing appropriate rules) to fit any given ontology vocabulary.

4 Semantic and Interactive Gait Data Visualization and Analysis

We have applied GaitViewer to a dataset consisting of 5 subjects. Each subject has undergone three acquisition sessions: i) before the total knee replacement (TKR) surgery (`PRE`), 3 months later (`POST3M`) and 6 months (`POST6M`) after the surgery. During each acquisition session, in order to improve the quality of measurements, a subject was required to perform the same walking pattern three times, thus for each acquisition session we have three sequences (e.g., `S1-10, S1-11, S1-12`). To perform analysis of data we have focused on joint angles, joint moments variations and on muscle activation (EMG) data. Each type of a biomedical measurement (signal) was recorded with the help of a marker placed on subject's body. Each marker represents a specific anatomical entity (e.g. left hip, gastrocnemius muscle).

Aggregation pipeline for the analysis of gait data. One of the main requirements for a gait analysis framework was the ability to group data over multiple trials per subject, and aggregate computed parameters (features) values for further analysis. We thus proposed to employ the elements of *split-apply-combine* strategy for data analysis [16]. To do so, GaitViewer leverages the aggregation pipeline framework which operates on JSON objects, and is available in some NoSQL databases. In Figure 5 we give an informal and a visual demonstration of the two main operators $f, g : O \mapsto O$ acting on a set of JSON objects O, such that f *filters* objects which satisfy a certain predicate p, while g *aggregates* filtered objects $o_1, \dots o_n$, into one object o. The filter and aggregation functions are operators acting on JSON objects and thus can be composed (i.e., $f \circ f' \circ g$) to produce a rich calculus for a wide range of data processing tasks. A formal treatment and semantics of such calculi for NoSQL languages can be found in Benzaken et al [4].

The aggregation pipeline could be applied to many use-cases, common in data analysis in the human motion research field. We present a use-case in which a user wants to first filter (or group by) each signal by a specific acquisition session, and then to compute (aggregate) average critical points per trial (sequence). We apply (Figure 5) the filter f with a predicate p which checks if the attribute value is equal to `"PRE"` (acquisition session which took place before the total knee replacement), returning the filtered set of objects O_p with all its elements satisfying p, corresponding to the three trials of the same walking pattern (sequences: `S1-10, S1-11, S1-12`). We then apply the aggregation function g on O_p, which computes the mean value of the critical

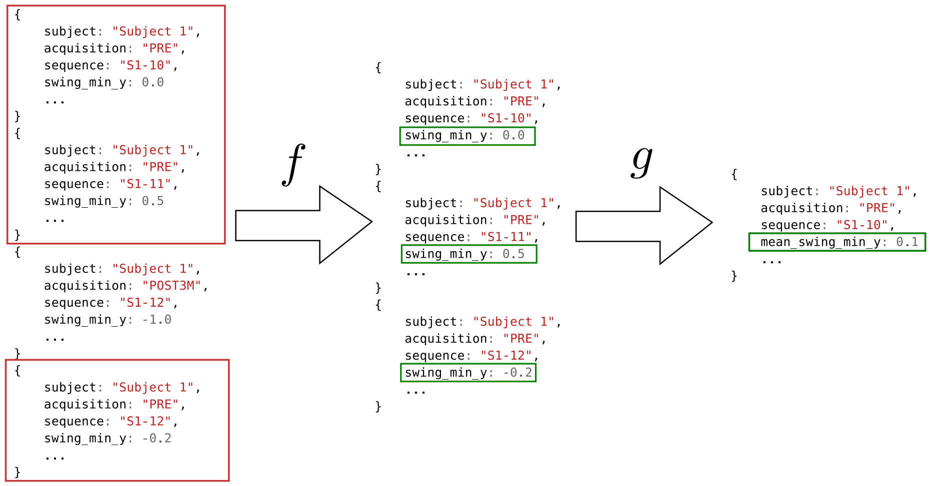
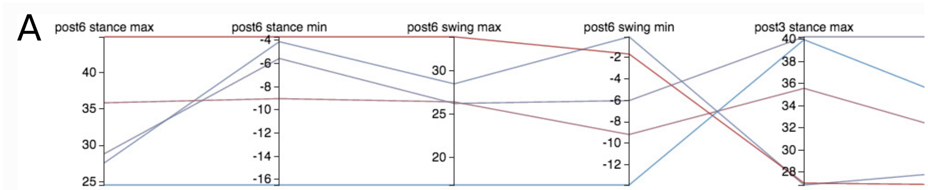


Fig. 5. Application of the aggregation pipeline to compute the average critical points of the swing phase for several trials of a specific acquisition session. Composition of the filter operator f with the aggregation operator g on the gait dataset is schematically presented.

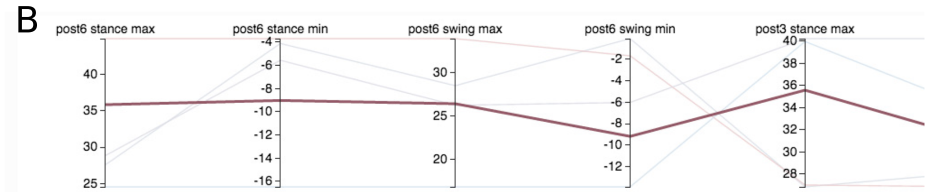
point in the swing phase (`swing_min_y`). Manually, such a data inspection task would have taken relatively small amount of time, however the aggregation pipeline is a more reproducible and a scalable solution.

Visual analytics for gait data correlation with interactive parallel coordinates plots. A recurrent task in movement science is to correlate relevant differences among the subjects, which requires grouping of data over multiple trials per subject, as well as the aggregation of computed parameters (features) values for further analysis. We opt for visual analytics techniques to perform such correlation tasks, for a more intuitive and interactive experience. Since there might be many features or dimensions to consider, we use parallel coordinates [8], a visual analytics tool to summarize this multivariate dataset. For demonstration purposes in this paper, we selected two acquisitions: 3 months and 6 months after the total knee replacement. We characterize all the angles variations of the left hip of 5 subjects during flexion or extension of the hip. We use the aggregation pipeline to average the results. Parallel coordinates (Figure 6 A) help us visually summarize these signals for 5 subjects in a compact manner (due to space limitations we do not show all the coordinates (dimensions)). We us the interactive *focus+context* visualization technique to separate visually one subject from all others. Whenever the user hovers on a specific subject, the other subjects are becoming more transparent (In Figure 6, B, the `Subject 2` is the focus, while all the other subjects are the context).

Visual interface for the aggregation pipeline. We provide a visual interface for the aggregation pipeline, which is implemented via the brushing tool technique, usual to data analysts using parallel coordinates. In Figure 7 we show how the user can identify visual regions on each coordinate, which are then translated into the composition



name	post6 stance max	post6 stance min	post6 swing max	post6 swing min	post3 stance max	post3 stance min
Subject 1	27.58	-4.11	28.46	-0.09	26.81	-14.31
Subject 2	35.83	-9.03	26.39	-9.19	35.57	-9.29
Subject 3	44.9	-3.69	33.86	-1.69	26.98	-16.71
Subject 4	24.54	-16.49	16.85	-13.89	39.98	-4.54
Subject 5	28.84	-5.56	26.21	-6.03	40.24	6.9

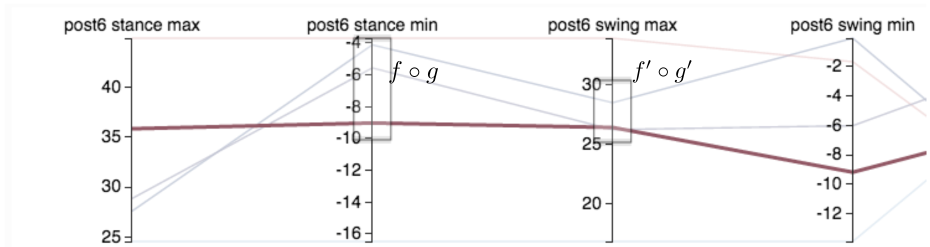


name	post6 stance max	post6 stance min	post6 swing max	post6 swing min	post3 stance max	post3 stance min
Subject 1	27.58	-4.11	28.46	-0.09	26.81	-14.31
Subject 2	35.83	-9.03	26.39	-9.19	35.57	-9.29
Subject 3	44.9	-3.69	33.86	-1.69	26.98	-16.71
Subject 4	24.54	-16.49	16.85	-13.89	39.98	-4.54
Subject 5	28.84	-5.56	26.21	-6.03	40.24	6.9

Fig. 6. A: Parallel coordinates visualization of aggregated signals for all 5 subjects. B: Interactive *focus+context* visualization, where `Subject 2` signals is *focus* and all other subjects' signals are *context* drawn more transparent.

of filter and aggregation operators. For instance, in Figure 7 there are two selected ranges $f \circ g, f' \circ g'$ on coordinates `post6 stance min` and `post6 swing max` respectively. Each visually selected range represents a composition of filter and aggregation operators. $f \circ g$, first filters (f) all signals corresponding to the acquisition session 6 months after the total knee replacement (`post6`), and whose minimum value in the stance phase falls in the period $min(post6_{stance}) \in [-4, -10]$. Finally, g computes the averages for the filtered signals for all trials. f', analogously first selects `post6` signals and those whose maximum value in the swing phase falls in the period $max(post6_{swing}) \in [25, 31]$, and g' computes the averages for all the filtered signals. Since all operators act on and return JSON objects, we can take the union of the results $f \circ g \cup f' \circ g'$, which contains all the signals satisfying the filtering and aggregation constraints imposed by $f \circ g$ and $f' \circ g'$. Note, that the table with all the values is updated accordingly, *i.e.,* only the signals satisfying the constraints are visualized both in the parallel coordinate plot and in the table. Interactive visual selections and parallel coordinates interface, and the underlying aggregation pipeline are available at `http://gaitviewer.plumdeq.xyz/par_coords/`.

Follow-up analysis of gait data. The proposed aggregation pipeline and interactive multivariate visualization via parallel coordinates can be applied for the follow-up analysis of gait data of the same subject. In the follow-up analysis we plot critical values per acquisition session, i.e., the rows are signals corresponding to different acquisition sessions. In Figure 8 we can observe how the critical values of the left hip angles differ



name	post6 stance max	post6 stance min	post6 swing max	post6 swing min
Subject 1	27.58	-4.11	28.46	-0.09
Subject 2	35.83	-9.03	26.39	-9.19
Subject 5	28.84	-5.56	26.21	-6.03

$(f \circ g) \cup (f' \circ g')$

Fig. 7. Visual interface to the aggregation pipeline. Visually selected ranges on the coordinates are translated into data processing operators $f \circ g \cup f' \circ g'$.

in three different acquisition sessions: prior to the total knee replacement, three months after and six months after the surgery. By analyzing the influence of the surgery on subject's gait pattern, i.e., variations of the critical values for different acquisition sessions, we can monitor the recovery progress of the subject. Aggregation pipeline and interactive multivariate visualization could support the practitioners in the follow-up analysis of gait data.

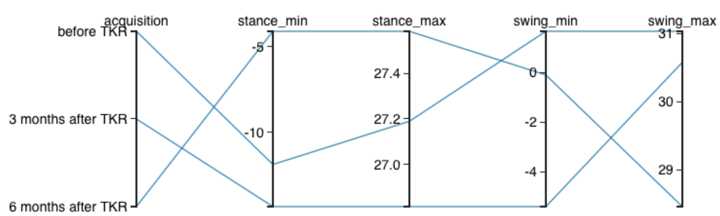


Fig. 8. Variation of the critical values of left hip angles for the subject 1 per three acquisition sessions: before, 3 and 6 months after the total knee replacement (TKR).

5 Conclusion and Discussion

GaitViewer is a collaborative initiative between biomedical engineers, practicing clinical gait analysis, and computer scientists working on biomedical applications of semantic data analysis and visualization. In this work we studied the potential of applying semantic integration techniques, coupled with knowledge discovery and visual analytics techniques for scenarios of collaborative diagnosis of movement disorders. Gait data addressed in this work: joint angles variations, muscle activation - are unstructured text data, each recorded differently, and containing both meta and numerical information.

The heterogeneity of these human motion measurements hinders data interoperability and integrated analysis of gait data for a more complete assessment of patient's locomotion.

We proposed a semantic integration methodology, which separates heterogeneous and mixed numerical and meta information in gait data, such that ontology concepts represent the separated meta information, while numerical information is stored in a NoSQL database. Numerical information is stored as JSON objects in a NoSQL database and is annotated with concepts from the ontology, enabling thus semantic data interoperability of gait data. Parallel coordinates visual analytics technique has been used as an interface to the analytics tools proposed by the NoSQL database. We tailored GaitViewer for two common use-cases in clinical gait analysis: correlation of measured signals for different subjects, and follow-up analysis of the same subject.

The experts feedback and their opinion on the methodology and the approach have been continuously taken into account during the design. From the technical point of view, the experts particularly liked the *interactivity* of GaitViewer. Indeed, the interactive visual selections and parallel coordinates provide an easy to use interface, which hides all the technical complexity of the underlying aggregation pipeline. The data processing operators are composed interactively and intuitively, and the users are not lost in the complex syntax of NoSQL querying languages. Users also liked that, unlike other state of the art gait analysis and visualization tools, which focus on gait data of one subject at a time, GaitViewer provides an integrated environment for analysis and visualization of multiple subjects, taking into account different acquisitions, multiple trials and semantic signal characterizations (stance and swing phases of a gait cycle).

The role of the semantic integration methodology for a large scale open gait data analysis has been regarded as a good potential direction for the human movement research. However, a large-scale adoption of frameworks such as GaitViewer will require involvement of different actors in the field. First, the clinical gait analysis depends on well-established commercial acquisition systems, and their standards for representing gait data. Adoption of a shared semantic layer (ontology of gait data acquisition in this case) will highly depend on the flexibility of the manufacturers, and possible creation of an international consortium responsible for issuing recommendations for the open gait data standards. Second, creation of an online and open portal for gait datasets will require approval from medical ethical committees, and creation of rules for online publishing of anonymized subjects datasets.

Acknowledgements This work was partially supported by the EU Marie Curie ITN MultiScaleHuman (FP7-PEOPLE-2011-ITN, Grant agreement no.: 289897) and by the CNR project DIT.AD009.006 Modelling and Analysis of anatomical shapes for computer assisted diagnosis. All authors would like to thank anonymous reviewers for their suggestions, and the whole Shape Modelling Group at IMATI-CNR for the useful comments to improve the paper.

References

1. Baker, R.: Gait analysis methods in rehabilitation. Journal of Neuroengineering and Rehabilitation 3, 4 (2006)

2. Barre, A., Armand, S.: Biomechanical toolkit: Open-source framework to visualize and process biomechanical data. Computer methods and programs in biomedicine 114(1), 80–87 (2014)
3. Belleau, F., Nolin, M.A., Tourigny, N., Rigault, P., Morissette, J.: Bio2rdf: Towards a Mashup to Build Bioinformatics Knowledge Systems. J. of Biomedical Informatics 41(5), 706–716 (Oct 2008), `http://dx.doi.org/10.1016/j.jbi.2008.03.004`
4. Benzaken, V., Castagna, G., Nguyen, K., Siméon, J.: Static and Dynamic Semantics of NoSQL Languages. In: Proceedings of the 40th Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages. pp. 101–114. POPL '13, ACM, New York, NY, USA (2013), `http://doi.acm.org/10.1145/2429069.2429083`
5. Berthelsen, E.: Why nosql databases are needed for the internet of things. Research Note, Machina Research (Apr 2014), `https://machinaresearch.com/report/research-note-why-nosql-databases-are-needed-for-the-internet-of-things/`
6. Burger, B., Toiviainen, P.: Mocap toolbox-a matlab toolbox for computational analysis of movement data. In: Proceedings of the Sound and Music Computing Conference 2013, SMC 2013, Logos Verlag Berlin, Stockholm, Sweden. ISBN 978-3-8325-3472-1. Logos Verlag Berlin (2013)
7. Federolf, P., Tecante, K., Nigg, B.: A holistic approach to study the temporal variability in gait. Journal of Biomechanics 45(7), 1127–1132 (2012), `http://dx.doi.org/10.1016/j.jbiomech.2012.02.008`
8. Heinrich, J., Weiskopf, D.: State of the Art of Parallel Coordinates. In: Eurographics 2013-State of the Art Reports. pp. 95–116. The Eurographics Association (2013)
9. Kogalovsky, M.R.: Ontology-based data access systems. Programming and Computer Software 38(4), 167–182 (Jul 2012), `http://link.springer.com/article/10.1134/S0361768812040032`
10. Manal, K., Stanhope, S.J.: A novel method for displaying gait and clinical movement analysis data. Gait & posture 20(2), 222–226 (2004)
11. MSH: MSH ontology: deliverable reports D8.2 (m24, m36) and OWL file. Accessed July 28, 2015, `http://multiscalehuman.miralab.ch/repository/Public\_download/D8.2\_MSD-Ontology/`
12. Mummolo, C., Mangialardi, L., Kim, J.H.: Quantifying dynamic characteristics of human walking for comprehensive gait cycle. Journal of Biomechanical Engineering 135(9), 91006 (Sep 2013)
13. Plugge, E., Hawkins, T., Membrey, P.: The Definitive Guide to MongoDB: The NoSQL Database for Cloud and Desktop Computing. Apress, Berkely, CA, USA, 1st edn. (2010)
14. Simon, S.R.: Quantification of human motion: gait analysis-benefits and limitations to its application to clinical problems. Journal of Biomechanics 37(12), 1869–1880 (Dec 2004)
15. Spahić, D., Cardiff, P., Flavin, R., Karač, A., Zenica, F.: Visualization of human motion using paraview-open source scientific visualization. In: Proceedings of the Trends in the Development of Machinery and Associated Technology Conference, Prague, Czech Republic (2011)
16. Wickham, H.: The split-apply-combine strategy for data analysis. Journal of Statistical Software 40(1), 1–29 (2011)
17. Winter, D.A.: Biomechanics and motor control of human movement. John Wiley & Sons (2009)
18. Yang, M., Zheng, H., Wang, H., McClean, S., Newell, D.: igait: An interactive accelerometer based gait analysis system. Computer methods and programs in biomedicine 108(2), 715–723 (2012)