

LinkedECG: Uma Abordagem para a Integração e Publicação de Dados de Eletrocardiograma

Douglas Torquato¹, Daniel Rodrigues¹, José Maria Monteiro¹, João Paulo Madeiro², Angelo Brayner¹, Vânia Vidal¹, Narciso Arruda¹, Tiago Vinuto¹

¹MDCC – Universidade Federal do Ceará (UFC)
Fortaleza – CE – Brasil

²IEDS – UNILAB
Redenção, CE – Brasil

{douglas,daniel,monteiro,brayner,vvidal,narciso,tiagosv}@lia.ufc.br

jpaulo.vale@unilab.edu.br

Abstract. *The electrocardiogram (ECG) is a widespread medical procedure in the cardiology field due to the fact that it is a fast, low-cost and non-invasive examination, and its analysis allows anomalies to be detected and interpreted by health experts. However, the data that comprise or may be extracted from the ECG signals are quite complex, heterogeneous and do not follow a single standardization. In this context, the present work proposes an approach, named LinkedECG, for ECG feature extraction (resulting from signal processing), ECG data integration and publishing following the main standards related to open data sharing on the web. The proposed methodology yields that data extracted from a set of collected signals and/or from publicly available signal records be integrated and published, in order to support, using Web environment, complex queries, execution of mining algorithms, and also enable collaboration among specialists.*

Resumo. *O eletrocardiograma (ECG) é um exame bastante difundido na área de cardiologia, devido ao fato de ser um procedimento simples, de baixo custo e não-invasivo. O seu estudo permite que doenças e anomalias cardíacas possam ser detectadas por especialistas da área. Contudo, os dados que compõem ou que podem ser extraídos dos sinais ECG são bastante complexos, heterogêneos e não seguem uma única padronização. Neste contexto, este trabalho propõe uma abordagem, denominada LinkedECG, para a extração de parâmetros decorrentes do processamento do sinal, integração e publicação de dados de ECG, seguindo os principais padrões para o compartilhamento de dados abertos na Web. A metodologia proposta permite que dados extraídos de um conjunto de sinais coletados e/ou disponíveis em bases de dados públicas sejam integrados e publicados, de forma a oferecer suporte, no próprio ambiente Web, a consultas complexas, execução de algoritmos de mineração, além de possibilitar a colaboração entre especialistas.*

1. Introdução

A Eletrocardiografia é uma técnica utilizada para registrar as alterações de potencial elétrico produzidas pela atividade cardíaca. A evolução temporal das referidas alterações é chamada de sinal eletrocardiograma (ECG) [Geselowitz 1989], o qual é o teste mais difundido na Cardiologia para o diagnóstico de doenças e anomalias cardíacas [Gonçalves et al. 2007]. A análise do comportamento do sinal ECG permite extrair informações diversificadas, as quais podem subsidiar a identificação de uma grande variedade de doenças cardíacas.

Devido ao grande interesse no estudo e interpretação do conteúdo do sinal ECG, diversos padrões de armazenamento dos dados foram criados [Gonçalves et al. 2007]. Esses padrões têm por objetivo: (1) permitir o armazenamento de arquivos de sinais ECG em prontuários eletrônicos (EHR); (2) facilitar a comunicação/intercâmbio dos resultados de exames cardíacos entre diferentes profissionais e/ou hospitais [Gonçalves et al. 2007]. Dentre os principais padrões criados, destacam-se: (i) AHA/MIT-BIH (Physionet) [Goldberger et al. 2000]; (ii) SCP-ECG [Mandellos et al. 2010] e (iii) HL7 aECG [Bond et al. 2011]. Apesar desses padrões tratarem de um mesmo domínio do conhecimento, os conceitos adotados são heterogêneos e distintos entre si. Este cenário dificulta, portanto, a integração de dados de ECG que estão armazenados em padrões diferentes.

Por outro lado, o estudo sobre a Web de Dados vem crescendo a cada ano. Essa nova Web se baseia em um conjunto de melhores práticas para publicação e consumo de dados estruturados, mais conhecido como *Linked Data*, possibilitando a integração entre itens de diferentes bases de dados para formar um único espaço de dados global [Heath and Bizer 2011]. As premissas de *Linked Data* fundamentam-se nas tecnologias da Web Semântica e permitem reduzir a complexidade de integrações devido às ligações estabelecidas e descritas entre os conjuntos de dados. O uso de um modelo de dados padronizado (RDF) e um mecanismo padronizado de consulta (linguagem de consulta SPARQL) simplificam ainda mais a integração de dados.

O principal objetivo deste trabalho consiste em propor uma abordagem de extração automática de parâmetros, decorrentes do processamento de sinais ECG, integração e publicação de dados de sinais ECG no formato RDF, com base nas tecnologias da Web Semântica e nos padrões de *Linked Data*. Esta abordagem possibilita a realização de consultas na linguagem SPARQL, bem como a execução de inferências sobre os dados.

O artigo está estruturado como segue. Na seção 2, são discutidos os trabalhos relacionados. Na seção 3, são detalhadas as diferentes etapas da abordagem proposta. Na seção 4, é apresentado um estudo de caso com a finalidade de avaliar a abordagem proposta. Finalmente, na seção 5, são apresentadas as conclusões e trabalhos futuros.

2. Trabalhos Relacionados

Gonçalves [Gonçalves et al. 2007] propôs uma ontologia para representar sinais de ECG, a qual independe da aplicação e da linguagem de codificação. A ontologia proposta é descrita através da linguagem OWL (Web Ontology Language). Adicionalmente, o autor descreve como mapear cada classe de cada um de diferentes formatos de dados ECG (AHA/MIT-BIH, SCP-ECG e FDA XML/HL7 aECG) com a correspondente classe na ontologia proposta. Raimond L. Winslow disponibilizou um *website* chamado *BioPortal*¹, que publica diversas ontologias biomédicas. Contudo, a estrutura

das ontologias disponíveis apresenta poucas semelhanças com a ontologia proposta em [Gonçalves et al. 2007]. Em [Tanantong et al. 2011] encontra-se uma ontologia para realizar inferências sobre tipos de arritmias com base em dados de ECG. Um fragmento da ontologia de [Gonçalves et al. 2007] é utilizado neste artigo.

Em [Ngo and Veeravalli 2014], os autores propõem uma plataforma baseada nas tecnologias da Web Semântica que permitem o armazenamento de parâmetros extraídos do sinal ECG em uma base de dados. Em [Trigo et al. 2012], Jesús Daniel Trigo et al. propõem o projeto e o desenvolvimento de um sistema para o gerenciamento interoperável de diferentes formatos de registros digitais de sinais ECG. Os autores propõem uma combinação de diferentes formatos de armazenamento de registros de ECG. Em [Khumrin and Chumpoo 2016], os autores propõem uma abordagem para a integração de dados eletrocardiográficos, originados de diferentes formatos, utilizando-se um formato de dados de referência baseado em classes Java.

3. A Abordagem LinkedECG

A Figura 1 ilustra as etapas da abordagem proposta (LinkedECG): extração de parâmetros, extração de dados e publicação de dados. Essas etapas serão descritas a seguir.

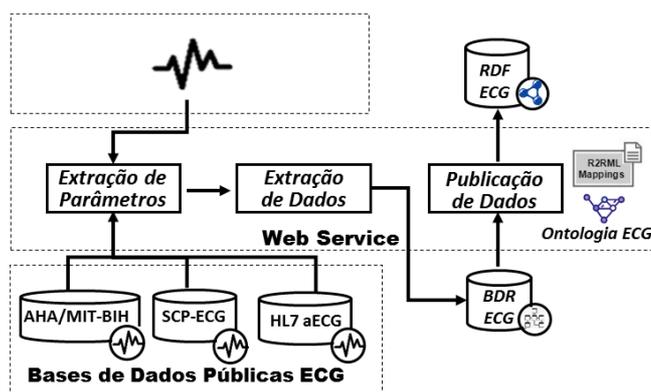


Figura 1. Visão Geral da Abordagem LinkedECG

3.1. Extração de Parâmetros (Features)

Esta etapa recebe como entrada um sinal ECG bruto, o qual pode ter sido originado de um biosensor ou de bases públicas previamente existentes (como por exemplo, AHA/MIT-BIH, SCP-ECG e HL7 aECG). Em seguida, um conjunto de algoritmos de processamento digital de sinais é executado, incluindo: filtragem para eliminação de ruído e interferências, aplicação de transformada *Wavelet* (análise tempo-frequência) para realce seletivo do complexo QRS e das ondas P e T, detecção dos picos e delineamento das formas de onda, extração de parâmetros (intervalos e amplitudes) [Madeiro et al. 2012]. Como resultado desta etapa, os seguintes parâmetros são extraídos: amplitude e duração de cada complexo QRS, intervalos entre batimentos, amplitude e duração de cada onda P, amplitude e duração de cada onda T, intervalos entre as diferentes formas de onda.

3.2. Extração de Dados

Esta etapa recebe como entrada uma matriz contendo os parâmetros extraídos do sinal ECG. Em seguida, esta matriz é processada e um conjunto de dados sobre o sinal ECG é

extraído e armazenado em banco de dados relacional, cujo esquema é ilustrado na Figura 2. A abordagem de armazenar os dados do sinal ECG em um banco relacional foi adotada com o objetivo de simplificar o processo de publicação dos dados, uma vez que já existem ferramentas que possibilitam a criação de *dumps* dos dados relacionais em formato RDF e a disponibilização desses *dumps* em *triplostore* RDF, de forma semiautomática.

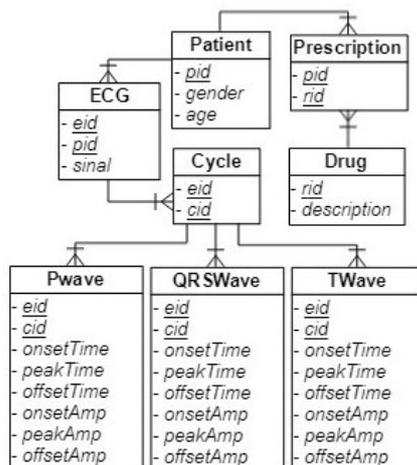


Figura 2. Esquema do banco de dados relacional.

3.3. Publicação de Dados

Nesta etapa, o banco relacional contendo as informações extraídas dos sinais ECG é acessado e os seus dados são exportados para o formato RDF. Dividimos esse processo em dois passos: No primeiro, cria-se um *dump* dos dados relacionais em formato RDF. Para realizar este passo, utilizamos a ferramenta *D2RQ*¹ junto com os mapeamentos na linguagem *R2ML*², os quais relacionam o esquema do banco de dados relacional com o vocabulário da ontologia adotada para representar os sinais ECG. No segundo passo, os dados em formato RDF presentes no *dump* gerado anteriormente são materializados em um *triplostore* RDF, mais especificamente no *Virtuoso*³, de forma semiautomática. O *Virtuoso* disponibiliza um *SPARQL endpoint* que possibilita realizar consultas semânticas.

3.3.1. Adaptação de uma Ontologia de Sinais ECG

Inicialmente, buscamos utilizar uma ontologia de sinais ECG já existente. A ontologia proposta em [Gonçalves et al. 2007] foi a que mais se aproximou dos requisitos definidos para a abordagem *LinkedECG*. Desta forma, utilizamos os vocabulários *ecg* :< <http://nemo.inf.ufes.br/biomedicine/ecg.html> >, disponível no trabalho de [Gonçalves et al. 2007], e *health* :< <https://health-lifesci.schema.org/> >⁴ (termos adicionais). O vocabulário *ecgo* :< <http://www.arida.ufc.br/ecg> > contém termos criados pela abordagem proposta. A Figura 3 apresenta a ontologia utilizada nesta etapa.

¹<http://d2rq.org/>

²<https://www.w3.org/TR/r2rml/>

³<https://virtuoso.openlinksw.com/rdf/>

⁴Disponível em <https://health-lifesci.schema.org/Patient>

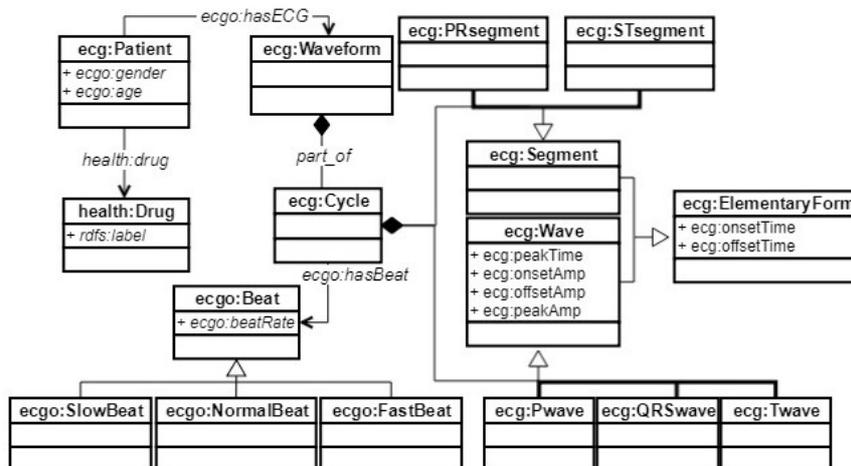


Figura 3. Ontologia utilizada na etapa de publicação dos dados.

4. Estudo de Caso

Realizar consultas sobre dados de sinais ECG é uma tarefa bastante complexa, uma vez que os dados são armazenados em arquivos de texto não estruturados. Por outro lado, a abordagem LinkedECG possibilita a execução de consultas semânticas. Para demonstrar este fato, apresentamos duas consultas que foram executadas sobre uma base de dados RDF gerada aplicando-se a abordagem LinkedECG.

Consulta Sparql 01: Quais pacientes do sexo masculino com mais de 60 anos tiveram algum batimento acelerado (acima de 100 batimentos por minuto) no ECG?

```

SELECT ?paciente
WHERE {?paciente a health:Patient ; health:age ?idade ;
          health:gender ?sexo ; ecga:hasECG ?ecg .
          ?ciclo ecgo:part_of ?ecg .
          ?ciclo ecgo:hasBeat ?bat .
          ?bat a ecgo:FastBeat.
FILTER (?age >60 && ?sexo = "male")}
```

Figura 4. Consulta Sparql 01.

Consulta Sparql 02: Quais pacientes tomaram o medicamento *Aldomet* e apresentaram algum batimento cardíaco lento (abaixo de 60 batimentos por minuto) no ECG?

```

SELECT ?paciente
WHERE {?paciente a health:Patient ; health:drug ?med .
          ?med rdfs:label ?nomemed .
          ?paciente ecgo:hasECG ?ecg .
          ?ciclo ecgo:part_of ?ecg ; ecgo:hasBeat ?bat .
          ?bat a ecgo:SlowBeat.
FILTER (?nomemed = "Aldomet")}
```

Figura 5. Consulta Sparql 02.

5. Conclusões

Neste trabalho apresentamos uma abordagem, denominada LinkedECG, para a integração e publicação de dados de sinais ECG. A abordagem proposta possibilita a criação de uma

base de conhecimento pública, que pode ser usada para dar suporte a consultas complexas. Um estudo de caso foi realizado com a finalidade de comprovar que a LinkedECG facilita a execução de consultas que dificilmente seriam executadas sobre os arquivos contendo os sinais ECG brutos.

Como trabalhos futuros, iremos realizar testes de desempenho com a finalidade de avaliar o impacto proporcionado pela manipulação de dados RDF no tempo de resposta das consultas, bem como a escalabilidade da abordagem proposta. Além disso, iremos adicionar novas bases públicas de sinais ECG e explorar a utilização de algoritmos de mineração para a classificação/reconhecimento de arritmias e outros eventos adversos.

Referências

- Bond, R. R., Finlay, D. D., Nugent, C. D., and Moore, G. (2011). A review of ECG storage formats. *International journal of medical informatics*, 80(10):681–697.
- Geselowitz, D. B. (1989). On the theory of the electrocardiogram. *Proceedings of the IEEE*, 77(6):857–876.
- Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C.-K., and Stanley, H. E. (2000). Physiobank, physiotoolkit, and physionet. *Circulation*, 101(23):e215–e220.
- Gonçalves, B., Guizzardi, G., and Pereira Filho, J. G. (2007). An electrocardiogram (ECG) domain ontology. In *Workshop on Ontologies and Metamodels for Software and Data Engineering, 2nd, João Pessoa, Brazil*, pages 68–81.
- Heath, T. and Bizer, C. (2011). Linked data: Evolving the web into a global data space. *Synthesis lectures on the semantic web: theory and technology*, 1(1).
- Khumrin, P. and Chumpoo, P. (2016). Implementation of integrated heterogeneous electronic electrocardiography data into maharaj nakorn chiang mai hospital information system. *Health informatics journal*, 22(1):34–45.
- Madeiro, J. P., Cortez, P. C., Marques, J. A., Seiseddos, C. R., and Sobrinho, C. R. (2012). An innovative approach of QRS segmentation based on first-derivative, hilbert and wavelet transforms. *Medical engineering & physics*, 34(9):1236–1246.
- Mandellos, G. J., Koukias, M. N., Styliadis, I. S., and Lymberopoulos, D. K. (2010). e-scp-ecg+ protocol: An expansion on SCP-ECG protocol for health telemonitoring—pilot implementation. *International journal of telemedicine and applications*, 2010:1.
- Ngo, D. and Veeravalli, B. (2014). Applied semantic technologies in ecg interpretation and cardiovascular diagnosis. In *Bioinformatics and Biomedicine (BIBM), 2014 IEEE International Conference on*, pages 17–24. IEEE.
- Tanantong, T., Nantajeewarawat, E., and Thiemjarus, S. (2011). Towards continuous electrocardiogram monitoring based on rules and ontologies. pages 327–330.
- Trigo, J. D., Martínez, I., Alesanco, A., Kollmann, A., Escayola, J., Hayn, D., Schreier, G., and García, J. (2012). An integrated healthcare information system for end-to-end standardized exchange and homogeneous management of digital ECG formats. *IEEE Transactions on Information Technology in Biomedicine*, 16(4):518–529.