

PLATAFORMAS DE PROCESSAMENTO E ARMAZENAMENTO DE GRANDES VOLUMES DE DADOS DE REGISTROS ELETRÔNICOS EM SAÚDE

Charles Jefferson Rodrigues Alves e Alves¹

¹Mestrando do Programa de Pós-Graduação em Modelagem Computacional de Sistemas – UFT. e-mail: charles.alves@ifto.edu.br

Resumo: Os registros eletrônicos em saúde capazes de realizar agrupamento para posterior análise das informações em saúde têm recebido crescente atenção por parte de uma ampla gama de organizações privadas e governamentais. Para alcançar esse objetivo de forma satisfatória é necessário transpor uma série de barreiras tais como: uso de diferentes terminologias em saúde, a grande dificuldade inerente à implementação da interoperabilidade semântica, o agrupamento, manipulação e a mineração de grandes volumes de dados de forma segura, entre outros. O objetivo desse estudo foi o de possibilitar uma visão sobre o estado da arte das publicações cuja finalidade era a concepção, implementação ou a avaliação de plataformas de processamento e armazenamento de grandes volumes de dados de registros eletrônicos em saúde, utilizando o método de mapeamento sistemático da literatura. A busca pelas publicações foi realizada nas bases de dados: ACM, Elsevier, IEEE, Emerald, ISI Web of Science, Wiley InterScience e PubMed, no período de 2005 a 2014

Palavras-chave: banco de dados distribuídos, gestão da informação, mapeamento sistemático, registros eletrônicos em saúde

1. INTRODUÇÃO

As técnicas associadas à computação distribuída, à computação em nuvem, à big data e à conexão segura, estável e rápida são aplicáveis a diversos ramos de atividades, tais como grandes bancos, grandes estudos científicos que exigem enorme demanda computacional, receita federal, etc. No caso da área da saúde essas técnicas podem ser utilizadas para agrupar dados provenientes de prontuários eletrônicos de diferentes pacientes, espalhados geograficamente, na forma de uma grande base de dados chamada de registro eletrônico em saúde. Esse registro eletrônico em saúde, por sua vez, é alimentado pelos dados provenientes de diversos prontuários eletrônicos do paciente. O prontuário eletrônico do paciente é um conjunto de informações individualizadas por paciente onde consta toda a sua história clínica para uma dada instituição de saúde. É a base para as principais atividades realizadas em hospitais, tanto na área de saúde quanto administrativa. Esse prontuário, por determinação legal, fica sob a responsabilidade da instituição de saúde onde o paciente foi atendido. Tem como principais peças a anamnese, a prescrição de medicamentos e o resultado de exames¹.

O objetivo desse estudo foi responder a seguinte questão utilizando o método de mapeamento sistemático da literatura: Qual é o estado da arte das publicações cuja finalidade é a concepção, implementação ou a avaliação de plataformas de processamento e armazenamento de grandes volumes de dados de registros eletrônicos em saúde?

2. MATERIAL E MÉTODOS

O mapeamento sistemático da literatura emprega um processo metódico para identificar, avaliar e interpretar todas as evidências científicas disponíveis e relevantes relacionadas a um tema específico de interesse². Dessa forma, nós realizamos uma extensa pesquisa da literatura nas bases de dados: ACM Digital Library (portal.acm.org), ScienceDirect Elsevier

(www.elsevier.com), IEEE Xplorer (ieeexplore.ieee.org), Emerald (www.emeraldinsight.com), ISI Web of Science (www.isiknowledge.com), Wiley InterScience (www.interscience.wiley.com) e PubMed (www.pubmed.gov), com o objetivo específico de avaliar o estado da arte das pesquisas que tratam das estratégias e tecnologias empregadas no processamento e armazenamento distribuído para registros eletrônicos em saúde.

Foram utilizadas as seguintes combinações de palavras-chave: electronic health record AND distributed database, electronic health record AND database architecture, electronic health record AND big data, electronic health record AND distributed data bank, electronic health record AND data bank architecture, electronic medical record AND distributed database, electronic medical record AND database architecture, electronic medical record AND big data, electronic medical record AND distributed data bank, electronic medical record AND data bank architecture, cloud computing AND electronic health record, cloud computing AND electronic medical record, cloud computing AND distributed database, cloud computing AND distributed data bank, distributed database AND big data, distributed data bank AND big data.

Os artigos primários identificados nesse estudo foram obtidos a partir de buscas realizadas através das combinações das palavras-chave em cada base de dados, conforme descrito acima. Essa busca rendeu 1794 estudos primários. Através da leitura dos títulos e/ou resumos desses 1794 estudos primários, baseando-se em critérios de inclusão e exclusão, foram selecionados 24 artigos com possibilidade de serem incluídos nesse mapeamento sistemático. Os critérios de inclusão foram: a) possuir uma relação com o tema; b) estar no idioma inglês; c) publicações em revistas, conferências, simpósios, *workshops*, periódicos e outros eventos científicos; d) estudos com ano de publicação igual ou superior a 2005; e) artigos que descreveram com profundidade as arquiteturas propostas ou testadas para registro eletrônico em saúde e big data. Os critérios de exclusão foram: a) não estar relacionado ao tema proposto; b) trabalhos duplicados existentes em diferentes versões, onde será incluída a versão mais completa do estudo. Os 24 artigos selecionados obedeceram a todos os itens de inclusão e não atenderam a nenhum critério de exclusão. Por fim, após a leitura completa dos 24 estudos primários selecionados, incluímos 13 artigos nesse mapeamento sistemático, conforme Tabela 1.

Tabela 1: Número de publicações encontradas nas bases de dados

Bases	Publicações Identificados	Publicações Não Selecionados	Publicações Selecionados	Publicações Excluídos	Publicações Incluídos
ACM	894	890	4	1	3
Elsevier	8	6	2	1	1
IEEE	654	639	15	7	8
Emerald	4	4	0	0	0
ISI Web of Science	0	0	0	0	0
Wiley InterScience	211	211	0	0	0
PubMed	23	20	3	2	1
TOTAL	1794	1770	24	11	13

3. RESULTADOS E DISCUSSÃO

Nos resultados coletados e análises realizadas, deu-se ênfase para os conteúdos direcionados a responder os objetivos do mapeamento sistemático.

Na Figura 1 observa-se a distribuição dos artigos incluídos no estudo, agrupados por base de dados. É mister afirmar que os artigos incluídos repetidos em bases diferentes foram contados quantas vezes aparecessem na coleta.

Neste trabalho mapeamos oito diferentes nacionalidades com trabalhos relevantes. Nessa relação de países, dois deles se destacam (China e Estados Unidos), somando juntos 54% do total das publicações no período analisado, Figura 1.

Os estudos incluídos foram sondados acerca da formação de todos os seus autores. Após análise da formação dos mesmos, percebe-se que a grande maioria dos trabalhos incluídos nesse estudo apresentava apenas autores da área da ciência da computação (62%), uma menor parcela dos trabalhos apresentava apenas autores da área da saúde (23%) e uma parcela ainda menor apresentava autores de ambas às áreas (15%), Figura 1.

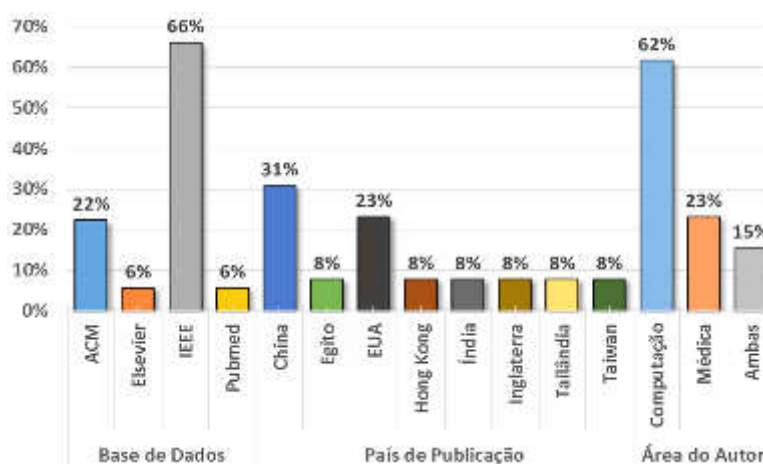


Figura 1 - Classificação dos resultados da revisão

Outro dado importante extraído da nossa avaliação mostra claramente a tendência dos grupos de pesquisa compostos exclusivamente por profissionais da área da computação em publicar em revistas de tecnologia da informação indexadas principalmente nas bases IEEE e ACM. Além disso, encontramos poucos trabalhos publicados por grupos de pesquisa mistos (compostos por profissionais da área da saúde e de computação).

Tabela 2: Revistas científicas

Nome do Periódico	Total de Artigos	Citações por Revista	Base de Dados
International Symposium on IT in Medicine and Education	2	6	IEEE
International Conference on e-Health Networking, Applications & Services	2	4,5	ACM/IEEE
International Conference on Internet Computing for Science and Engineering	1	3	ACM/IEEE
International Joint Conference on Computer Science and Software Engineering	1	0	IEEE
International Conference for Internet Technology and Secured Transactions	1	3	IEEE
International Conference on Social Computing	1	3	IEEE
International Multitopic Conference	1	3	IEEE
Journal of Biomedical and Health Informatics	1	17	IEEE
International Conference on Information Management Innovation Management and Industrial Engineering	1	9	IEEE
Computer Methods and Programs in Biomedicine	1	11	Elsevier/PubMed/ACM
Computer	1	0	IEEE

O total dos 13 artigos incluídos nesse estudo foi oriundo de 10 periódicos distintos indexados nas bases de dados sondadas. Na Tabela 2 são mostrados os nomes de todos os periódicos que tiveram seus artigos incluídos, a soma do número de citações de cada artigo incluído proveniente desse periódico dividido pela quantidade de artigos do mesmo periódico e a base de dados onde se encontra o periódico. Essa tabela mostra que atualmente não temos um periódico especializado, específico e único para esse tipo de publicação.

A Figura 2 apresenta a evolução dos artigos encontrados conforme o local de sua publicação. Interessante notar que a sua maioria é proveniente de conferências ou simpósios e apenas um artigo oriundo de revista científica, significando que área em estudo não está amadurecida.

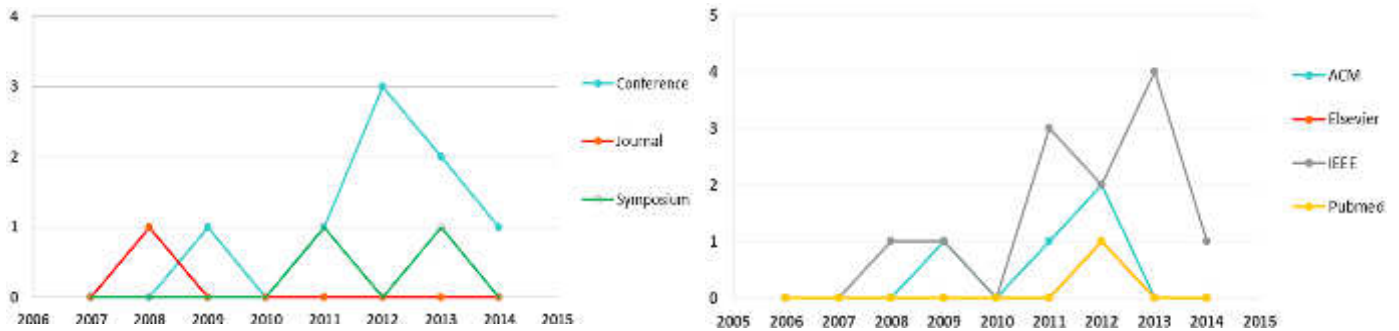


Figura 2 - Evolução dos estudos por tipo de publicação e por base de dados

Com a finalidade de descobrir se existe uma rede de pesquisa mundial para a realização de estudos nessa área, classificamos os grupos de pesquisa que participaram da produção de todos os 13 artigos envolvidos nesse trabalho. Os grupos de pesquisa classificados foram conectados entre si através dos artigos científicos que eles tinham publicado em comum. A Figura 3 mostra claramente que esses grupos de pesquisa classificados não formam uma rede de pesquisa ampla e mundial que seja voltada para discussão do tema objeto de estudo desse mapeamento. Podemos deduzir que os estudos mais citados são constituídos por grupos formados por membros unicamente da ciência da computação^{S8, S9}. Os grupos de pesquisa elencados na figura abaixo são representados pelas bolinhas com a inicial “D” e encontram-se listados detalhadamente no Quadro 1. Os retângulos azuis indicando os artigos primários utilizados em nosso mapeamento sistemático, juntamente com a quantidade de citações na base Google Scholar entre parênteses, também se encontram listados detalhadamente nas referências do artigo para consulta.

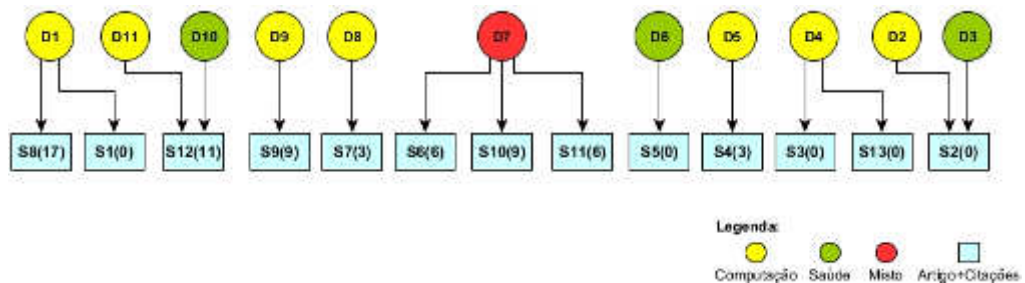


Figura 3 - Rede de grupos de pesquisa conectados pelas publicações em comum

Por fim realizou-se um mapeamento dos artigos incluídos nesse estudo quanto à evolução anual do número de publicações na área por cada base de dados, conforme exposto na Figura 2. Através da figura percebe-se que o interesse em realizar investigações nessa área encontra-se em plena ascensão, com especial importância para a base de dados do IEEE.

Quadro 1 - Grupos de Pesquisas

6. CONCLUSÕES

Nesse trabalho, buscou-se mapear publicações que estudam ou propõe soluções para os problemas de armazenamento de “big data” em saúde, apresentando dados relevantes sobre o estado da arte nessa área. Os dados obtidos nas análises revelaram que esse campo de estudo vem crescendo ano após ano dentro das bases pesquisadas. Nesse processo, um número bem pequeno

- D1 - Department of Electrical and Computer Engineering - Georgia Institute of Technology - EUA
 D2 - College of Computer Science and Technology - Harbin Engineering University - China
 D3 - Carver College of Medicine - The University of Iowa - EUA
 D4 - College of Computer Engineering - Arab Academy of Science and Technology and Maritime Transport - Egito
 D5 - Graduate Institute of Biomedical Informatics - Taipei Medical University - Taiwan
 D6 - College of Medical Information Engineering - GuangDong Pharmaceutical University - China
 D7 - School of Engineering & Design - Brunel University - Inglaterra
 D8 - Department of Computer Science and Engineering - Indian Institute of Technology - Índia
 D9 - Centre for Integrative Digital Health, School of Nursing - The Hong Kong Polytechnic University - Hong Kong
 D10 - Information Technology Committee - Hong Kong Doctors Union - Hong Kong
 D11 - Department of Computer Engineering - San Jose State University - EUA

de periódicos começa a se destacar dos demais em importância quando se observam o número de artigos publicados na área do presente estudo e o número de citações desses artigos.

Nossos dados também apontaram que a maioria dos estudos na área é realizada por apenas dois países (China e EUA). As bases de dados com maior importância são aquelas relacionadas à engenharia e ciência da computação (IEEE e ACM) refletindo o fato de que a maior parte dos estudos está sendo realizado por grupos formados apenas por indivíduos com formação na área de ciência da computação. Os grupos que pesquisam nessa área não estão organizados na forma de uma rede global que se comunica entre si e discutem quais seriam as melhores estratégias para resolver um problema tão complexo e importante.

Ao analisar esse panorama percebemos que esse campo de pesquisa se encontra em fase de incubação, onde os grupos começam a se formar e realizar estudos individualmente ou associados a outros poucos grupos. O fato da maioria dos grupos não ser de formação mista prejudica o desenvolvimento da área pelo fato da mesma ser de natureza multidisciplinar. Acreditamos que pela importância do problema a tendência futura será a de conexão entre grupos que se encontram disperso atualmente, além do amadurecimento dos mesmos, se tornando grupos de formação mista. A tendência de aumento no número de publicações nessa área ano após ano, e o aumento de participantes nos Congressos de Informática em Saúde ao redor do mundo discutindo “big data” em saúde, servirão como catalisadores desse processo.

REFERÊNCIAS

- [1] Aduino TEIXEIRA, Josenir e SUANNES. Notas Sobre o Prontuário do Paciente. Pró-Saúde - Associação Beneficente de Assistência Social e Hospitalar. São Paulo, 1998.
- [2] KIETCHENHAM, B. A. Guidelines for performing Systematic Literature Reviews in Software Engineering. 2007.
- [S1] Madisetti, Vijay, and Arshdeep Bahga. "Cloud-Based Information Integration & Informatics Framework for Healthcare Applications." (2013): 1-1.
- [S2] Shen, Linshan, et al. "Framework of Distributed Medical Images Library for Medical Research and Education." Internet Computing for Science and Engineering (ICICSE), 2012 Sixth International Conference on. IEEE, 2012.
- [S3] Laohakangvalvit, Tipporn, and Tiranee Achalakul. "Cloud-based data exchange framework for healthcare services." Computer Science and Software Engineering (JCSSE), 2014 11th International Joint Conference on. IEEE, 2014.

- [S4] Sobhy, Dalia, Yasser El-Sonbaty, and M. Abou Elnasr. "MedCloud: healthcare cloud computing system." Internet Technology And Secured Transactions, 2012 International Conference for. IEEE, 2012.
- [S5] Yang, Cheng-Yi, and Chien-Tsai Liu. "Developing IHE-Based PHR Cloud Systems." Social Computing (SocialCom), 2013 International Conference on. IEEE, 2013.
- [S6] Jin, Yang, Tang Deyu, and Zheng Xianrong. "Research on the distributed electronic medical records storage model." IT in Medicine and Education (ITME), 2011 International Symposium on. Vol. 2. IEEE, 2011.
- [S7] Jalal-Karim, A., and W. Balachandran. "The optimal network model's performance for sharing Electronic Health Record." Multitopic Conference, 2008. INMIC 2008. IEEE International. IEEE, 2008.
- [S8] Bahga, Arshdeep, and Vijay Krishna Madiseti. "A Cloud-based Approach for Interoperable Electronic Health Records (EHRs)." Biomedical and Health Informatics, IEEE Journal of 17.5 (2013): 894-906.
- [S9] Patra, Debkumar, et al. "Achieving e-health care in a distributed EHR system." e-Health Networking, Applications and Services, 2009. Healthcom 2009. 11th International Conference on. IEEE, 2009.
- [S10] Jin, Yang, Tang Deyu, and Zhou Yi. "A Distributed Storage Model for EHR Based on HBase." Information Management, Innovation Management and Industrial Engineering (ICIII), 2011 International Conference on. Vol. 2. IEEE, 2011.
- [S11] Jin, Yang, Tang Deyu, and Zheng Xianrong. "Research on the distributed electronic medical records storage model." IT in Medicine and Education (ITME), 2011 International Symposium on. Vol. 2. IEEE, 2011.
- [S12] Lee, Ken Ka-Yin, Wai-Choi Tang, and Kup-Sze Choi. "Alternatives to relational database: comparison of NoSQL and XML approaches for clinical data storage." Computer methods and programs in biomedicine 110.1 (2013): 99-109.
- [S13] Yu, Weider D., et al. "Computer Engineering Department, San Jose State University, (Silicon Valley), California, USA, 95192-0180." e-Health Networking, Applications & Services (Healthcom), 2013 IEEE 15th International Conference on. IEEE, 2013.