



DINCON 2011

10^a Conferência Brasileira de Dinâmica, Controle e Aplicações

28 de agosto a 1^o de setembro de 2011



ADOÇÃO DE MEDIDAS DE QUALIDADE DE DADOS PARA O DESENVOLVIMENTO DE SISTEMAS BIOMÉDICOS

Wilson Jung^{1,2}, *Huei Diana Lee*^{1,2,3}, *Adrieli Cristina da Silva*¹, *Luiz Henrique Dutra da Costa*^{1,2}, *Bianca Espindola*³,
*Cláudio Saddy Rodrigues Coy*³, *João José Fagundes*³, *Feng Chung Wu*^{1,3}

¹Laboratório de Bioinformática — Universidade Estadual do Oeste do Paraná, Foz do Iguaçu, Brasil

²Programa de Pós-Graduação em Engenharia de Sistemas Dinâmicos e Energéticos
Universidade Estadual do Oeste do Paraná, Foz do Iguaçu, Brasil, {wilsonfoz, hueidianalee}@gmail.com

³Programa de Pós-Graduação em Ciências da Cirurgia — Faculdade de Ciências Médicas
Universidade Estadual de Campinas, Campinas, Brasil

Resumo: A Qualidade de Dados (QD) pode apresentar um impacto considerável sobre as áreas econômica, social e médica. O objetivo deste trabalho é aplicar princípios de QD ao desenvolvimento de sistemas biomédicos e propor um processo para monitoração do nível de QD, auxiliando na adoção de medidas de correção adequadas.

Palavras-chave: Aplicações em Medicina, Protocolo de Cirurgia Coloproctológica, Qualidade de Dados.

1. INTRODUÇÃO

Com a evolução da tecnologia, diversas áreas têm utilizado sistemas computacionais para auxiliar no desenvolvimento de suas atividades. Consequentemente, um grande volume de dados é gerado e parte deles é, em geral, usado como base para a tomada de decisões. Assim, problemas de Qualidade de Dados (QD) podem gerar perdas econômicas e sociais [1] e repercutem desde reveses mínimos até desastres, como os incidentes relacionados ao ônibus espacial *Challenger* em 1986 e ao *Iranian Airbus* em 1988 [2].

Os problemas de QD podem ser considerados, basicamente, em relação aos dados e ao usuário. Sob a perspectiva de dados, entende-se que a informação¹ deve estar de acordo com suas especificações e seus requisitos. Já sob a perspectiva de usuário, o foco está na adequação da informação às necessidades do usuário. Desse modo, o conceito de QD pode ser definido como a busca pela informação livre de defeitos e com as características desejadas pelos usuários [3].

Nesse contexto, os problemas de QD podem ser causados por diversos motivos, como problemas de qualidade de *software* e o processo de inserção de informações em sistemas computacionais, podendo causar grande impacto, especial-

mente sobre a área de saúde. Como exemplo, pode-se citar o caso que provocou danos à saúde e mortes por meio da aplicação de doses fatais de radiação em pacientes que realizaram tratamento com as máquinas Therac-25 no ano de 1986, devido a um problema de *software* do equipamento [4]. Outro caso relevante foi a *overdose* aplicada por uma unidade de radioterapia Cobalto-60, prejudicando cerca de 28 pessoas no Panamá, entre os anos de 2000 e 2001 [5].

Assim, a utilização de métodos eficientes de planejamento e construção de sistemas computacionais é importante não somente com o intuito de evitar a presença de erros, mas com o objetivo de auxiliar a prover a QD nesses sistemas.

Devido à grande quantidade de informação gerada por meio de procedimentos médicos, pode ser interessante aplicar processos de Mineração de Dados (MD) para auxiliar em uma posterior análise mais completa dos dados. Tendo em vista a importância de uma base de dados (BD) sólida e consistente para a realização desse processo, foram realizados estudos que apontaram os efeitos positivos da QD no processo de MD [6–8].

O objetivo deste trabalho é apresentar um estudo de caso, em andamento, da aplicação de conceitos relacionados à QD no desenvolvimento de um sistema biomédico, com a proposição do modelo de um processo para monitoração da QD, permitindo uma visão do nível de QD do sistema. O desenvolvimento desse sistema também visa a construção de uma BD robusta e adequada para aplicação de processos de MD.

Uma das maneiras de garantir a QD é assegurar a correlação entre os dados e as mudanças que ocorrem no mundo real [9]. Nesse contexto, a monitoração da QD, como proposta neste trabalho e pouco frequente na literatura relacionada a sistemas biomédicos, auxilia na identificação de problemas que podem comprometer a representação correta de uma situação do mundo real.

¹Neste trabalho os termos dado e informação são usados indistintamente.

Este trabalho faz parte do projeto de Análise Inteligente de Dados em uma parceria entre o Laboratório de Bioinformática (LABI) da Universidade Estadual do Oeste do Paraná (UNIOESTE)/Foz do Iguaçu, o Serviço de Coloproctologia da Faculdade de Ciências Médicas (FCM) da Universidade Estadual de Campinas (UNICAMP), o Laboratório de Inteligência Computacional da Universidade de São Paulo (USP)/São Carlos e o Grupo Interdisciplinar em Mineração de Dados e Aplicações da Universidade Federal do ABC.

O restante deste trabalho está organizado da seguinte maneira: na Seção 2 são abordadas questões relacionadas à QD; na Seção 3 é apresentado o protótipo proposto; na Seção 4 é apresentado o material e o método utilizado; na Seção 5 são discutidos o trabalho realizado e as perspectivas futuras do projeto e na Seção 6 é apresentada a conclusão.

2. AVALIAÇÃO DA QD

Para representar características específicas dos dados foram criadas diversas dimensões [1, 10, 11], que podem estar conectadas entre si e a um conjunto de métricas, que caracterizam diferentes problemas de QD [3]. Esta relação é apresentada na Figura 1.

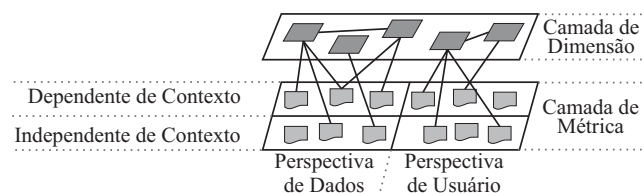


Figura 1 – Relação entre os componentes da avaliação da QD: dimensões, métricas, perspectiva e contexto (modificado de [3]).

Diversas dimensões são de natureza multivariável, sendo as variáveis dependentes ou não de contexto. Desse modo, deve-se determinar quais variáveis serão avaliadas, de acordo com o ambiente em que os dados se inserem [12]. Entre as dimensões de QD apresentadas na literatura, algumas das mais citadas incluem [6, 10, 12]:

Acurácia: permite avaliar a fidelidade da representação do mundo real através dos dados obtidos pelos sistemas de informação;

Completo: consiste na capacidade do sistema de informação em representar cada estado significativo de uma entidade do mundo real, para que atenda às necessidades do usuário;

Consistência: representa a presença de dados redundantes ou o nível de consistência entre dois elementos relacionados, como endereço e código postal;

Pontualidade: representa se os dados estão suficientemente atualizados para atender às necessidades dos usuários.

Como mencionado, a QD pode ser considerada sob duas perspectivas: dados e usuário. Sendo assim, os problemas de QD apresentados na literatura, considerando essas duas perspectivas e a dependência do contexto, podem ser classificados em quatro categorias [3]:

Perspectiva de Dados/Independente de Contexto: são problemas de qualidade presentes em BD e referem-se a qual-

quer conjunto de dados;

Perspectiva de Dados/Dependente de Contexto: são problemas de qualidade que violam as especificações de negócio e podem ser detectados por regras contextuais;

Perspectiva de Usuário/Independente de Contexto: são problemas de qualidade que ocorrem no processamento da informação;

Perspectiva de Usuário/Dependente de Contexto: são problemas de qualidade que indicam que os dados não estão adequados para as necessidades dos usuários.

Entre as causas mais comuns para surgimento de problemas de QD está o preenchimento manual no sistema de dados provenientes de, por exemplo, formulários e prontuários [13, 14]. Por este motivo, é importante que os usuários sejam conscientizados e entendam a responsabilidade de inserir corretamente os dados no sistema.

3. DESCRIÇÃO DO PROTÓTIPO PROPOSTO

Os conceitos abordados sobre QD estão sendo aplicados a um estudo de caso em desenvolvimento, em parceria com o Serviço de Coloproctologia da FCM/UNICAMP.

Atualmente, no Serviço de Coloproctologia da FCM/UNICAMP o armazenamento dos dados relacionados a procedimentos cirúrgicos é realizado através de um formulário eletrônico (sistema legado) criado por meio do *Microsoft Access*². Porém, esta solução não atende às necessidades de seus usuários e apresenta problemas de QD.

Para que um paciente seja submetido a um procedimento cirúrgico é necessário obter informações sobre diversos aspectos, como o histórico pessoal e familiar de enfermidades e procedimentos cirúrgicos, os hábitos sociais e o quadro clínico. Além dessas informações, os resultados dos exames pré-operatórios realizados, também devem ser armazenados visando reduzir o risco de complicações cirúrgicas que possam ocorrer.

Adicionalmente, as informações relacionadas aos pacientes podem servir como base para a aplicação de processos como a MD, para realizar uma análise mais completa e auxiliar na descoberta de padrões implícitos nos dados [7].

Porém, para oferecer uma BD adequada para o processo de MD são necessários mecanismos que permitam avaliar a QD do sistema, de modo que seja possível corrigir eventuais problemas nos dados para que não influenciem o processo. A monitoração da QD também proporciona mais confiabilidade ao sistema e oferece maior segurança ao usuário para tomar decisões com base nos dados.

Desse modo, nesse trabalho é proposto um sistema *web* para gerenciamento de Protocolo de Cirurgia Coloproctológica, que oferece um ambiente adequado, apoiado pelo processo de monitoração de QD, para armazenar informações relacionadas à enfermidade colorretal que o paciente apresenta e, ao procedimento cirúrgico ao qual foi submetido. Esse sistema permitirá também reunir informações sobre diversos aspectos do paciente que podem contribuir para futuros estudos.

Para garantir a segurança e a privacidade das informações

²<http://office.microsoft.com>

dos pacientes, é fundamental a utilização de métodos que bloqueiem o acesso aos dados por pessoas não autorizadas e que assegurem a privacidade dos dados transmitidos pela rede. Outra característica importante do sistema, consiste na organização da informação de acordo com os elementos clássicos da anamnese [15], facilitando a localização de um dado específico por um profissional da saúde.

Na Figura 2 é apresentada uma visão de alto nível do sistema, após o levantamento e a especificação dos requisitos.

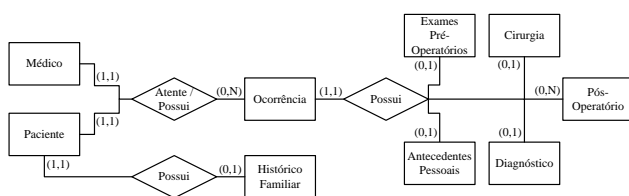


Figura 2 – Modelo entidade relacionamento do protótipo.

O preenchimento das informações no sistema é realizado por meio da transcrição dos dados presentes nos prontuários dos pacientes. No entanto, devido à rotina movimentada do ambiente hospitalar, a transcrição de informações nem sempre é realizada completamente. O preenchimento de dados incorretos também influencia negativamente na QD.

Sendo assim, como é apresentado na Figura 3, na primeira fase da implantação de medidas visando a QD validam-se os dados contemplados pelos campos presentes na interface gráfica, tal como a obrigatoriedade do preenchimento de campos específicos e a exigência de padrões para determinados campos, não permitindo, por exemplo, que campos numéricos recebam valores diferentes de números.

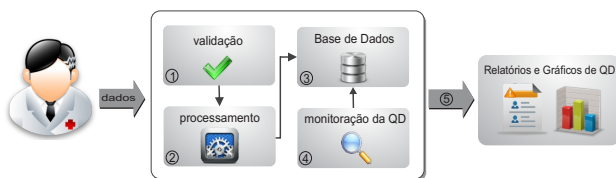


Figura 3 – Modelo do processo de monitoração da QD.

Após a validação e a persistência dos dados, através do processo de monitoração é possível avaliar os registros armazenados e emitir relatórios sobre os pacientes com dados irregulares, auxiliando ao profissional de saúde obter uma visão mais completa sobre os problemas existentes. Assim, a monitoração da QD auxilia no processo de revisão dos prontuários, indicando os problemas específicos de cada paciente, e aumenta a confiabilidade do sistema.

Outra importante funcionalidade desse módulo é a possibilidade de gerar gráficos que demonstram os índices de QD de acordo com as dimensões estabelecidas. Para avaliar a QD estão sendo adotadas as dimensões representadas por acurácia, completude, consistência e pontualidade, pois são amplamente discutidas na literatura e têm grande influência no processo de Mineração de Dados [8].

4. MATERIAL E MÉTODO

O levantamento dos requisitos para o desenvolvimento do sistema foi realizado junto aos especialistas do domínio e serviram, primeiramente, para a construção de um protótipo de telas, possibilitando uma visão prévia da organização das informações e auxiliando a identificar as funcionalidades que devem compor o sistema.

O desenvolvimento do sistema proposto utiliza a prototipagem como modelo de processo de *software* [16]. Nesse processo, a cada versão do protótipo, revisam-se os requisitos e as funcionalidades em conjunto com os especialistas do domínio, em ciclos sucessivos, até que o protótipo atenda a todos os requisitos.

O sistema está sendo desenvolvido como uma aplicação *web*, sendo necessário apenas um navegador *web* para utilizá-lo. A implementação está sendo realizada com o *framework Ruby on Rails*, que utiliza o modelo *Model View Controller* para o desenvolvimento de aplicações. Esse modelo oferece mecanismos auxiliares para a validação dos dados inseridos no sistema, além de permitir o desenvolvimento de rotinas de teste para avaliar as funcionalidades implementadas [17].

Entre outras características, o *framework Ruby on Rails* também oferece mecanismos para automatizar diversas tarefas no processo de desenvolvimento e manutenção de um sistema *web* [17]. Este *framework* permite a utilização da linguagem *JavaScript*, que pode contribuir significativamente no desenvolvimento de interfaces gráficas interativas e com efeitos visuais, auxiliando na validação de formulários e permitindo utilizar um fluxo de dados assíncrono [18].

Os dados serão armazenados e gerenciados pelo Sistema Gerenciador de Banco de Dados (SGBD) PostgreSQL³. Esse sistema, além de ser *opensource*, é amplamente utilizado e garante a integridade dos dados nele armazenados.

Para garantir a segurança dos dados transmitidos durante a utilização do sistema, estão sendo empregadas técnicas, como o *Secure Sockets Layer (SSL)*, a qual criptografa os dados para que não possam ser interpretados caso exista uma interceptação não autorizada [19].

5. DISCUSSÃO E PERSPECTIVAS

O modelo proposto neste trabalho está sendo desenvolvido com o acompanhamento constante dos especialistas do domínio, de maneira a entender e atender às necessidades dos usuários e ao processo envolvido na aquisição dos dados.

Também com o auxílio dos especialistas foi elaborado um protótipo de telas, que permitiu analisar previamente a organização da informação e as funcionalidades necessárias. A criação de interfaces mais interativas e elaboradas pode auxiliar no preenchimento correto das informações, contribuindo também com a QD [13].

Após o desenvolvimento do protótipo, os dados existentes no sistema legado serão migrados para o novo sistema. Uma análise da estrutura dos dados do sistema legado permitiu identificar eventuais falhas na modelagem e no relacionamento entre as entidades. Atualmente estão sendo analisadas as dimensões e as métricas que irão compor a avaliação da

³<http://www.postgresql.org/>

QD do sistema legado. Também será necessário adotar medidas de correção para adequar os dados ao novo sistema.

Até onde se conhece o problema, os sistemas na área biomédica apresentam, quando existente, apenas avaliações simples sobre a validade dos dados e não apresentam ferramentas para auxílio à monitoração da QD, como o proposto neste trabalho, por meio da utilização de medidas objetivas [12].

Além da preocupação com a QD sob a perspectiva dos dados, é importante avaliar se o sistema proposto fornece informações adequadas às necessidades do usuário. Por esse motivo, a avaliação da QD também será realizada de maneira subjetiva [12], por meio de formulários adaptados a partir de modelos existentes, que serão direcionados aos usuários do sistema na área biomédica.

6. CONCLUSÃO

A introdução de conceitos de QD no desenvolvimento de sistemas, especificamente neste trabalho em relação à sistemas biomédicos, visa estabelecer medidas preventivas para evitar problemas de QD, tendo em vista que há uma lacuna em relação a pesquisas que abordam este tema [20]. Medidas de correção para problemas de QD são necessárias, mas implicam em grandes custos. Assim sendo, o ideal é atingir um equilíbrio entre medidas preventivas e corretivas.

O novo sistema, apresentado neste trabalho, servirá como uma ferramenta de apoio mais adequada ao gerenciamento de informações sobre procedimentos cirúrgicos realizados no Serviço de Coloproctologia da FCM/UNICAMP, fornecendo mecanismos que garantam a privacidade dos dados dos pacientes, como controle de acesso, criptografia dos dados que trafegam pela rede e uma BD com elevado número de atributos, que podem contribuir para pesquisas futuras nesta área.

Por fim, a monitoração da QD, inexistente na grande maioria dos sistemas, permite adotar medidas de correção à medida em que se identificam problemas. Uma BD com qualidade, além de possibilitar tomada de decisão mais segura com base nos dados, pode contribuir para o processo de MD, otimizando a etapa de pré-processamento dos dados, uma das mais custosas do processo [7].

AGRADECIMENTOS

À Fundação Parque Tecnológico Itaipu e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico pelo apoio por meio da linha de financiamento de bolsas.

REFERÊNCIAS

[1] R. Y. Wang and D. M. Strong, “Beyond accuracy: What data quality means to data consumers,” *J. Manage. Inf. Syst.*, Vol. 12, No. 4, pp. 5–33, 1996.

[2] C. W. Fisher and B. R. Kingma, “Criticality of data quality as exemplified in two disasters,” *Information & Management*, Vol. 39, No. 2, pp. 109–116, 2001.

[3] M. Ge and M. Helfert, “A Review of Information Quality Research,” in *Proceedings of the ICIQ*. USA: MIT, 2007, pp. 76–91.

[4] N. G. Leveson and C. S. Turner, “An investigation of the Therac-25 accidents,” *Computer*, Vol. 26, No. 7, pp. 18–41, 1993.

[5] C. Borrás, “Overexposure of radiation therapy patients in Panama: problem recognition and follow-up measures,” *Rev. Panam. Salud Pública*, Vol. 20, No. 2-3, pp. 173–187, 2006.

[6] C. Batini and M. Sannapieco, *Data quality: Concepts, methodologies and techniques*. USA: Springer-Verlag, 2006.

[7] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, 2nd ed. USA: Elsevier, 2006.

[8] R. Blake and P. Mangiameli, “The effects and interactions of data quality and problem complexity on classification,” *J. Data and Information Quality*, Vol. 2, pp. 8:1–8:28, 2011.

[9] K. Orr, “Data quality and systems theory,” *Commun. ACM*, Vol. 41, No. 2, pp. 66–71, 1998.

[10] Y. Wand and R. Y. Wang, “Anchoring data quality dimensions in ontological foundations,” *Commun. ACM*, Vol. 39, No. 11, pp. 86–95, 1996.

[11] T. C. Redman, *Data quality for the information age*. USA: Artech House, 1996.

[12] Y. W. Lee, L. L. Pipino, J. D. Funk, and R. Y. Wang, *Journey to Data Quality*. USA: MIT Press, 2006.

[13] A. Maydanchik, *Data Quality Assessment*. USA: Technics Publications, LLC, 2007.

[14] W. Zalewski, H. D. Lee, A. C. Lorena, M. J. F. A. Caetano, A. G. Maletzke, J. J. Fagundes, C. S. R. Coy, and F. C. Wu, “Evaluation of models for the recognition of handwritten digits in medical forms,” in *Proceedings of the Advances in Bioinformatics and Computational Biology*, Vol. 5167. Germany: Springer, 2008, pp. 178–181.

[15] C. C. Porto, *Exame Clínico*. Brasil: Guanabara Koogan, 1982.

[16] S. L. Pfleger, *Engenharia de Software: Teoria e Prática*, 2nd ed. Brasil: Prentice Hall, 2004.

[17] S. Ruby, D. Thomas, and D. H. Hansson, *Agile Web Development with Rails*, 4th ed. USA: Pragmatic Bookshelf, 2010.

[18] D. Goodman and M. Morrison, *JavaScript Bible*, 6th ed. USA: Wiley Publishing, 2007.

[19] R. Oppliger, *SSL and TLS: theory and practice*. USA: Artech House, 2009.

[20] C. Guerra-García, I. Caballero, and M. Piattini, “A systematic literature review of how to introduce data quality requirements into a software product development,” in *Proceedings of the ENASE*, Greece, 2010, pp. 12–19.