

EnDia 2.0: Sistema Web para Construção de Bases de Dados a partir de Exames de Endoscopia Digestiva Alta

Newton Spolaôr¹, Everton Alvares Cherman¹, Neimar Neitzel¹,
Luiz Henrique Dutra da Costa¹, Carlos Andres Ferrero¹, André Gustavo Maletzke¹,
Daniel de Faveri Honorato^{1,2}, Renato Bobsin Machado^{1,2}, Huei Diana Lee¹,
João José Fagundes³, Cláudio Saddy Rodrigues Coy³, Feng Chung Wu^{1,2,3}

¹Laboratório de Bioinformática — LABI,
Universidade Estadual do Oeste do Paraná — UNIOESTE, Brasil

²Fundação Parque Tecnológico Itaipu — FPTI, Brasil

³Serviço de Coloproctologia da Faculdade de Ciências Médicas — FCM,
Universidade Estadual de Campinas — UNICAMP, Brasil

Resumo - O avanço tecnológico atual possibilita o processamento e o registro de um volume crescente de dados. Esse fenômeno é observado em áreas como a medicina, na qual hospitais e clínicas armazenam uma quantidade expressiva de exames de pacientes, freqüentemente, em documentos impressos. A análise manual desses grandes conjuntos de dados pode ser complexa devido ao alto custo de tempo envolvido, de modo que processos de aquisição automática e semi-automática de conhecimento, como a descoberta de conhecimento em bases de dados, são propostos para auxiliar nessa tarefa. Além desses fatos mencionados, o avanço tecnológico permite ainda o crescimento da capacidade de processamento de sistemas *web*, os quais estão se tornando cada vez mais populares. O objetivo deste trabalho é apresentar o EnDia 2.0, um sistema *web* em desenvolvimento que permite o gerenciamento de informações referentes ao exame de endoscopia digestiva alta. O sistema foi planejado com o intuito de, posteriormente, aplicar-se processos computacionais para a aquisição de conhecimentos que auxiliem em processos de tomada de decisão no domínio considerado.

Palavras-chave: Descoberta de Conhecimento em Bases de Dados, Gastroenterologia, Modelo Cliente-Servidor.

Abstract - Technological advancement allows processing and registering of a growing amount of data. This phenomenon is observed in areas such as medicine, in which hospitals and clinics store an expressive quantity of patients' exams, often in printed documents. Manual analysis of these large data sets may be complex due to the high consumption of time. Automatic and semi-automatic knowledge acquisition process, as Knowledge Discovery from Databases, can provide support for this task. Technological advancement has also influenced on the processing capacity of web systems, which are becoming very popular. The objective of this work is to present EnDia 2.0, a web system under development planned to give support to the management of upper digestive endoscopy exams. Later, computational process for knowledge acquisition, which help specialists on the decision making process, will be applied on the information registered in the structured database by the system.

Key-words: Knowledge Discovery from Databases, Gastroenterology, Client-Server Model.

Introdução

Com o avanço tecnológico, cada vez mais hospitais e clínicas médicas estão armazenando e recuperando informações relacionadas a pacientes e a processos laboratoriais. As informações armazenadas podem ser analisadas de modo que padrões sejam extraídos para auxiliar especialistas no processo de tomada de decisão. Essa análise pode ser auxiliada por processos computacionais, como o de Descoberta de Conhecimento em Base de Dados – DCBD [1]. Esse processo tem como obje-

tivo a identificação de informações úteis armazenadas implicitamente nas Bases de Dados – BD e é composto por três etapas: (1) Pré-processamento, (2) Extração de padrões e (3) Pós-processamento. A etapa de pré-processamento tem como objetivo preparar os dados para a próxima etapa, por meio de tarefas como limpeza, transformação e redução do conjunto de dados. A etapa de extração de padrões tem como objetivo a construção de modelos que auxiliem na identificação de padrões encontrados nos dados. Essa tarefa pode ser realizada por meio da utilização de métodos de Inteligência Artificial, especificamente

da área de Aprendizado de Máquina. Os padrões identificados devem ser analisados e validados na fase de pós-processamento, por meio de medidas estatísticas e também utilizando o conhecimento de especialistas do domínio. Nesse sentido, foi proposto em outro trabalho o sistema EnDia 1.0, o qual tem como objetivos o gerenciamento e a construção de BD estruturadas sobre informações de Endoscopia Digestiva Alta – EDA para a aplicação do processo de DCBD [2].

Nesse trabalho, em andamento, é apresentado o sistema *web* EnDia 2.0, o qual é parte de uma metodologia para construção de BD médicas, especificamente, de exames de EDA. Esse sistema faz parte do projeto de Análise Inteligente de Dados, o qual é desenvolvido em uma parceria entre o Laboratório de Bioinformática/LABI da Universidade Estadual do Oeste do Paraná/UNIOESTE – Foz do Iguaçu, o Serviço de Endoscopia Digestiva Alta do Hospital Municipal de Paulínia/HMP e o Serviço de Coloproctologia da Faculdade de Ciências Médicas da Universidade Estadual de Campinas/UNICAMP [3, 4, 5, 6, 7].

O EnDia 2.0 é um sistema em desenvolvimento que possui como objetivo o gerenciamento e armazenamento de maneira consistente, a partir de uma interface *web*, das informações de exames de EDA em um formato adequado (atributo-valor) para que possam ser utilizadas no processo de DCBD. Para o desenvolvimento desse sistema, foram necessários um estudo do domínio da aplicação e a interação com especialistas do domínio. Um aspecto que deve ser ressaltado é que a EDA é um exame importante para o diagnóstico de doenças gastroduodenais, como úlceras e gastrites, as quais representam uma das entidades em gastroenterologia de maior incidência na população mundial [8, 9].

Este trabalho está organizado da seguinte maneira: na segunda seção é apresentado o processo de desenvolvimento do sistema, bem como as tecnologias utilizadas; na terceira seção são discutidos os benefícios da aplicação dessas tecnologias e do processo no desenvolvimento do EnDia 2.0; e no final são apresentadas as conclusões do trabalho.

Materiais e Métodos

O exame de EDA contempla informações sobre o esôfago, o estômago e o duodeno. De modo complementar, também são consideradas as informações sobre o exame de Anatomia Patológica e o teste da urease, as quais contribuem para o diagnóstico da bactéria *Helicobacter Pylori*, presente em doenças pépticas.

O gerenciamento dessas informações pode ser realizado pelo sistema EnDia 2.0, o qual é imple-

mentado a partir do processo de desenvolvimento *Rational Unified Process* – RUP adaptado às características da equipe. Desse modo, a elaboração do EnDia 2.0 é constituída de quatro fases:

Fase 1. Análise de requisitos e prototipação;

Fase 2. Modelagem do sistema;

Fase 3. Implementação do sistema;

Fase 4. Validação e verificação.

Na Fase 1 é realizado o levantamento dos requisitos do sistema por meio do estudo de conceitos do domínio na literatura e de reuniões com especialistas. Nesta fase também é utilizada a técnica de prototipação, a qual corresponde a um processo iterativo e interativo utilizado para auxiliar na identificação de requisitos adicionais e na validação dos requisitos já identificados. O principal produto elaborado nesta fase é o documento de visão, no qual são registradas informações como o objetivo do sistema e os requisitos funcionais e não funcionais [10]. Após a definição dos requisitos e a validação do protótipo, junto aos especialistas do domínio, inicia-se a Fase 2, na qual é realizada a modelagem da solução computacional. A descrição dos diagramas e documentos nesta fase é realizada utilizando a linguagem *Unified Modeling Language* – UML [11]. Desse modo, alguns documentos e diagramas podem ser elaborados para auxiliar na modelagem do EnDia 2.0:

- Diagrama de casos de uso;
- Detalhamento dos casos de uso;
- Diagrama de classes de negócio;
- Modelo Entidade-Relacionamento;
- Diagrama de classes do sistema;
- Descrição de classes do sistema;
- Diagramas de seqüência.

Além desses documentos também são definidas, na Fase 2, as três camadas que compõem a arquitetura de desenvolvimento implementada na próxima fase: apresentação, negócio e persistência. Essas camadas são construídas de acordo com o modelo cliente-servidor [12], de modo que as camadas de persistência e negócio são alocadas no computador servidor, enquanto que a camada de apresentação é carregada nos computadores clientes, geralmente a partir de um navegador *web*. A plataforma Java *Enterprise Edition* – EE [13], associada com o servidor de aplicação *Sun Java System Application Server*

9.1¹ são utilizados na Fase 3 para dar suporte a aspectos relacionados à infra-estrutura de um sistema *web* nesse modelo. A camada de apresentação tem como objetivos disponibilizar a interação do usuário com o sistema e permitir a comunicação com a camada de negócio. Para a implementação dessa camada são utilizadas a linguagem *Java Server Pages – JSP*² e o ambiente de desenvolvimento de interfaces gráficas *Rich Web*³ OpenLaszlo 4.1.1⁴. A camada de negócio utiliza a tecnologia *Enterprise Java Beans – EJB* 3.0⁵, a qual possibilita, dentre outras tarefas, o compartilhamento dos serviços providos por essa tecnologia para distintas camadas de apresentação. A persistência dos dados na última camada é realizada pelo Sistema Gerenciador de Banco de Dados – SGBD MySQL 5.01⁶.

Os testes de validação e verificação do sistema EnDia 2.0 são aplicados na Fase 4. A validação é realizada com auxílio de especialistas e tem como objetivo avaliar a adequabilidade do sistema para a utilização no domínio e para o atendimento dos requisitos. Os testes de verificação correspondem à análise da consistência dos documentos e diagramas gerados nas fases anteriores do desenvolvimento.

Resultados e Discussão

O EnDia 2.0 permite, a partir de uma interface *web*, o armazenamento de informações relacionadas ao exame de EDA em BD no formato atributo-valor, o qual é adequado à aplicação de processos computacionais, como o DCBD.

O desenvolvimento do EnDia 2.0 está sendo realizado de acordo com as quatro fases apresentadas anteriormente. As tarefas propostas nas Fases 1 e 2 estão concluídas, ao passo que a Fase 3 está sendo finalizada. A partir da consolidação da Fase 3 podem ser aplicados os testes propostos na Fase 4 com o auxílio de especialistas do domínio.

As tarefas realizadas na Fase 1 permitiram adquirir os requisitos do sistema, os quais foram descritos em uma formalidade acessível aos desenvolvedores e aos usuários no documento de visão.

Durante essa fase foi identificada uma quantidade extensa de atributos correspondentes às informações de EDA. Para facilitar o acesso a esses atributos pelos usuários, as interfaces gráficas do EnDia 2.0 estão sendo elaboradas de modo a agrupar as informações conforme a afinidade que apresentam entre si. Desse

modo, foram definidas interfaces específicas para o gerenciamento de exames de EDA e de Anatomia Patológica.

Algumas interfaces gráficas correspondentes ao exame de EDA são constituídas por abas, o que facilita a organização de subconjuntos de informações, verificadas no exame, relacionadas aos órgãos do esôfago, do estômago e do duodeno. Outros atributos presentes nessas telas são a história clínica e informações correspondentes ao paciente e ao médico. Em relação a cada órgão analisado no exame de EDA, é realizada outra subdivisão de abas de acordo com a região considerada.

A tela que apresenta as informações correspondentes ao estômago, por exemplo, é constituída por abas específicas para as regiões do fundo, corpo e antro gástrico, nas quais estão dispostos os respectivos atributos considerados. A partir dessas telas, opções como o cadastro e a consulta de exames de EDA e de Anatomia Patológica também podem ser utilizadas pelo usuário. Um protótipo da tela correspondente às informações do corpo gástrico é apresentado na Figura 1.

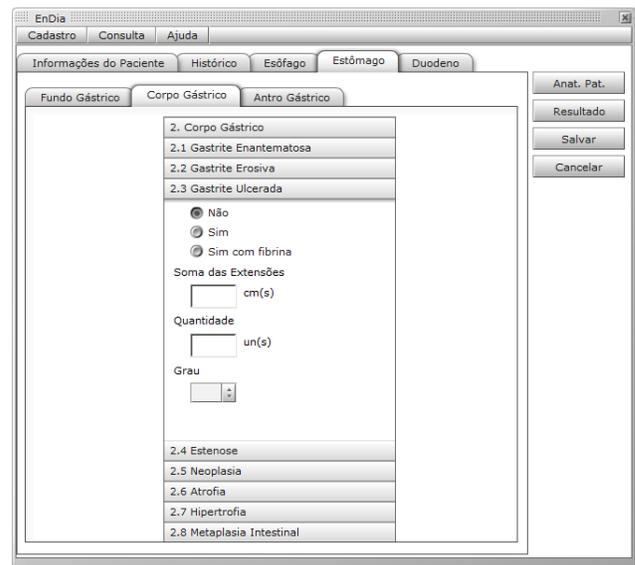


Figura 1: Protótipo de tela do EnDia 2.0

Na Fase 2, a linguagem UML foi utilizada para a construção de diagramas que auxiliam os desenvolvedores e os usuários a compreender a composição do sistema. O diagrama de casos de uso ilustrado na Figura 2 possibilita a representação abstraída dos

¹<http://developers.sun.com/appserver/>

²<https://java.sun.com/products/jsp>

³<http://www.theserverside.com/tt/articles/article.tss?l=RIA>

⁴<http://www.openlaszlo.org/>

⁵<http://java.sun.com/products/ejb/>

⁶<http://www.mysql.com/>

conceitos de EDA considerados no EnDia 2.0. A partir desses conceitos são construídos diagramas e documentações que provêm suporte à definição da BD e de outros componentes do sistema.

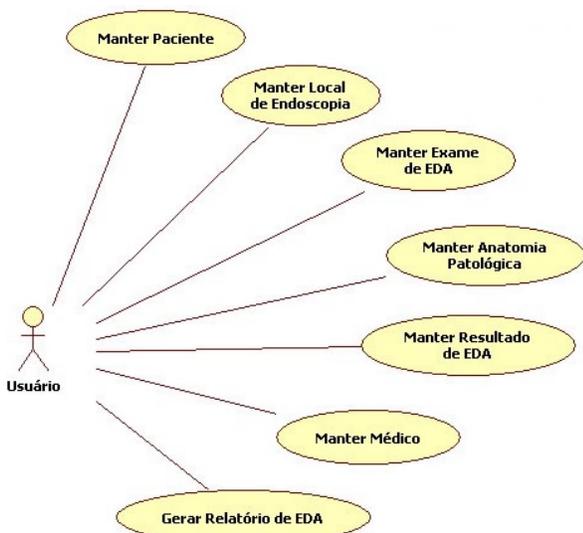


Figura 2: Diagrama de casos de uso do EnDia 2.0

Nesta figura é possível observar que os conceitos de EDA considerados mais importantes pelos especialistas e desenvolvedores correspondem ao exame de EDA, às informações relacionadas ao paciente e ao médico e o exame de Anatomia Patológica, dentre outros. A definição dos casos de uso é importante ainda para possibilitar a estimativa do tamanho do sistema, o que auxilia no planejamento de cronogramas e na distribuição de tarefas entre os desenvolvedores.

O EnDia 1.0 apresentava como objetivo principal o armazenamento de dados em um formato estruturado, o qual é apto à aplicação de processos computacionais como o DCBD. O sistema foi implementado para facilitar o gerenciamento das informações relacionadas ao exame de EDA, anatomia patológica e o teste da urease. Essas informações podem dar suporte, posteriormente, à extração de conhecimentos que auxiliem, por exemplo, em processos de tomada de decisão.

O avanço tecnológico na capacidade de processamento de grandes conjuntos de dados por redes de computadores, como a Internet, vem motivando o desenvolvimento de plataformas *web* para diversos sistemas. O EnDia 2.0 está sendo desenvolvido de acordo com essa tendência, a partir de algumas das técnicas e ferramentas mais aplicadas nesse sentido. O acesso remoto que o sistema pode disponibilizar via Internet e intranet é útil, por exemplo, para redes hospitalares com unidades geograficamente distantes e para uma clínica médica com diversos computadores em um mesmo recinto, respectivamente.

As práticas propostas no modelo cliente-servidor contribuem para a implementação do EnDia 2.0, visto que permitem agregar ao sistema aspectos como a centralização dos dados e a portabilidade em relação à plataforma.

A centralização dos dados em um servidor pode contribuir no controle dos investimentos necessários para a integridade dos dados, a manutenção de *hardwares* e a atualização de *softwares*. A maior parte dos recursos destinados a esses objetivos, em um sistema de saúde, poderiam ser concentrados em um servidor, enquanto que os demais computadores do sistema atuariam como terminais, utilizados apenas para a coleta e o envio das informações para esse servidor. Desse modo, é possível utilizar computadores clientes de baixo custo e, conseqüentemente, otimizar o uso de recursos econômicos e financeiros.

O gerenciamento centralizado dos dados disponibilizada pelo EnDia 2.0 auxilia também na integração das informações registradas nas clínicas afiliadas a um sistema de saúde, assim como nos computadores que compõem uma mesma clínica, em uma BD compartilhada. Um aspecto importante nessa integração corresponde a adoção de políticas de privacidade, as quais podem ser aplicadas, se desejadas, por um sistema de saúde. A implementação de tais políticas pode ser realizada pela restrição de acesso, para determinadas clínicas afiliadas, a alguns conjuntos de dados presentes na BD compartilhada.

A portabilidade, de acordo com a plataforma, nos computadores clientes e servidores é possível devido à natureza *web* do EnDia 2.0. Essa portabilidade permite que diversos sistemas operacionais possam ser utilizados em um mesmo ambiente ou sistema de saúde, o que pode proporcionar uma redução de custos se aplicações *freeware* forem consideradas. A condição necessária nesse sentido é o suporte a um navegador *web*, por parte do sistema operacional utilizado, que atenda aos requisitos das interfaces gráficas presentes no EnDia 2.0.

As interfaces gráficas do EnDia 2.0 apresentam estruturas não disponíveis em páginas *web* no padrão *Hyper Text Markup Language* – HTML, como interações amigáveis ao usuário, o que permite agregar ao sistema aspectos comuns a *Rich Web*. Essas interações apresentam animações intuitivas, incluindo estruturas como botões e menus, que são semelhantes a determinados aplicativos *desktop*.

O modelo cliente-servidor e o processo de desenvolvimento utilizado no EnDia 2.0 são estruturados de modo a permitir o desenvolvimento de outras interfaces além da *web*. A plataforma Java EE, associada ao EJB 3.0, possibilitam o desenvolvimento de interfaces do EnDia 2.0 específicas para equipamentos como *Personal Digital Assistants* – PDA e celulares.

Conclusão

Neste trabalho, em andamento, foi apresentado o sistema *web* EnDia 2.0, bem como o seu processo de desenvolvimento. A plataforma *web* inerente ao sistema possibilita às redes hospitalares e às clínicas médicas um gerenciamento centralizado dos dados correspondentes aos exames de EDA e outros exames associados.

Como trabalho futuro se propõe a conclusão das Fases 3 e 4 do sistema, de modo a permitir que o aplicativo EnDia 2.0 possa ser consolidado e, posteriormente, utilizado para a construção de BD, a partir de exames de EDA, adequadas à extração de conhecimentos. Esses conhecimentos podem auxiliar em processos de tomada de decisão associados a esse domínio.

Agradecimentos

Ao Programa de Desenvolvimento Tecnológico Avançado – PDTA/FPTI-BR pelo auxílio financeiro.

Referências

- [1] Fayyad UM, Platestsky-Shapiro G, Smyth P. From data mining to knowledge discovery: an overview. In: American Association for Artificial Intelligence; 1996. p. 1–30.
- [2] Ferrero CA, Maletzke AG, Lee HD, Wu FC, Neto AP, Fagundes JJ, et al. EnDia – Uma Ferramenta para Construção de Repositórios de Dados de Endoscopia Digestiva Alta. In: Anais do X Congresso Brasileiro de Informática na Saúde. Florianópolis, SC – Brasil; 2006. p. 1–6.
- [3] Honorato DDF, Cherman EA, Lee HD, Monard MC, Wu F. Construção de uma Representação Atributo-valor para Extração de Conhecimento a partir de Informações Semi-estruturadas de Laudos Médicos. In: Anais do Conferencia Latinoamericana de Informática - CLEI. San José - Costa Rica; 2007. p. 1–12.
- [4] Honorato DDF, Lee HD, Wu FC, Machado RB, Neto AP, Fagundes JJ, et al. *H.pylori*-MINDSys: Um Protótipo de Sistema Baseado em Conhecimento para Auxílio na Predição da Bactéria *Helicobacter pylori* em Doenças Pépticas. In: Anais do V Workshop de Informática Médica. Porto Alegre, RS; 2005. p. 1–1.
- [5] Lee HD, Monard MC, Wu FC. Seleção de Atributos Relevantes e Não Redundantes usando a Dimensão Fractal do Conjunto de Dados. In: Anais do V Encontro Nacional de Inteligência Artificial, XXV Congresso da Sociedade Brasileira

de Computação. Porto Alegre, RS; 2005. p. 444–453.

- [6] Ferro M, Lee HD, Esteves SC. Intelligent Data Analysis: A Case Study of the Diagnostic Sperm Processing. In: Proc. of the International Conference on Computer Science, Software Engineering, Information Technology, e-Business and Applications. Foz do Iguaçu, PR; 2002. p. 116–120.
- [7] Esteves SC, Lee HD, Monard MC. Inteligência Artificial Aplicada à Andrologia: Um Estudo de Caso do Processamento de Sêmen Diagnóstico. In: Anais do XXVII Congresso Brasileiro de Urologia. Fortaleza, CE; 2001. p. 1–1.
- [8] Pellicano R, Fagoonee S, Palestro G, Rizzetto M, Figura N, Ponzetto A. The Diagnosis of *Helicobacter Pylori* Infection: Guidelines from the Maastricht 2-2000 Consensus Report. Minerva Gastroenterol Dietol. 2004;vol. 50(2):125–33.
- [9] Cordeiro F. Endoscopia Digestiva. Editora Média e Científica Ltda.; 1994.
- [10] Pressman RS. Engenharia de Software. Rio de Janeiro, RJ: McGrawHill; 2002.
- [11] Booch G, Rumbaugh J, Jacobson I. The Unified Modeling Language User Guide. Indianapolis, IN: Addison-Wesley; 1998.
- [12] Tanenbaum AS. Redes de Computadores. 3rd ed. Campus; 1997.
- [13] Alur D, Crupi J, Malks D. Core J2EE Patterns: As melhores práticas e estratégias de design. 2nd ed. Elsevier; 2004.

Contato

N. Spolaôr — newtonspolaor@gmail.com
E. A. Cherman — everttoncherman@gmail.com
N. Neitzel — mar.neitzel@gmail.com
L. H. D. Costa — lhdc90@gmail.com
C. A. Ferrero — anfer86@gmail.com
A. G. Maletzke — andregustavom@gmail.com
D. F. Honorato — dfaverih@gmail.com
R. B. Machado — rmachado@itaipu.gov.br
H. D. Lee — hueidianalee@gmail.com
J. J. Fagundes — jfagundes@mpcnet.com.br
C. S. R. Coy — ccoy@terra.com.br
F. C. Wu — wufc@pti.org.br
Laboratório de Bioinformática — LABI, Universidade Estadual do Oeste do Paraná — UNIOESTE, Parque Tecnológico Itaipu — PTI, Av. Tancredo Neves, 6731, CEP 85866-900, Foz do Iguaçu — PR.