

Ontologia na representação de metamodelos para data mining: estudo de caso Hospital da Cruz Vermelha

Carlos Kuretzki¹, Diovane Rinaldin¹, Silvio Bortoleto¹, Thiago T. Angélico¹

¹Centro Universitário Positivo – UNICENP
Rua Prof. Pedro Viriato P. de Souza, 5.300 – 81.280-330 – Curitiba – PR – Brazil

Carlos@zenithti.com, diovane@nilko.com.br, silvio.bortoleto@uol.com.br
, angelico100@gmail.com

Abstract. *The data mining looks for the discovery of standards, to assist the diverse branches. In this work we give emphasis to the medicine, but so that this happens it is necessary to make a mapping of all the steps that the patient covers, leaving of the handbook until the comparisons between the treatments of one same illness. In this article ontology model is created to represent one metamodel capable to visible the submodels that compose it until arriving at the information. Metamodel is divided in submodels that represent distinct areas, but keeps the Inter-relationships, where it is prevailed by a knowledge model, that can answer questionings set in motion distinct algorithms.*

Resumo. *A mineração de dados procura na descoberta de padrões, auxiliar os diversos ramos. Neste trabalho damos ênfase à medicina, mas para que isto aconteça é necessário fazer um mapeamento de todos os passos que o paciente percorre, partindo do prontuário até as comparações entre os tratamentos de uma mesma doença. Neste artigo um modelo ontológico é criado para representar um metamodelo capaz de enxergar os submodelos que o compõe até chegar à informação. O metamodelo é dividido em submodelos que representam áreas distintas, mas mantém os inter-relacionamentos, onde é regido por um modelo de conhecimento, que pode responder questionamentos acionado algoritmos distintos.*

1. Introdução

O campo de pesquisa está na descoberta de padrões, como a utilização de redes bayesianas para diagnósticos de doenças pulmonares (HRUSCHKA JR., E. R 1997). visando auxiliar profissionais da área.

Esta tarefa não se resume apenas na aplicação de algoritmos sobre a massa de dados. Deve descobrir conhecimento que denote um processo de uma representação equivalente na procura de um valor válido, que se encaixe nos padrões estabelecidos, não triviais (descoberta real) e que encontrem relação ao modelo descrito (Sarker, Abbas e Newton 2002), ou seja, é preciso modelos bem definidos e que traduzam a realidade

através da informação disponível, portanto esses modelos devem ser tratados segundo as regras que regem o negócio.

A pesquisa foi executada com a utilização da base de dados do Hospital da Cruz Vermelha da cidade de Curitiba-PR. Foi dividido o modelo de dados em três grandes áreas dos protocolos hospitalares: atendimento paciente, prescrição médica e solução. Cada área possui conhecimentos e regras distintas, existindo a necessidade uma definição global de metamodelo (*modelo principal*). Os sub-modelos estão vinculados a este metamodelo, criando uma rede de conhecimento e potencialidade na aplicação da informação.

O metamodelo é uma definição que abrangente os dados que virão após eles, os sub-modelos possuem os menores grânulos onde estão os dados de cada unidade de negócio, sendo assim entende-se que o metamodelo irá ser um agente de relacionamento entre os sub-modelos que tem ciência de cada tipo de informação em seus sub-modelos e que assim pode relacioná-los (Felipe Nery R. Machado 2004).

As possibilidades são diversas. O objetivo é de encontrar informações que não estejam explícitas e que o sistema atual não é capaz de oferecer. O foco está nos processos da empresa e nas pessoas envolvidas nestes processos, que são os estruturadores de informação (Fayyad, Piatetsky-Shapiro, Smyth, Uthurusamy 1996) e quem pode definir quais dados e conjuntos as informações são válidos.

As tarefas foram de organizar as informações que estavam disponíveis e remodelar partes do sistema, adotando um que pudesse ser trabalhado. Demonstrações do problema foram extensivamente realizadas para que se pudesse extrair o máximo possível de conhecimento e que pudesse disseminar informação ao maior número de pessoas.

O sistema que alimenta atualmente o banco de dados é o Tasy. Surgiu a partir da idéia de quatro médicos em criar um sistema capaz de gerenciar a tarefa hospitalar em seus diversos níveis como: farmácia, atendimento clínico, internamentos e outros. Por ser altamente customizável, está implantado em diversos hospitais do Brasil, em Curitiba além do hospital Cruz Vermelha, o hospital Nsa. das Graças e o Santa Cruz são hospitais que fazem uso do Tasy para gerenciar suas tarefas diárias.

Cerca de 60% das informações imputadas no sistema não são tratadas e inseridas devidamente, gerando uma séria inconsistência no banco de dados. Este fato dificulta a pesquisa, uma vez que se necessita de padrões para elaborar os modelos ontológicos. Segundo ALVARENGA (2003), um modelos ontológicos “é o responsável pelo mapeamento das informações em um nível superior, mais próximo ao entendimento de todos os envolvidos, além da sua validação conforme o entendimento de cada. Vale ressaltar que o modelo ontológico não contém dado algum e sim o significado e relação entre cada dado a fim de representar a realidade de uma informação”.

2. Metodologia

Precisa-se encontrar um padrão para representar o conhecimento dentro dos sub-modelos normalizados. Para isto, foi utilizada a linguagem OWL (*Web Ontology Language*), que já possui uma normalização de regras. Também foi utilizada a ferramenta Protege. Ela é capaz de representar a ontologia do negocio, usando as regras

da OWL. Isto é feito através de classes, propriedades e demais características, que tornam o trabalho menos custoso, além de ser uma ferramenta gratuita.

Sua importância se deu ao fato de que os demais profissionais envolvidos no trabalho, como médicos, enfermeiras, entre outros, precisarem compreender o que está sendo trabalhado, a estrutura gerada e o diagrama da ontologia.

Foi também utilizada a ferramenta ErWin, onde foi executada a reengenharia do banco de dados (apenas tabelas). Após esta etapa, foram divididas as áreas que serão mostradas mais a frente (sub-modelos). Por fim, a ferramenta DbDesigner foi utilizada para criar os sub-modelos com as tabelas e atributos efetivamente úteis para nós. A partir do conceito dos dados pode-se construir o modelo e seus sub-modelos, seguindo uma lógica de relacionamento entre eles (Silberschatz, A., Korth, H. F. and Sudarshan, S. 2005).

3. Ontologia do negócio

Para se chegar a esse nível de informação foi construído um sub-modelo para cada área do hospital. Cada um desses sub-modelos refere-se a uma área de negócio que é movimentada durante os estados em que o paciente pode se encontrar, desde a sua entrada no hospital até a sua saída. Então se definiu o significado de cada atributo dos sub-modelos em um modelo ontológico.

Estudando em paralelo o modelo e o sistema de Informação utilizado como *front-end* pelos colaboradores (Tasy), identificou-se quais dados seriam realmente relevantes para a construção de sub-modelos para a mineração de dados no prontuário hospitalar. Desse ponto em diante, já há uma estrutura de funcionamento do sistema que irá direcionar as buscas dos caminhos percorridos pelas informações ocultas dentro da base, conforme representado pela figura 1.

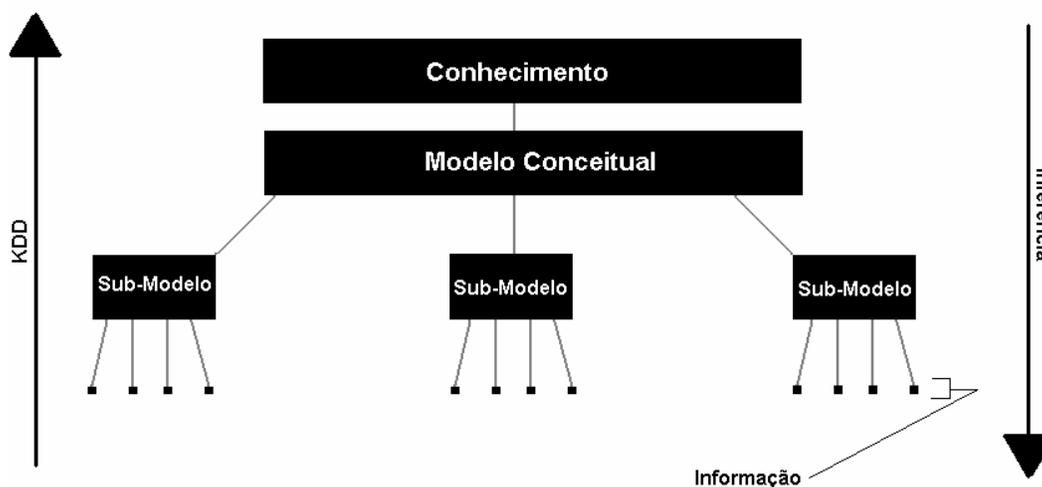


Figura 1. Estrutura lógica da arquitetura.

Num primeiro momento definiu-se qual o conhecimento sobre o negócio construindo um modelo representativo desse conhecimento (Figura 2), e que pudesse desenhar os sub-modelos condizentes com a realidade do Hospital.

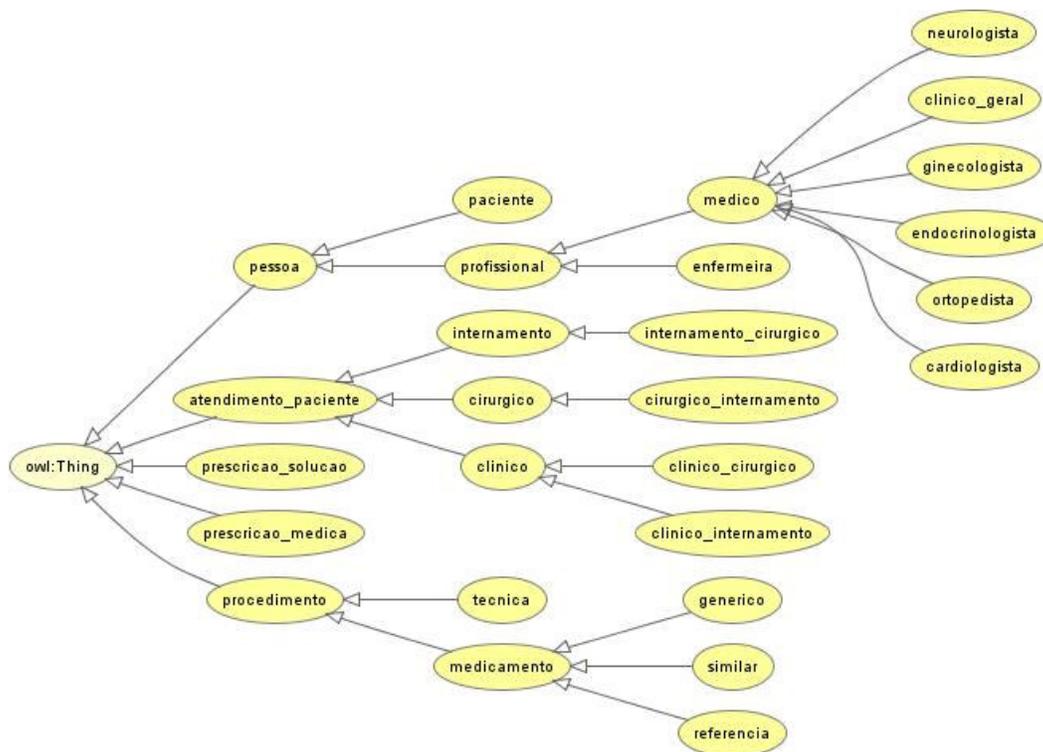


Figura 2. Representação ontológica do negócio.

Cada uma das classes refere-se a um componente do sistema e está acompanhada de seu significado, representada na linguagem OWL. No caso das especializações do atendimento ao paciente, por exemplo, as classes referem-se ao estado inicial de cada evento, podendo ocupar outros estados que são representados por outras classes. Porém, são aplicadas regras para se distinguir cada estado que aparentemente se repita. Por exemplo, não há a possibilidade de um determinado paciente dar entrada para um atendimento clínico e ao ser encaminhado ao internamento, mudar de classe representativa, ele continuará como clínico porém se encaixará na especialização de internamento, mantendo a ordem do processo.

3.1. Sub-modelo de atendimento paciente

Como o nome sugere, faz referência às informações a respeito do paciente dentro do hospital, como laudos, movimento no estoque (medicação utilizada, por exemplo). Sua nutrição e demais informações relacionadas ao atendimento dispensado ao paciente. Possui classificação segundo o tipo de atendimento dispensado, sendo este modelo importante para mapear a estadia do paciente segundo alguns aspectos, como seu plano de saúde em função de idade e sexo. Nesses parâmetros se definir quais as incidências de doenças em uma determinada época do ano. Se criando uma classe de CONVENIOS e sub-classes dentro dela de IDADE e SEXO. O SVM é capaz de treinar continuamente e com o tempo se especializando nessa tarefa.

Tipo	Descrição	Evolução
Clinico	Atendimento feito no consultório ou Pronto Socorro do hospital.	clinico_cirurgico clinico_internamento
Cirúrgico	Paciente entra no hospital para uma intervenção cirúrgica.	cirurgico_internamento
Internamento	Paciente entra para ficar no hospital	internamento_cirurgico

Tabela 1. Tipos de atendimentos dos pacientes.

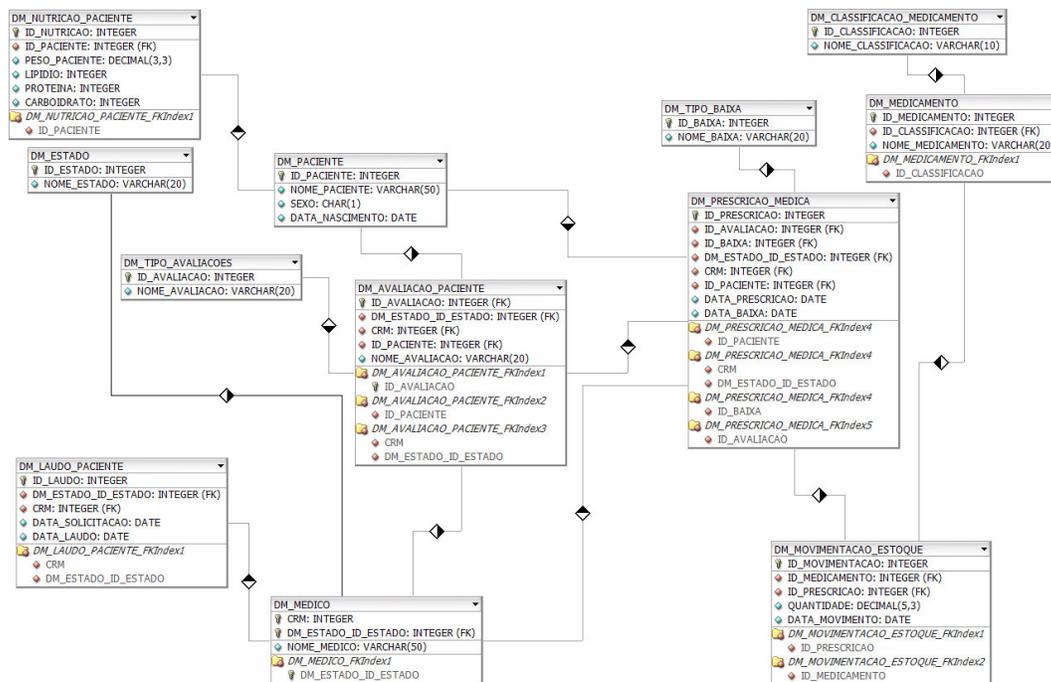


Figura 3. Sub-modelo de atendimento paciente.

3.2. Sub-modelo de Prescrição médica

É quando o médico prescreve um determinado medicamento ao paciente em questão. Normalmente o medicamento indicado não é destinado a solucionar o caso atual do paciente, mas sim definir um medicamento que não altere ou não agrave outros males ou deficiências do paciente. Um dos principais sub-modelos trabalhados, devido seu amplo co-relacionamento entre outras regras e entidades.

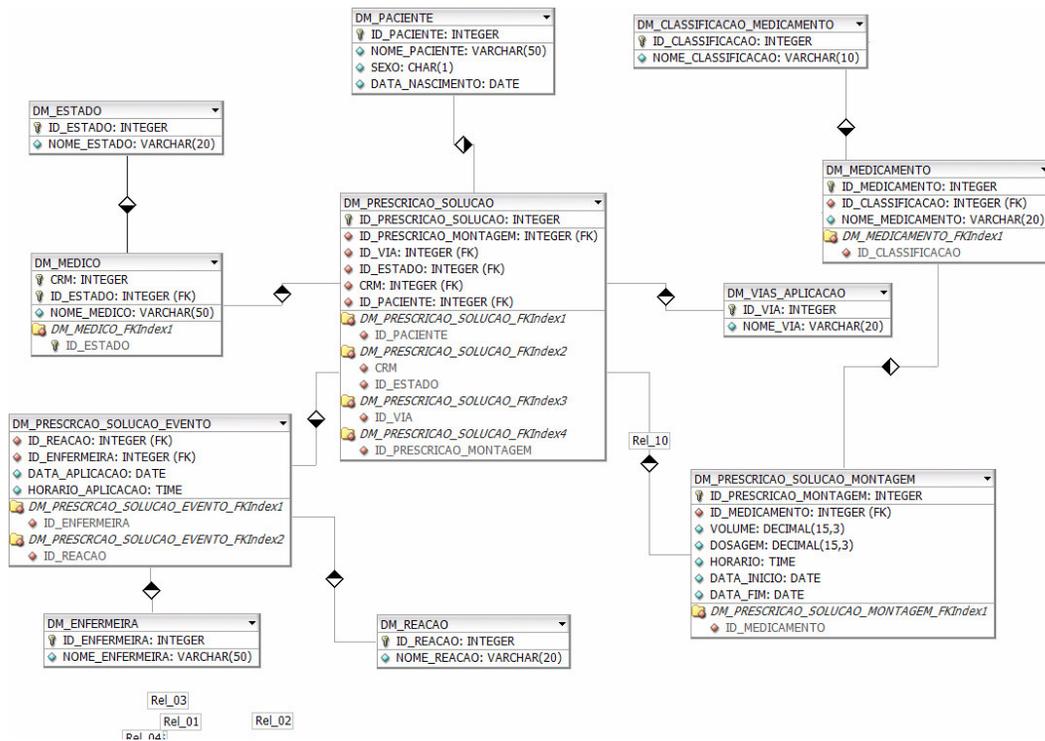


Figura 5. Sub-modelo a Prescrição Solução.

3.4. Pessoa

Dois grupos distintos que movimentam as ações, profissionais de saúde e pacientes. Enquanto os pacientes são pessoas que necessitam de serviços do hospital, com alguma suspeita de patologia ou trauma, para serem atendidos, os profissionais de saúde vêm em auxílio dos pacientes, os acompanhando e auxiliando, são classificados como médicos ou enfermeiras, onde o primeiro tem o papel principal de geração de prescrições, laudos e outras movimentações, enquanto o segundo tem a tarefa de por em prática essas orientações.

Tipo	Especialidade
Enfermeira	-
Medico	Clínico Geral
	Cardiologista
	Endocrinologista

Tabela 2. Especializações dos profissionais de saúde, alguns exemplos.

3.5. Procedimento

É o método de tratamento a ser executado no paciente em busca de sua cura. Divide-se na parte de medicamentos a serem ministrados e os hábitos a serem seguidos, como alimentação, exercícios, entre outros. Os medicamentos são ministrados aos pacientes pelos profissionais auxiliares segundo recomendação dos médicos. São reconhecidos segundo sua classificação de fórmula:

- a) genérico;
- b) referencia;
- c) similar.

Os dois últimos tópicos não se referem a áreas de interesse no negócio e sim a um conhecimento que é necessário para a extração de informação nos sub-modelos e por isso estão sendo mostradas.

4. Algoritmos

Neste momento já era conhecido o modelo de dados e a forma que os dados estavam distribuídos dentro da base que foi trabalhada. Era conhecido o nível de dados vazios e de dados faltantes. Os dados espalhados já haviam sido tratados na remodelagem para a criação dos sub-modelos. Esse tratamento foi comparado com as informações alimentadas no Tasy. Cada algoritmo possui uma peculiaridade no tratamento de cada uma dessas imperfeições, assim como podem gerar saídas inesperadas, caso esses problemas não sejam solucionados satisfatoriamente. Portanto, houve uma etapa de preparação dos dados que iriam popular o modelo proposto. Como foi utilizado o Oracle, foram aproveitados os algoritmos do ODM (*Oracle Data Mining*) na aplicação dos modelos. Foram utilizados algoritmos supervisionados como o Nave Bayes e o SVM (*Support Vector Machine*) para classificação dos dados.

Algoritmo	Tipo	Impacto de dados fora de padrão
Nave Bayes	CLASSIFICAÇÃO	Os dados se concentram em poucos agrupamentos (em casos extremos em até um agrupamento). Como resultado, o poder de discernimento dos algoritmos é extremamente reduzido.
SVM	CLASSIFICAÇÃO	Faz com que muitos dados se concentrem em uma pequena faixa de valores. Isso faz com que o aprendizado seja custoso e o treinamento demorado

Tabela 3. Algoritmos utilizados, suas definições e funções no sistema.

5. Conclusão

O levantamento do negócio em forma de ontologia foi um ponto crucial para o sucesso do trabalho, pois norteou a divisão das áreas de forma mais clara e também foi responsável pelo entendimento de como se construir os sub-modelos dessas áreas, no tocante aos atributos e entidades envolvidas, onde através das definições conseguimos testar os algoritmos e verificar a praticidade do metamodelo nas definições dos classificadores, e ainda pudemos observar o funcionamento do Tasy e dele obter, quais dados já estavam disponíveis, quais métodos de preparação de dados eram necessários para serem aplicados naquela massa a fim de obtermos o uso posterior dos algoritmos e também criar um conjunto de testes validadores. O mais importante está na melhoria dos processos de trabalho, com a compreensão dos processos e métodos. A adição desta ferramenta auxilia de forma substancial aos profissionais envolvidos na área, visando uma melhoria no desempenho do diagnóstico com aplicação dos conhecimentos gerados através dos submodelos.

6. Bibliografia

- OWL Web Ontology Language Reference (W3C Proposed Recommendation 15 December 2003) - <http://www.w3.org/TR/2003/PR-owl-ref-20031215/>
- Silberschatz, A., Korth, H. F. and Sudarshan, S. (2005) "Sistema de banco de dados", São Paulo, MAKRON Books
- SARKER, Ruhul, ABBASS, Hussein, NEWTON, Charles (2002) "Heuristics and optimization for knowledge discovery", Hershey, Idea Group.
- Propagação de Evidências em Redes Bayesianas: Diagnóstico sobre Doenças Pulmonares, MSc. Thesis, CIC-Universidade de Brasília, 1997 - <http://www.cic.unb.br/pg/mestrado/teses/rd-estevam-9570527.html>
- Protégé - <http://protege.stanford.edu/doc/users.html>
- FAYYAD, Usama M., PIATETSKY-SHAPIRO, Gregory, SMYTH Padhraic, UTHURUSAMY, Ramasamy (1996) "Advances in Knowledge Discovery and Data Mining",
- Inmon, W. H. (2002) "Building the Data Warehouse", New York, WILEY
- Johnson, J. L. (1997) "Database Models, Languages, Design", New York, Oxford
- MACHADO, Felipe Nery R. (2004) "Tecnologia e Projeto de Data Warehouse", São Paulo, Érica
- Encontros Bibli – Representação do Conhecimento na Perspectiva da Ciência da Informação em Tempo e Espaço Digitais, UFSC, 2003 - <http://redalyc.uaemex.mx/redalyc/pdf/147/14701503.pdf> - Acessado dia 05 de maio de 2007.