

METODOLOGIA PARA DESENVOLVIMENTO DE SISTEMAS DE SUPORTE A DECISÃO CRISP-DM UTILIZANDO A NOTAÇÃO UML UMA ABORDAGEM APLICADA À GERÊNCIA DE ATENDIMENTO HOSPITALAR

EDILBERTO M. SILVA¹, TIAGO G. VIOTT¹, CINDY C. ALVES¹, BRUNA A. C. PEREIRA¹

¹ *Uniplac - União Educacional do Planalto Central, Faciplac - Faculdade de Ciência e Tecnologia do Planalto Central - Campus II - Gama/DF*

E-mails: edilms@yahoo.com; tiago.viott@gmail.com; cindycastroalves@gmail.com; brunamenza@gmail.com

Abstract - Extract excellent standards of data contributes to subsidize the managers from based scenes. The CRISP-DM (*Cross Industry Standard Process for Data Mining*) is a methodology consecrated for development of support systems the decision (SSD). It consists of a set of phases and processes standards, not rigid and independent of the area of business and the used tools, of structuralized and methodical form. This model standardizes and registers the development all. This work demonstrates as the Unified Modeling Language (UML) can be applied in set with this methodology with intention to facilitate the documentation of projects. It was looked to use the graphical notation of the UML in the phases of the CRISP-DM since this manages the execution of its stages of literal form. One justifies the study of new forms of if improving the SSD construction and thus to allow the discovery of knowledge for the organizations.

Keywords - CRISP-DM, UML, Methodology, Decision Support Systems

Resumo - Extrair padrões relevantes de dados contribui para subsidiar os gestores a partir de visões fundamentadas. A CRISP-DM (*Cross Industry Standard Process for Data Mining*) é uma metodologia consagrada para desenvolvimento de sistemas de suporte a decisão (SSD). Consiste de um conjunto de fases e processos padrões, não rígidos e independentes da área de negócio e das ferramentas utilizadas, de forma estruturada e metódica. Este modelo padroniza e documenta todo o desenvolvimento. Esse trabalho demonstra como a Linguagem de Modelagem Unificada (UML) pode ser aplicada em conjunto com esta metodologia com intuito de facilitar a documentação de projetos. Procurou-se utilizar a notação gráfica da UML nas fases da CRISP-DM visto que esta gerencia a execução de suas etapas de forma totalmente textual. Justifica-se o estudo de novas formas de se melhorar e dinamizar a construção de SSD's e assim permitir a descoberta de conhecimento para as organizações.

Palavras-chave - CRISP-DM, Metodologia, Sistemas de Suporte a Decisão.

1 Introdução

Competitividade é uma palavra cada vez mais presente nas organizações. A demanda por informações de qualidade, respostas imediatas e soluções eficientes geram a necessidade crescente por avaliações rápidas e precisas sobre o negócio.

As organizações têm necessidade muito grande por dados, mas freqüentemente não possuem os mecanismos adequados para tratá-los. O volume de dados torna-se irrelevante se não estiver organizado de forma a agregar valores, e em uma organização isso significa transformar dados em informações úteis.

Neste trabalho será abordado o estudo de caso de desenvolvimento de um Data Warehouse por meio da metodologia CRISP-DM - *Cross Industry Standard Process for Data Mining* aplicando em suas fases a UML - *Unified Modeling Language*, uma vez que a CRISP-DM não dispõem de recurso gráfico para documentação de suas fases.

2 CRISP-DM

Com o intuito de promover a padronização de conceitos e técnicas na busca de informações específicas para a tomada de decisões, foi criado em 1996 o grupo de trabalho que propôs a criação de uma metodologia que auxiliaria os administradores e responsáveis no processo geral de planejar e executar a mineração de dados (SILVA, 2002). A essa metodologia deu-se o nome de CRISP-DM.

Essa metodologia consiste em um conjunto de fases e processos padrões para desenvolver projetos de *Data Mining*, independentemente da área de negócio e das ferramentas utilizadas, de uma forma estruturada e metódica. Tem como objetivos principais: transformar necessidades de negócio em tarefas de *Data Mining*, sugerir transformações nos dados e técnicas adequadas, utilizar métricas para avaliar a qualidade dos resultados obtidos e documentar o projeto.

O ciclo de vida de um projeto de Data Mining é dividido em fases, cada uma concluída por um marco principal. A transição entre as fases depende do resultado de sua avaliação. A avaliação satisfatória permite que o projeto passe para a próxima fase.

Na versão 1.0 recomendam-se seis fases:

2.1 Compreensão do negócio

Essa fase inicial procura identificar as metas, os objetivos e as necessidades numa perspectiva de negócio e converter esse conhecimento, numa tarefa de *Data Mining*, num plano inicial de ataque ao problema.

2.2 Compreensão dos dados

Essa fase compreende a identificação da informação que possa ser relevante para o estudo e uma primeira familiarização com o conteúdo, descrição, qualidade e utilidade dos dados.

2.3 Preparação dos dados

Consiste em uma série de atividades destinadas a obter um conjunto final de dados, a partir do qual, será criado e validado o modelo. Nesta fase são utilizados programas de extração, limpeza e transformação dos dados.

2.4 Modelagem

Na fase de modelagem são selecionadas e aplicadas as técnicas de descoberta de conhecimento mais apropriadas em relação aos objetivos pretendidos. Assim, essa fase consiste em selecionar os algoritmos a serem utilizados e o efetivo processamento do modelo.

2.5 Avaliação

Essa fase consiste em avaliar os modelos, para verificar se os resultados vão ao encontro dos objetivos do negócio, definidos inicialmente. A partir daí, é verificado se o processo encerra, devendo-se então passar para a fase de avaliação, ou se deverão ser efetuadas correções no processo, voltando para as fases anteriores, ou mesmo reiniciar um novo processo.

2.6 Aplicação

Essa fase consiste em um conjunto de ações que conduzam à utilização dos resultados (padrões valiosos) oriundos das técnicas/soluções aplicadas no negócio.

3 UML

A UML é uma linguagem padrão para visualizar, especificar, construir e documentar artefatos de um sistema baseado em software, tendo como objetivo prover as necessidades de desenvolvedores de software com uma linguagem visual completa.

Como é inviável representar um sistema na sua completude por meio de um único diagrama, é necessário um conjunto de recursos que expresse os diversos aspectos que compõem o sistema. Pensando nesse contexto, a UML possibilita empregar várias notações gráficas que buscam caracterizar o sistema na sua totalidade.

O sistema é descrito em facetadas (visões), em cada qual se observa um aspecto particular do sistema e, as junções dessas visões mostram o sistema na sua totalidade. Cada visão está composta por um conjunto de diagramas que retratam a particularidade enfatizada pela visão (TONSIG, 2003).

Os diagramas utilizados pela UML são compostos de nove tipos: diagrama de caso de uso, diagrama de classe, diagrama de objetos, diagrama de seqüência, diagrama de colaboração, diagrama de estado, diagrama de atividade, diagrama de componente, diagrama de implantação.

4 Resultados

O projeto fundamentou-se na junção de diagramas da UML às fases da CRISP-DM, onde foi feito um estudo apurado da metodologia, buscando verificar aonde poderiam ser aplicados os diagramas da UML.

A metodologia proposta foi validada por meio do desenvolvimento de um Data Warehouse voltado para gerência de atendimento hospitalar utilizando como base a CRISP-DM e aplicando às suas fases diagramas da UML.

As três primeiras fases da CRISP-DM: Compreensão do Negócio, Compreensão dos Dados e Preparação dos Dados representam cerca 70% de todo o desenvolvimento, sendo as fases que exigem maior esforço. Nestas, a pesquisa possibilitou o a aplicação de três diagramas da UML que contribuíram significativamente para a melhor visualização de todo o contexto dessas fases.

Na fase de Modelagem que é o núcleo do desenvolvimento foram agregados mais dois diagramas.

As duas últimas fases da CRISP-DM, avaliação e aplicação, versam sobre a análise dos resultados obtidos por meio de relatórios. Como o objetivo da pesquisa era associar a visualização gráfica às fases da CRISP-DM por meio de diagramas da UML, verificou-se que não seria possível essa análise por meio de diagramas.

A figura 1 demonstra a visualização da estrutura proposta para a CRISP-UML. Durante o

desenvolvimento do sistema de suporte a decisão verificou-se que a maior demanda de trabalho está concentrada nas três primeiras fases da CRISP-DM, então com a junção de cinco diagramas da UML houve uma contribuição significativa para melhor visualização de todo o contexto dessas fases cobrindo assim mais de 85% de todo o desenvolvimento do SSD.

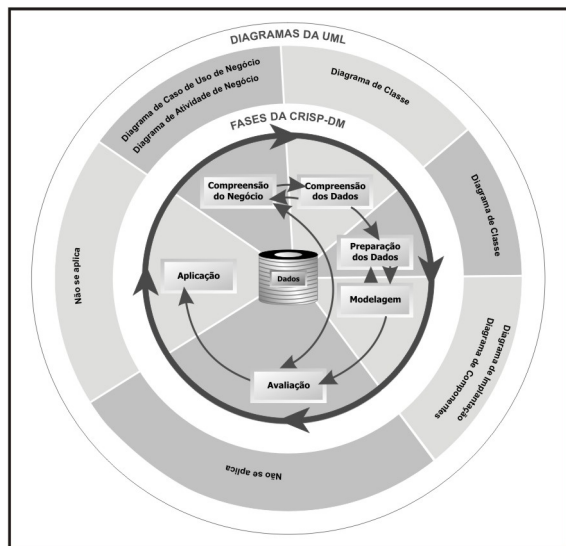


Figura 1. CRISP-UML: junção da CRISP-DM com a UML

4.1 Modelagem do Negócio

Os diagramas da UML são voltados para o desenvolvimento de software, mas existe na CRISP-DM a fase de compreensão do negócio que como o próprio nome sugere, abrange a visão do negócio aonde será desenvolvido o SSD.

A solução encontrada para representar graficamente essa fase foi a utilização dos diagramas de negócio explicitados no RUP (*Rational Unified Process*). Esses diagramas são idênticos aos diagramas da UML o que os diferencia é a abordagem onde um foca o desenvolvimento do software e o outro a visão do negócio do cliente.

O caso de uso de negócios consiste em uma seqüência de atividades que, juntas, produzem algo para o ator de negócios. A sua finalidade para os analistas de sistemas é poder compreender o contexto do sistema, e o cliente poderá certificar-se que o contexto do negócio foi entendido. O fluxo de trabalho de um caso de uso de negócios descreve o que o negócio deve fazer para fornecer o valor que o ator de negócios servido requer (Figura 2).

Um diagrama de atividades de um fluxo de trabalho explora a ordem das tarefas ou das atividades que realizam as metas do negócio. Uma atividade pode ser uma tarefa manual ou automatizada que completa uma unidade de trabalho (Figura 3).

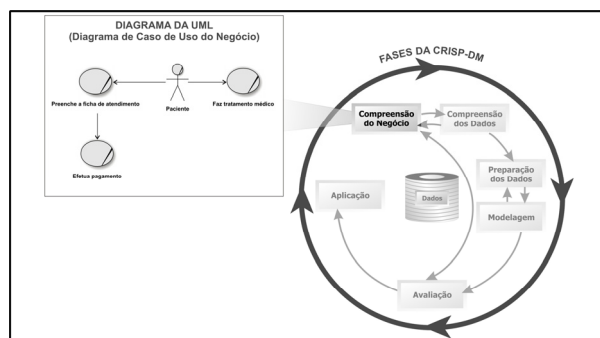


Figura 2. Aplicação do caso de uso de negócio na Compreensão do Negócio

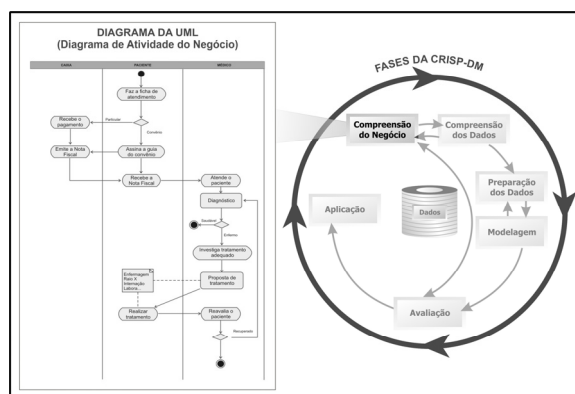


Figura 3. Aplicação do diagrama de atividade do negócio na Compreensão do Negócio

4.2 Diagrama de Classe

Analisando o diagrama de classe apresentado pela UML, percebeu-se que este pode ser aplicado para demonstrar, na compreensão dos dados, as tabelas do banco com o seu respectivo conteúdo, aplicado antes do refinamento das tabelas.

Isso se justifica, pois, esse diagrama é o único na UML que modela cada recurso em termos de sua estrutura, relacionamentos e comportamentos, esses recursos podem ser pessoas, materiais ou informações.

Outra característica importante do diagrama é a possibilidade de visualizar o nome, os atributos e as operações de cada classe, já que no entendimento dos dados busca-se justamente a compreensão do legado. De posse dessa informação e com o conhecimento das próximas fases da CRISP-DM, será possível planejar ações futuras com relação a modificações nos dados, se necessário (figura 4).

O diagrama de classe na preparação dos dados é mais específico, pois os dados já estão refinados sendo possível verificar se os dados disponíveis são suficientes para responder as perguntas do objetivo da mineração

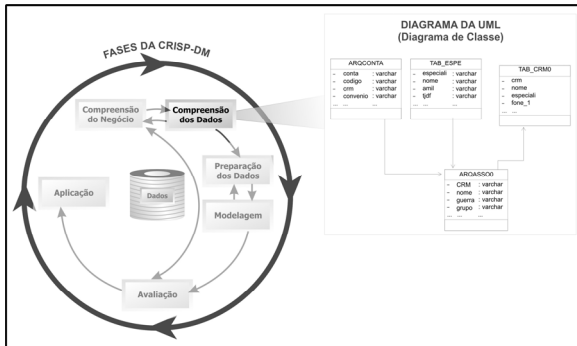


Figura 4. Aplicação do diagrama de classe na Compreensão dos Dados

A definição de quais classes utilizar somente é possível devido à compreensão do negócio, onde foi identificado o que o cliente almeja do sistema de suporte a decisão (figura 5).

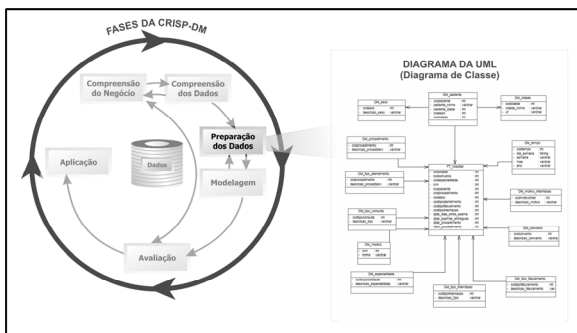


Figura 5. Aplicação do diagrama de classe na Preparação dos Dados

O diagrama de classe por ser apurado, é possível sua utilização também no desenvolvimento do modelo multidimensional (figura 6). Esse modelo não é um diagrama da UML, mas é uma importante ferramenta para a modelagem de sistemas de suporte a decisão.

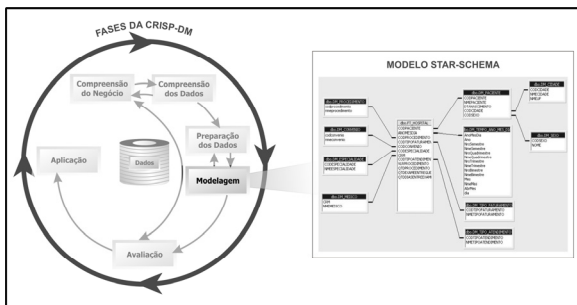


Figura 6. Modelo Star-Schema

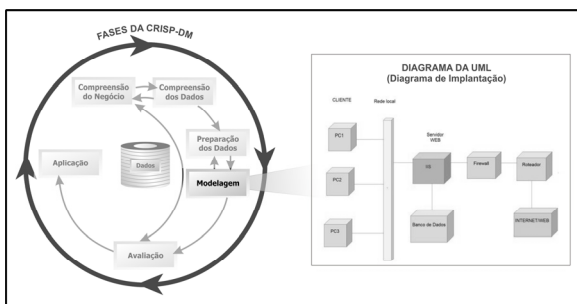


Figura 7. Aplicação do diagrama de implantação na Modelagem

4.3 Modelagem

A modelagem na CRISP-DM representa a fase central da mineração, incluindo escolha, parametrização e execução de técnica(s) sobre o conjunto de dados visando à criação de um ou vários modelos (SILVA, 2002).

Na pesquisa foi utilizado o diagrama de implantação (figura 7), para representar a arquitetura física do sistema, e o diagrama de componentes (figura 8), que mostra os vários componentes de software de um sistema e suas dependências.

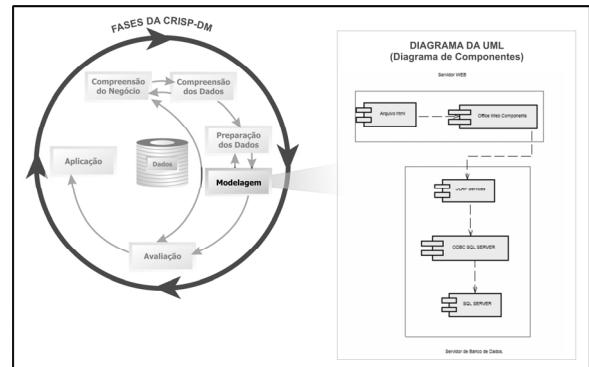


Figura 8. Aplicação do diagrama de componentes na Modelagem

5 Conclusão

O estudo resultou em uma nova abordagem intitulada CRISP-UML proporcionando um entendimento mais conciso tanto para os desenvolvedores, quanto para o cliente. Contribui para desenvolvimento de sistemas de suporte a decisão mais eficiente, com recurso valioso da UML - a notação gráfica.

A principal contribuição é a proposta de um modelo de desenvolvimento de sistemas de suporte a decisão mais ágil e clara, por meio da CRISP-DM associada aos diagramas da UML.

A CRISP-UML contribui na documentação, que é uma tarefa onerosa no desenvolvimento de Sistemas de Suporte a Decisão - SDD.

Permite que a documentação do projeto seja mais compreensível para clientes e técnicos envolvidos, visto que é possível a utilização de recursos visuais em grande parte do processo de desenvolvimento do SSD.

Para os clientes a CRISP-UML torna as fases de homologação do levantamento de requisitos e validação do escopo do projeto mais simples, visto que os diagramas são de fácil entendimento.

Para os desenvolvedores esta nova proposta permite diminuir o tempo e esforço na execução das fases de compreensão do Negócio até a preparação (cerca de 70% do esforço). Além disso, melhora substancialmente a apresentação final do documento por meio da notação gráfica.

O resultado da pesquisa tende a proporcionar benefícios tangíveis para empresas especializadas em

desenvolvimento de sistemas de suporte a decisão, bem como para gestores de projetos, pois o processo de validação da solução proposta poderá ser realizado de forma mais dinâmica.

Referências Bibliográficas

- BARBIERI, Carlos. BI – Business Intelligence, Modelagem & Tecnologia. Rio de Janeiro: Axcel Books, 2001.
- BEZERRA, Eduardo. Princípios de Análise e Projeto de Sistemas com UML: Um guia prático para modelagem de sistemas orientados a objetos através da Linguagem de Modelagem Unificada. Rio de Janeiro : Campus, 2003.
- BOOCH, Grady; RUMBAUGH, James; JACOBSON, Ivar. UML – Guia do Usuário. 2ª ed. Rio de Janeiro : Campus, 2006.
- BOOCH, Grady; RUMBAUGH, James; JACOBSON, Ivar. UML – Guia do Usuário. 1ª ed. Rio de Janeiro : Campus, 2000.
- BRAUNER, Daniela Francisco. O Processo de Descoberta de Conhecimento em Banco de Dados: Um estudo de caso sobre os dados da UFPel. 2003. 75f. Trabalho de conclusão de curso (Graduação) - Universidade Federal de Pelotas, Rio Grande do Sul - RS.
- CALAÇO, Methanias Junior. Projetando Sistemas de apoio à decisão baseados em Data Warehouse. Rio de Janeiro: Axcel Books, 2004.
- CRISP-DM Cross Industry Standart Process for Data Mining. Disponível em:< <http://www.crisp-dm.org/>>. Acessado em: dez. 2005.
- DAVENPORT, T. H., PRUSAK, L. Conhecimento Empresarial: como as organizações gerenciam o seu capital intelectual. Rio de Janeiro: Campus, 1998.
- FURLAN, José Davi. Modelagem de objetos através da UML, Análise e desenho orientados a objetos. São Paulo: Makron Books, 1998.
- PENDER, Tom. UML: A bíblia. 2ª ed. Rio de Janeiro: Campus, 2006.
- SILVA, Edilberto Magalhães. Descoberta de Conhecimento com o uso de Text Mining: Cruzando o Abismo de Moore. 2002. 175f. Dissertação (Mestrado), Universidade Católica de Brasília, Brasília - DF. 2002.