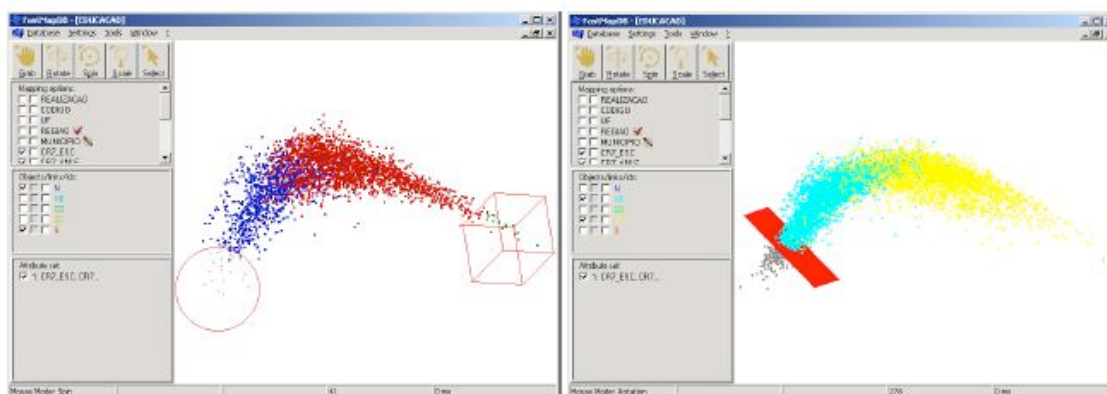


**ESTENSÃO DA CAPACIDADE DE PREDIÇÃO EM AMBIENTE DE TRANSFORMAÇÃO DOS DADOS PARA EXTRAÇÃO DE CONHECIMENTO ATRAVÉS DE INTERFACE GRÁFICA E INTERATIVA.** Felipe de Campos Fabel, Carlos Roberto Valêncio. Bacharelado em Ciência da Computação – Departamento de Ciência de Computação e Estatística – Instituto de Biociências, Letras e Ciências Exatas – Campus São José do Rio Preto.

O acelerado crescimento na coleta e armazenamento de dados trouxe consigo o conceito de *Data Mining*, que sugere a extração de conhecimentos pouco explícitos de uma base de dados. A etapa de seleção e transformação dos dados é uma das chaves para o bom andamento do processo de descoberta de conhecimento, - *KDD (Knowledge Discovery in Databases)* - e conseqüentemente do *Data Mining*. Considerando este fato, o projeto especificado busca aprimorar esta etapa, possibilitando a confecção de tabelas mais moldáveis às necessidades do usuário da ferramenta *FastMapDB*.

A ferramenta apresentada neste trabalho tem por objetivo realizar a projeção dos dados presentes em uma dada relação. Essa projeção é feita em um ambiente interativo de maneira que os dados selecionados para a prospecção, são visualizados em um gráfico tri-dimensional como mostra a figura 1. Diversos são os recursos oferecidos pelo *FastMapDB* para manipulação dos dados a serem minerados. Um problema enfrentado é que na maioria das vezes os dados desejados numa prospecção de *Data Mining*, estão espalhados pelas diversas tabelas da base na qual estão contidos. Sendo assim, o módulo de junções proposto por este projeto tem o intuito de operar sobre as diversas tabelas, de maneira que as informações desejadas sejam reunidas numa única tabela.



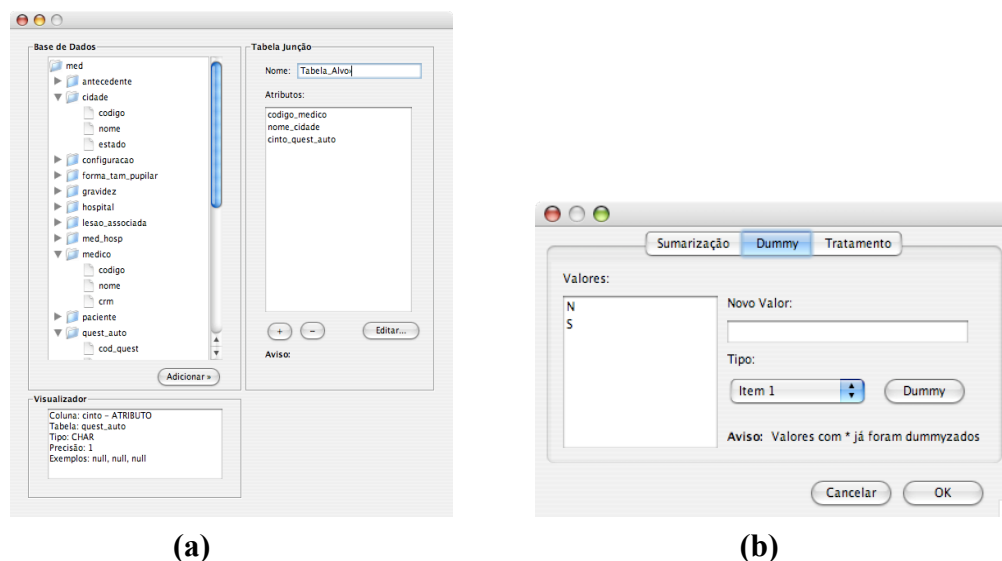
**Figura 1.** Representação gráfica de dados na ferramenta FastMapDB

Entendem-se por tabelas mais moldáveis, a possibilidade de junções entre tabelas de maneira simplificada – que devem ser realizadas a partir da restrição do algoritmo *FastMap* trabalhar sobre um tabela apenas - e criação de diversas variáveis derivadas. Os resultados possibilitarão maior eficácia na sequência do processo, que inclui as fases de predição e classificação de dados, no que diz respeito à escalabilidade, precisão e desempenho.

Tem-se que o resumo de dados em suas diversas formas, fornece muito mais informações sobre o comportamento de uma variável do que a própria tabela original de dados [BUS\_03]. Muitas vezes, deseja-se resumir ainda mais estes dados, apresentando um, ou alguns valores que sejam representativos da série toda. Assim, a possibilidade de criação de medidas-resumo foi contemplada no intuito de trazer melhores resultados na mineração.

A criação de modelos multidimensionais, resultante das junções, possibilitará mudanças estruturais na base de dados, dando suporte ao esquema de *Data Warehouse*, que por sua vez melhorará o curso do algoritmo *FastMap*, e também a visualização dos resultados dos dados prospectados.

O objetivo central do projeto, no que tange a uma boa interação com o usuário, é trazer a ele transparência e autonomia na manipulação de seus dados. Como pode ser observado na figura 2.a, basta ao usuário selecionar as linhas das tabelas e executar os tratamentos e operações desejados, como mostra a figura 2.b. Depois disso, a nova tabela é formada e seus dados podem ser levados ao módulo de visualização para que a prospecção seja realizada.



**Figura 2.** Telas do sistema. (a) Engenharia Reversa; (b) Ferramentas para tratamento de dados

Unindo a otimização de junções e o esquema de *data warehouse* gerado, a gama de aplicações que utilizam o *FastMapDB* pode se tornar muito mais ampla. Isso se deve ao fato de que mesmo aplicações muito grandes, e com muitos dados, poderão ser contempladas, pois a tarefa de junção será realizada de modo a buscar uma maneira ótima para executá-la. Além disso, as mais diversas áreas poderão ter seus dados prospectados independentemente da organização, pois o módulo de junções acoplado à ferramenta eliminará esta barreira.

### Referências Bibliográficas:

MOLINA, H.G.; ULLMAN, J.D.; WIDOM, J. Implementação de Sistemas de Bancos de Dados, 1a. Edição, Editora Campus Ltda., 2001, cap. 7, p. 391-451

MCHUGH, J. A. Algorithmic Graph Theory, 1st Edition, Prentice-Hall, Inc., 1990, 326p.

HAN, J.; KAMBER, M. Data mining: concepts and techniques, San Francisco: MKP, 2001. 550p.

**ÁREA:** Ciências Exatas

**SUBÁREA:** 1.03 Ciência da Computação