

CENTRO UNIVERSITÁRIO FEEVALE

GUILHERME COIMBRA MOSTARDEIRO

DESCOBERTA DE CONHECIMENTO EM UMA BASE DE DADOS  
CRIADA A PARTIR DE KEYLOGGERS  
(Título Provisório)

Anteprojeto de Trabalho de Conclusão

Novo Hamburgo, Setembro de 2009

GUILHERME COIMBRA MOSTARDEIRO

gcm@feevale.br

DESCOBERTA DE CONHECIMENTO EM UMA BASE DE DADOS  
CRIADA A PARTIR DE KEYLOGGERS

(Título Provisório)

Centro Universitário Feevale  
Instituto de Ciências Exatas e Tecnológicas  
Curso de Sistemas de Informação  
Anteprojeto de Trabalho de Conclusão

Professor orientador: Ricardo Ferreira de Oliveira

Novo Hamburgo, Setembro de 2009.

## RESUMO

As técnicas de mineração de dados têm como objetivo principal a extração de informações úteis a partir de grandes quantidades de dados. O uso deste tipo de tecnologia surgiu justamente para assumir a responsabilidade de exame e interpretação de dados onde estas tarefas não são possíveis de serem executadas manualmente, ou seja, sem o uso de computação. Portanto torna-se imprescindível o desenvolvimento de ferramentas que auxiliem o homem, de forma automática e inteligente, na tarefa de analisar, interpretar e relacionar esses dados para que se possa desenvolver e selecionar estratégias de ação em cada contexto de aplicação. Sendo assim, este trabalho propõe o desenvolvimento de uma ferramenta que irá atuar na tarefa de mineração de dados, onde estes foram capturados anteriormente pela utilização de *keyloggers*. As informações capturadas por *keyloggers* são muito valiosas em relação a tarefa de medição da produtividade dos usuários. Com elas poderemos saber que informações o usuário inseriu no computador tiveram relação com o negócio da empresa, assim como as informações que não têm nenhuma relação com o ambiente empresarial, e acabam afetando a produtividade do mesmo. Também poderão ser extraídos dados estatísticos e feitas associações como, por exemplo, que tipo de informações são inseridas para cada tipo de aplicação que o usuário está utilizando, ou até mesmo associar por título de janela do sistema operacional.

Palavras-chave: Descoberta de conhecimento. Reconhecimento de padrões. Inteligência artificial. *Keyloggers*.

## SUMÁRIO

|                   |    |
|-------------------|----|
| MOTIVAÇÃO.....    | 5  |
| OBJETIVOS.....    | 8  |
| METODOLOGIA.....  | 9  |
| CRONOGRAMA.....   | 10 |
| BIBLIOGRAFIA..... | 12 |

## MOTIVAÇÃO

Atualmente, o mundo está “transbordando” de informações, e este fato só continua a aumentar sem previsão de fim, se é que terá algum dia. Os computadores tornaram tão fácil a tarefa de guardar dados que as pessoas acabam guardando dados que provavelmente já teriam sido apagados caso não fosse a tecnologia. A facilidade está na obtenção de mais “espaço” para guardar coisas, assim informações não precisam mais ser descartadas. Dentre todas estas informações que poderiam ser descartadas, podem haver muitas informações importantes e potenciais “escondidas” de forma que se forem analisadas corretamente podem ser de muito valor (FRANK, WITTEN. 2009).

No início, os computadores serviam para nos livrar das tarefas de rotina: absorviam dados, faziam contas e imprimiam relatórios. Como alguns desses dados precisavam ser constantemente acessados para consulta e atualização por diferentes tipos de usuários, sendo muitas vezes, compartilhados por mais de um sistema automatizado de informação, surgiu e, depois de alguns anos, passou a ser amplamente utilizada a tecnologia de banco de dados. E a partilha de dados foi se intensificando progressivamente com as ligações através de redes (GOLDSMITH, PASSOS. 2003).

Durante todo esse tempo, um sonho de todo especialista em sistemas foi o de ver as pessoas se tornarem mais racionais na tomada de decisões, pelo menos a partir do dia em que pudessem contar com um acesso rápido e fácil a esses vastos repositórios de dados. Mas o sonho nem sempre se realiza, infelizmente. Examinar grandes volumes de dados brutos requer esforço, interpreta-los corretamente pode exigir talento e um nível de capacidade técnica não trivial, até mesmo para os especialistas (GOLDSMITH, PASSOS. 2003).

A mineração de dados (também conhecida pelo termo inglês *Data Mining - DM*), é o processo de explorar grandes quantidades de dados à procura de padrões consistentes, como regras de associação ou seqüências temporais, para detectar relacionamentos sistemáticos entre variáveis, transformando os dados processados em conhecimento útil.

Segundo Jerônimo, o DM vai muito além da simples consulta a uma banco de dados, pois possibilita aos usuários a captura de informações úteis a partir dos dados, tentando descobrir conhecimentos escondidos.

A tecnologia de mineração de dados é motivada pela necessidade de uso de novas técnicas que auxiliem na análise, entendimento e visualização de muitas informações armazenadas em bancos de dados capturadas de aplicações de tipos variados. É impulsionada pela descoberta de padrões e hipóteses, que são automaticamente extraídos da base de dados, para posteriormente serem utilizadas em novos dados na tentativa de detectar a mesma situação ou para auxiliar em decisões (KIM, 2004).

Basicamente, o tratamento que a mineração de dados aplica sobre os dados não tem nada de novo, as pessoas buscam por padrões em dados desde o início de sua existência. Caçadores buscam por padrões no comportamento de migração de animais, políticos buscam por padrões em opiniões de eleitores, fazendeiros buscam por padrões relacionados ao aumento de safra, e a partir do momento em que padrões são encontrados, estes podem ser utilizados futuramente em novas situações, ou seja, se baseando na experiência para identificar oportunidades (FRANK, WITTEN. 2009).

A descoberta de conhecimento em uma base de dados pela mineração de dados ultrapassa os limites de percepção de determinado usuário sobre a mesma. Seria praticamente impossível a classificação/extração manual de informações dentre milhões de registros existentes, ao contrário do que muito se pensava antigamente, simplesmente melhorando o acesso a informação.

Diante disto surgiu a possibilidade de automatização do processo de descoberta de conhecimento, utilizando-se de técnicas de mineração de dados, que realmente vieram em socorro, e acabaram redirecionando a responsabilidade do exame e interpretação dos dados ao computador.

A descoberta de conhecimento em bases de dados (Knowledge Discovery in Databases – KDD), vem despertando grande interesse junto às comunidades científica e industrial. O termo KDD foi formalizado em 1989 em referência ao amplo conceito de procurar conhecimento a partir de bases de dados (GOLDSMITH, PASSOS. 2003).

A análise de grandes quantidades de dados pelo homem é inviável sem o auxílio de ferramentas computacionais apropriadas. Portanto torna-se imprescindível o desenvolvimento de ferramentas que auxiliem o homem, de forma automática e inteligente, na tarefa de analisar, interpretar e relacionar esses dados para que se possa desenvolver e selecionar estratégias de ação em cada contexto de aplicação (GOLDSMITH, PASSOS. 2003).

O foco do presente trabalho será a aplicação de descoberta de conhecimento em uma base de dados que foi gerada a partir do uso de um *keyloggers*, programa que captura todo conteúdo digitado em determinado(s) computador(s), de determinada rede, independente de seu uso principal. Analisando o tema do trabalho proposto, podemos perceber o quão complicado seria caso fosse necessária uma classificação dos dados e extração/descoberta de conhecimento manualmente nesta suposta base de dados, por este motivo é que se fez necessária a utilização de técnicas de descoberta de conhecimento, caso contrário a análise de todas as informações capturadas ficaria inviável.

O desenvolvimento deste trabalho poderá ser aplicado na prática em ambientes organizacionais, como empresas, para verificar que tipo de informações são inseridas nos computadores por parte dos usuários, e em cima disto, poderemos fazer uma análise utilizando-se das técnicas de mineração de dados propostas, para que todas as informações possam ser classificadas sem que seja necessária uma interação por parte de algum usuário, haja visto que seria necessário muito tempo para a execução desta tarefa.

Diante disto, a proposta deste trabalho é, primeiramente, identificar, através de pesquisas, a melhor maneira, utilizando-se das técnicas de mineração de dados, para que seja possível o tratamento das informações adquiridas. Baseando nestas pesquisas e nos resultados obtidos, será proposta a modelagem e desenvolvimento de uma ferramenta capaz de executar a tarefa de mineração de dados, assim como a modelagem da ferramenta responsável pela captura dos dados que serão tratados.

## OBJETIVOS

### **Objetivo geral:**

O objetivo geral deste trabalho é propor a análise, modelagem e desenvolvimento de uma ferramenta que se utilize de técnicas de mineração de dados, para que possa ser utilizada na classificação de informações e descoberta de conhecimento em uma base de dados criada a partir de informações capturadas pelo intermédio de *keyloggers*.

### **Objetivos específicos:**

Dentre os objetivos, podem ser definidos como específicos os que se seguem:

- Apresentar fundamentação teórica a respeito do tema;
- Analisar técnicas de mineração de dados buscando identificar entre estas as que possam ser melhores aplicadas para o desenvolvimento da ferramenta;
- Analisar e realizar testes com algoritmos da técnica de mineração escolhida;
- Realizar modelagem das ferramentas em UML;
- Efetuar modelagem da ferramenta de captação das informações;
- Efetuar modelagem da ferramenta de mineração dos dados;
- Aplicar as ferramentas desenvolvidas em alguma rede organizacional, a fim de verificar sua desenvoltura.

## METODOLOGIA

Segue listagem de atividades que serão realizadas no decorrer deste trabalho:

1. Elaboração do anteprojeto após definição do assunto do trabalho juntamente com o professor orientador;
2. Revisão e entrega do anteprojeto após ter sido avaliado pelo orientador;
3. Pesquisa e fundamentação teórica em técnicas de mineração de dados a fim de encontrar o algoritmo que melhor satisfaça o tema proposto;
4. Pesquisa e fundamentação teórica focalizada na técnica de mineração que foi escolhida, para uso na ferramenta a ser desenvolvida;
5. Execução de testes técnicos, ou seja, voltados a aplicação de algoritmos de mineração de dados na plataforma de desenvolvimento de software escolhida;
6. Elaboração do relatório do Trabalho de Conclusão I;
7. Revisão e entrega do relatório do Trabalho de Conclusão I;
8. Desenvolvimento do protótipo responsável pela a captura dos dados que serão analisados posteriormente;
9. Desenvolvimento do protótipo de ferramenta responsável pela mineração de dados;
10. Execução de testes sobre ferramenta proposta, a fim de verificar resultados, corrigir problemas e identificar melhorias;
11. Aplicação do software desenvolvido em alguma organização buscando analisar resultados com situações reais;
12. Elaboração do relatório do Trabalho de Conclusão II;
13. Revisão e entrega do relatório do Trabalho de Conclusão II;
14. Elaboração de apresentação do Trabalho de Conclusão II.

## CRONOGRAMA

### Trabalho de conclusão I

| Etapas   | Meses |     |     |     |
|--|-------|-----|-----|-----|
|  | Ago   | Set | Out | Nov |
| Elaboração do anteprojeto após definição do assunto do trabalho juntamente com o professor orientador.   | ■     | ■   |     |     |
| Revisão e entrega do anteprojeto após ter sido avaliado pelo orientador.   |       | ■   |     |     |
| Pesquisa e fundamentação teórica em técnicas de mineração de dados a fim de encontrar o algoritmo que melhor satisfaça o tema proposto.                |       | ■   | ■   |     |
| Pesquisa e fundamentação teórica focalizada na técnica de mineração que foi escolhida, para uso na ferramenta a ser desenvolvida.                      |       |     | ■   |     |
| Execução de testes técnicos, ou seja, voltados a aplicação de algoritmos de mineração de dados na plataforma de desenvolvimento de software escolhida. |       |     | ■   | ■   |
| Elaboração de relatório do Trabalho de Conclusão I.  |       |     |     | ■   |
| Revisão e entrega do relatório do Trabalho de Conclusão I.   |       |     |     | ■   |

## Trabalho de conclusão II

| <b>Etapas</b>  | <b>Meses</b> |            |            |            |
|--|--------------|------------|------------|------------|
|  | <b>Mar</b>   | <b>Abr</b> | <b>Mai</b> | <b>Jun</b> |
| Desenvolvimento do protótipo responsável pela a captura dos dados que serão analisados posteriormente.                   |              |            |            |            |
| Desenvolvimento do protótipo de ferramenta responsável pela mineração de dados.  |              |            |            |            |
| Execução de testes sobre ferramenta proposta, a fim de verificar resultados, corrigir problemas e identificar melhorias. |              |            |            |            |
| Aplicação do software desenvolvido em alguma organização buscando analisar resultados com situações reais.               |              |            |            |            |
| Elaboração do relatório do Trabalho de Conclusão II.   |              |            |            |            |
| Revisão e entrega do relatório do Trabalho de Conclusão II.  |              |            |            |            |
| Elaboração de apresentação do Trabalho de Conclusão II.  |              |            |            |            |

## BIBLIOGRAFIA

FAYYAD, Usama M; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic; UTHURUSAMY, Ramasamy; WIEDERHOLD, Gio. **Advances in Knowledge Discovery and Data Mining**. California, USA: American Association for Artificial Intelligence, 1996.

FRANK, Eibe; WITTEN, Ian H.. **Data mining: Know it all**. Burlington, MA: Elsevier, 2009.

GOLDSCHMIDT, Ronaldo; PASSOS, Emmanuel. **Data Mining: Um Guia Prático: Conceitos, Técnicas, Ferramentas, Orientações e Aplicações**. Rio de Janeiro, RJ: Elsevier, 2005.

JERÔNIMO, Paulo Marcelo. **Estudo sobre Data Mining. Data Warehouse. Cases – Data Warehouse**. Novo Hamburgo, RS: Monografia (Bacharelado em Ciência da Computação) – Institutos de ciências exatas e tecnológicas, Centro Universitário Feevale, 2001.

KIM, Jin Sung. **Customized Recommendation Mechanism Based on Web Data Mining and Case-Based Reasoning. Intelligent Agents for Data Mining and Information Retrieval**. Hershey, PA: Idea Group, 2004.

KRANZ, Paulo Henrique. **Business Intelligence: Estudo Aplicado em Cooperativa Médica**. Novo Hamburgo, RS: 2004. Monografia (Bacharelado em Ciência da Computação) – Instituto de Ciências Exatas e Tecnológicas, Centro Universitário Feevale, 2004.

LUGER, George. **Inteligência Artificial: Estruturas e Estratégias para a Solução de Problemas Complexos**. Porto Alegre, RS: Bookman, 2004.

NEMATI, Hamid R.; BARKO, Christopher D. **Organizational data mining: leveraging enterprise data resources for optimal performance**. United States of America: Idea Group, 2004.

REZENDE, Solange Oliveira. **Sistemas Inteligentes: Fundamentos e Aplicações**. Barueri, SP: Manole, 2005.

WAGNER, Eduardo Antônio. **Mineração de dados aplicada na análise de pesquisas de clima organizacional**. Campinas, SP: 2008. Artigo. Universidade Estadual de Campinas.