**Slide 2**

Este trabalho tem como objetivos gerais propor o uso da mineração de dados na análise de documentos de patentes através da aplicação da técnica de extração de padrões frequentes e realizar o uso de matriz de similaridade e Textmining para obter um melhor resultado na busca.

**Slide 3**

OMPI - Proposito a promoção da proteção da propriedade intelectual ao redor do mundo através da cooperação entre estados.

INPI - Realiza o regulamento da propriedade industrial tendo em vista função social, econômica e jurídica.

UPTO - Órgão americano responsável pelo registro de patentes.

**Slide 5**

Patentes são a concessão de um título de propriedade temporária de invenções de produtos, equipamentos, ferramentas e até mesmo de procedimentos e métodos de negócio ou de produção, desde que estejam na fase de projeto, pois não é possível patentear uma ideia.

Documentos de patentes são ativos de propriedade intelectual categorizados e indexados a partir de uma Classificação Internacional de Patentes, conhecida pela sigla IPC (International Patent Classification).A IPC é o sistema de classificação internacional, criada a partir do Acordo de Estrasburgo (1971),cujas áreas tecnológicas são divididas nas classes A a H. Dentro de cada classe, há subclasses, grupos principais e grupos, através de um sistema hierárquico.

**Slide 9**

Investigar o uso da mineração de padrões frequentes como forma de obtenção de conhecimentos a partir de documentos de patentes.

Realizar um estudo de caso com a utilização uma data set formado por um subconjunto de documentos de patentes extraído da base de dados Derwent Innovations Index.

Realizar uma avaliação experimental aplicando a metodologia proposta na geração de padrões frequentes a partir de dados estruturados e semiestruturados contidos nos documentos de patentes.

Analisar os padrões encontrados e verificar se as informações obtidas possuem valor agregado para posterior utilização. Apresentar um formato de visualização alternativa para análise de regras de associação.

**Slide 10**

Poderosa ferramenta de pesquisa de patentes que combina alguns banco de dados de patentes e e atualizado semanalmente contendo mais de 16milhoes de invenções as informações de patentes são coletadas de 41 autoridades emissoras diferentes em todo o mundo e são classificadas em 3 categorias (Quimica,Engenharia e eletrica e eltronico).

Analisar determinado conjunto de dados (“conjunto de treinamento”) e organizá-los de acordo com padrões. O reconhecimento de padrões visa classificar dados baseados em conhecimento a priori (preliminar ou dedutivo) ou informações estatísticas extraídas de padrões.

**Slide 11**

Matriz de pontuação que expressa a similaridade entre dois pontos dados, fortemente relacionadas. Pode usar similaridade dos cossenos ou distancia euclidiana.

**Slide 12**

O objetivo maior e transformar o texto em dados para analise, por meio da aplicação do processamento de linguagem natural (PLN) e de Métodos analíticos.

**Slide 13**

Um grande desafio e trabalhar com a grande quantidade de dados na realização da comparação da matriz de similaridade quando nesse caso dependendo do tamanho da base pode gerar uma matriz de similaridade muito grande e durante o processo de comparação usando o textmining. Desta forma , pode gerar um tempo muito grande para processar esses dados além disso pode precisar de uma máquina com um poder de processamento muito grande para tratar essas informações.

Uma das ideias seria realizar o cálculo de forma individual e armazenar no banco de dados para consulta dos resultados a posteriori por exemplo. Ainda não foi pensado estamos ainda no processo de escolha e rodar padrões frequentes para verificar se o agrupamento produz uma vantagem satisfatória no tamanho dos dados.