

# USO DA ANÁLISE FORMAL DE CONCEITOS PARA O CÁLCULO DE SIMILARIDADES COMO SUPORTE PARA UM SISTEMA ESPECIALISTA DE DIAGNÓSTICOS VIA WEB

Vinicius Medina Fernandes, bolsista PIBIC e estudante do curso de Engenharia de Computação da UNICAMP  
Kleber Xavier Sampaio de Souza e Sílvia Maria Fonseca Silveira Massruhá, supervisores  
Leonardo Machado Carreiro, bolsista PIBIC, colaborador

V Mostra de Trabalhos de Estagiários e Bolsistas  
Campinas, SP – 26 a 30 de outubro de 2009

## Resumo

O raciocínio para a descoberta de um diagnóstico tanto para um sistema automatizado quanto para um especialista consiste em achar a correspondência de um conjunto de sintomas numa matriz (esparsa) contendo um conjunto completo das doenças e os sintomas. Sempre que um sintoma se manifesta, o especialista ativa nessa matriz as possíveis doenças que podem manifestar este sintoma. O raciocínio prossegue até o ponto que um grupo muito pequeno de doenças (idealmente, apenas uma) seja conhecido como o causador daqueles sintomas. Como geralmente o número de sintomas é grande, é necessário agrupar as doenças similares em conjuntos de forma que os sintomas mais freqüentes são perguntados em primeiro lugar. Dessa maneira, o raciocínio favorece a redução do universo de possíveis doenças, excluindo aquelas que não manifestam o sintoma. Este trabalho avalia a similaridade entre as doenças considerando um conjunto comum de doenças e distintos sintomas e propõe um método para estruturar a o espaço de pesquisa.

## Objetivos

Atualmente, o sistema de diagnóstico, *Diagnose Virtual* concebido e construído na Embrapa Informática Agropecuária, utiliza uma medida de entropia para selecionar a próxima pergunta no processo de diagnóstico. Este trabalho propõe uma avaliação do uso de similaridades, baseada na Análise Formal de Conceito, um método fundamentado sobre a Teoria dos Reticulados, para a estruturação do espaço de pesquisa no sistema especialista.

## Material e Métodos

- Uso da ferramenta Galicia 3 para entrada da matriz de contexto contendo a relação entre as doenças e os sintomas, construção e visualização do reticulado de conceitos.
- Aplicação da Análise Formal de Conceitos.
- Programa, implementado em Java, que através do arquivo XML contendo o reticulado de conceitos: gera a similaridade numérica entre os conceitos, armazena a matriz de similaridades no formato CSV e sua visualização através de uma árvore hiperbólica no formato HTML
- Ferramenta Eclipse para desenvolvimento em Java.
- Base de conhecimento contendo 41 doenças do milho e 75 sintomas.

## Discussão e Resultados

Na Análise Formal de Conceitos, a abstração dos conceitos presentes no pensamento humano, no qual conceitos são objetos com determinadas classes de atributos, está estruturada em um reticulado, o Reticulado de Conceitos. Neste reticulado, se A é um conceito acima de um conceito B, e os dois estão ligados, o conceito A é um conceito mais geral do que B e, como tal, carrega parte dos atributos de B. Como consequência, é correto afirmar sempre que B acontece, A também está presente, sugerindo uma vinculação lógica. No reticulado, não se enxerga apenas uma hierarquia de conceitos, mas também todo o conjunto binário de relações presentes entre os conceitos. Isso faz com que a análise visual dos dados possa ser obtida olhando para uma hierarquia de classes.

Se dois objetos foram colocados no mesmo nó (conceito), eles possuem os mesmos atributos e são, portanto, instâncias da mesma classe de objetos que possuem aquele conjunto de atributos.

O número de atributos em comum pode então ser ponderado em função do número de atributos que está presente em apenas um dos objetos para medir a similaridade entre dois objetos. Contudo, muitas vezes atributos podem aparecer em pares ou trios com a consequência de serem posicionados no mesmo nó do reticulado. Isso significa que do ponto de vista estrutural, os atributos não estão adicionando informações relevantes para diferenciar objetos.

A fim de contornar este problema, uma medida de similaridade estrutural foi proposta em (Souza e Davis, 2004). Essa medida considera alguns elementos especiais do reticulado, chamados *meet-irreducibles*. Estes elementos podem ser identificados facilmente no reticulado como os nós que possuem apenas uma aresta com as camadas superiores do reticulado. No reticulado da Figura 1, dos nós 2 ao no 7 reúne-se todos os elementos irreduzíveis. Todos os seis sintomas são posicionados no topo do reticulado. A similaridade medida considera o número de elementos *meet-irreducible* em comum, bem como o número de tais elementos que cada nó ( $n_i$ ), tem em separado

Objetos	Atributos					
	s1	s2	s3	s4	s5	s6
d1	x			x		
d2	x		x	x		
d3	x		x			
d4	x					x
d5		x	x	x		
d6		x	x			
d7		x			x	
d8				x	x	x
d9		x			x	

TABELA 1: Objetos e Atributos representados no diagrama da Figura 1.

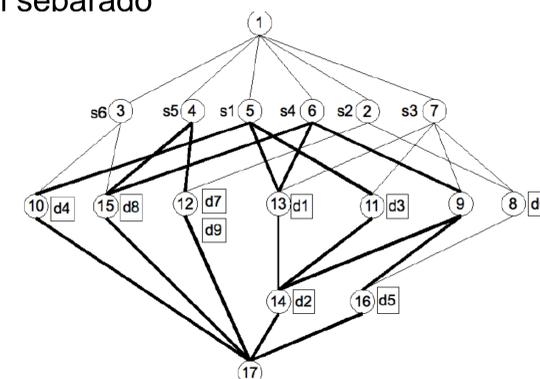


Figura 1: Diagrama correspondente ao reticulado obtido com a relação da tabela 1.

## Conclusões

Após testes realizados com doenças do milho que apresentaram resultados muito bons comparando a árvore de decisão desenhada por um fitopatologista e as similaridades calculadas usando essa medida de similaridade, foi elaborado um algoritmo que faz uso dessa medida para seleção da próxima pergunta.

O algoritmo desenvolvido resultou na implementação de um aplicativo em Java que percorre o reticulado, interpretado pelo programa através de um arquivo XML, considerando como ponto de partida para o início das perguntas um atributo estruturante selecionado pelo usuário. Após a seleção desse sintoma, todos os nós conceitos que o possuem como pai são armazenados e ordenados de acordo com a similaridade e dá-se início as perguntas através dos sintomas (atributos) de um conceito (agrupamento de atributos) com maior similaridade em relação ao sintoma estruturante. Caso confirmados, um a um, os sintomas de um conceitos, armazena-se os nós que possuem aquele conceito como pai e continua-se a percorrer recursivamente o reticulado e caso um dos sintomas não seja confirmado, o próximo nó a ser visitado é o que apresenta segunda maior similaridade com o pai e assim por diante até se esgotarem todas as possibilidades.