

UM MODELO DE ENGENHARIA DO CONHECIMENTO BASEADO EM ONTOLOGIA E CÁLCULO PROBABILÍSTICO PARA APOIO AO DIAGNÓSTICO

Luiz Fernando Lopes, Dr.¹

Alexandre Leopoldo Gonçalves, Dr.¹

José Leomar Todesco, Dr.¹

¹Universidade Federal de Santa Catarina (UFSC), Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Florianópolis, SC, Brasil

Resumo

O diagnóstico, como tarefa intensiva em conhecimento, é um processo complexo uma vez que existe uma grande variedade de elementos e circunstâncias a serem considerados para uma tomada de decisão. Incertezas geradas pela subjetividade, imprecisão e/ou falta de informações atualizadas existem em quase todos os estágios e interferem quanto à segurança e eficácia no resultado. Os dados e informações úteis, quando coletados e tratados adequadamente, provenientes de diagnósticos realizados e que permanecem em estado latente (despercebidos / adormecidos), podem tornar-se uma valiosa fonte de conhecimento se associados à experiência e observação do profissional que os utiliza. Assim, o objetivo deste artigo é propor um modelo de Engenharia do Conhecimento que possibilita a geração de novos conhecimentos para apoiar o processo de diagnóstico. As metodologias, métodos e técnicas da Engenharia do Conhecimento utilizados neste modelo para apoiar este processo são: CommonKADS, Ontologia, Cálculo Probabilístico e Sistemas de Descoberta Baseados na Literatura. Através da integração entre esses elementos, o modelo proposto é aplicado em um exemplo didático, o qual possibilita que evidências sejam destacadas e analisadas através de pesquisa literária como possíveis novos conhecimentos. Após a confirmação de um novo conhecimento, o processo de inferência é atualizado. Conclui-se, portanto, que, através desta pesquisa, o modelo proposto atende os requisitos para a geração de novos conhecimentos e contribui para o aperfeiçoamento da tarefa de diagnóstico.

Palavras-chave: Engenharia do Conhecimento. Sistemas Baseados em Conhecimento. Diagnóstico. Ontologia. Cálculo Probabilístico.

1. INTRODUÇÃO

O avanço tecnológico ocorrido nos últimos anos tem impulsionado o desenvolvimento de Sistemas Baseados em Conhecimento (SBC) que utilizam a inteligência computacional para auxiliar nas tarefas intensivas em conhecimento (ZHAO, YANXIANG & HUI, 2005; RODRÍGUEZ *et al.*, 2009a). Entretanto, alavancar o potencial destas aplicações, em relação ao processo de diagnóstico, tem sido uma ação crítica e desafiadora no tocante à confiabilidade destes sistemas (GARCÍA-CRESPO *et al.*, 2010).

Vários modelos correlatos têm sido propostos para auxiliar o profissional no processo de diagnóstico tais como: Modelo SOMKS, proposto por Minchin *et al.* (2006); Modelo FENZA proposto por Fenza *et al.* (2008); Modelo ODDIN, proposto por García-Crespo *et al.* (2010), porém, a precisão e a eficiência ainda são metas a serem alcançadas (GARCÍA-CRESPO *et al.*, 2010).

O diagnóstico, de uma forma geral é sempre um processo complexo que exige elevado nível de conhecimento e, tendo uma grande quantidade de variáveis a serem consideradas, muitas vezes dificulta o seu desempenho (SHORTLIFFE *et al.*, 2001).

Portanto, um grande desafio tem sido como apoiar este processo, onde os sistemas são projetados e desenvolvidos baseados em precisão e certeza. Eles proveem soluções irreais baseados em um ambiente fechado e para a maioria das aplicações reais as incertezas são inevitáveis e não podem ser ignoradas (YANG, 2007).

A Engenharia do Conhecimento (EC) é uma área que pode auxiliar este processo através de um conjunto de métodos, técnicas e ferramentas que oferecem suporte à Gestão do Conhecimento (GC) para a formalização e explicitação das atividades intensivas em conhecimento (SCHREIBER *et al.*, 2002).

Um destes instrumentos é a metodologia CommonKADS (*Knowledge Acquisition and Documentation Structuring*), que fornece um conjunto de etapas para o desenvolvimento de SBC (SCHREIBER *et al.*, 2002). No nível de contexto, o Modelo de Organização desta metodologia suporta a maioria dos recursos da organização para descobrir problemas e oportunidades, estabelecendo sua viabilidade e avaliando os possíveis impactos do desenvolvimento destes sistemas.

No nível de conceito, o Modelo de Conhecimento desta metodologia procura especificar tarefas intensivas de conhecimento, explicando em detalhes o tipo e a estrutura do conhecimento usado na execução de uma tarefa. Já ao nível de artefato, o Modelo de Projeto tem o objetivo de apresentar todos os detalhamentos necessários para a construção de um Sistema de Conhecimento (SCHREIBER *et al.*, 2002).

Algumas das etapas desta metodologia podem ser auxiliadas e aprimoradas utilizando a modelagem proposta por Bunge (2003) através do modelo CESH (*Composition – Environment – Structure – Mechanism*), o qual proporciona uma visão sistêmica e pode contribuir, significativamente, para a geração, elicitación e estruturação do conhecimento em um domínio específico.

No processo de geração do conhecimento, Calhoun e Starbuck (2003) comentam que esta pode ocorrer por meio da organização de conhecimento anterior em novas formas, da combinação de informações relevantes, ou mesmo de *insights* acerca da aplicação de conhecimento existente em novos contextos.

Gómez-Pérez (1999), dentro do processo de elicitación do conhecimento, destaca a técnica da ontologia como forma eficiente de representar o conhecimento através de um conjunto de termos ordenados hierarquicamente para descrever um domínio que pode ser usado como um esqueleto para uma base de conhecimento.

Uma das principais características da ontologia é a sua excelente capacidade de representar a estrutura organizacional de um grande domínio complexo e raciocinar sobre ela. Porém, sua aplicação é limitada devido à inabilidade de lidar com incertezas (KOLLER e PFEFFER, 1998; COSTA *et al.*, 2009). Esta deficiência pode ser contornada através do cálculo probabilístico que fornece um tratamento adequado às variáveis (YANG e

CALMET, 2005).

As redes Bayesianas, fundamentadas no cálculo probabilístico, são amplamente usadas para trabalhar com incerteza devido à sua excelente expressividade gráfica e poder computacional; entretanto, são inadequadas para representar o conhecimento em domínios grandes e complexos (MAHONEY e LASKEY, 1996).

Portanto, tem-se na ontologia associada às redes Bayesianas uma estrutura adequada complementar que possibilita aproveitar a potencialidade de cada uma delas no apoio ao processo de diagnóstico (YANG e CALMET, 2005).

A construção de um modelo de engenharia do conhecimento, envolvendo os conceitos e técnicas de representação de conhecimento, principalmente pelo uso de ontologias e do cálculo probabilístico, possibilita trabalhar dentro de um domínio do conhecimento equacionando as incertezas de forma a apoiar o processo do diagnóstico.

2. OBJETIVO

Este artigo tem por objetivo apresentar um modelo de engenharia do conhecimento de apoio ao diagnóstico envolvendo incertezas, subjetividade ou falta de informações adequadas que dificultam a realização ou até mesmo a impossibilita.

Também explora o refinamento probabilístico para sugerir informação relevante, a ser analisada através de descoberta baseada na literatura, como um possível novo conhecimento a ser incorporado à base de conhecimento.

Para atender seus objetivos, este trabalho está organizado como segue. Na seção 3, é abordada a EC e a metodologia CommonKADS como forma de estruturar um SBC. A seção 4 apresenta o modelo proposto. Na seção 5, um exemplo didático é construído e os resultados são apresentados e avaliados na seção 6. Na seção 7, é proposta uma discussão sobre a contribuição do modelo. A conclusão e futuros trabalhos são apresentados na seção 8.

3. ENGENHARIA DO CONHECIMENTO

A Engenharia do Conhecimento é uma área que tem por objetivo prover sistemas capazes de efetivar a explicitação e preservação do conhecimento organizacional. Inicialmente tratada como uma subárea da Inteligência Artificial (IA) na construção de SBC para a solução de problemas específicos, a EC transcendeu esta visão ao considerar todo o contexto sistêmico organizacional das atividades intensivas em conhecimento (SCHREIBER *et al.*, 2002).

Assim, a EC deve prover um conjunto de ferramentas que forneçam suporte à GC a partir da formalização e explicitação das atividades intensivas em conhecimento nas organizações (SCHREIBER *et al.*, 2002). Para isto, Struder *et al.* (1998) descrevem como fundamental a aplicação de metodologias, métodos e técnicas voltadas à modelagem do conhecimento e sua consequente representação em sistemas de conhecimento.

Neste contexto, a construção de um SBC deve ser realizada através de uma abordagem estruturada, com uma visão organizacional e sistêmica.

3.1 Engenharia do Conhecimento e a Metodologia CommonKADS

Uma valiosa fonte de princípios e técnicas de modelagem do conhecimento é a metodologia CommonKADS, a qual oferece um conjunto de etapas para o desenvolvimento de Sistemas Baseados em Conhecimento (SCHREIBER *et al.*, 2002). A metodologia é baseada em quatro princípios que fundamentam a Engenharia do Conhecimento:

- A Engenharia do Conhecimento aborda diferentes aspectos dos modelos humanos de conhecimento e não apenas a extração do conhecimento de um especialista;

- A modelagem do conhecimento deve estar focada inicialmente na estrutura conceitual e não nos detalhes de programação dos sistemas de conhecimento;
- O conhecimento tem uma estrutura interna estável, podendo ser categorizado em tipos ou papéis;
- Um projeto de SBC deve ser construído de forma espiral, a partir de uma metodologia que permita este tipo de abordagem.

Além disso, três questões são fundamentais no processo de modelagem de um SBC utilizando-se a metodologia: Por quê? O quê? Como? Para responder a estas questões, a metodologia é organizada em três níveis de modelos (SCHREIBER *et al.*, 2002):

- Nível de contexto: tem por objetivo identificar os elementos fundamentais do ambiente e do contexto do Sistema de Conhecimento;
- Nível conceitual: tem por objetivo explicitar e formalizar o conhecimento, bem como destacar de que modo ocorrem as interações entre os agentes envolvidos em sua utilização;
- Nível de artefato: tem por objetivo apresentar todos os detalhamentos necessários para a construção de um Sistema de Conhecimento.

A figura 1 apresenta os modelos de cada um dos níveis do CommonKADS. No Nível de Contexto, os Modelos de Organização, Tarefa e Agente são estruturados. No Nível de Conceito, são estruturados os Modelos de Conhecimento e Comunicação. No Nível de Artefato, é estruturado o Modelo de Projeto.

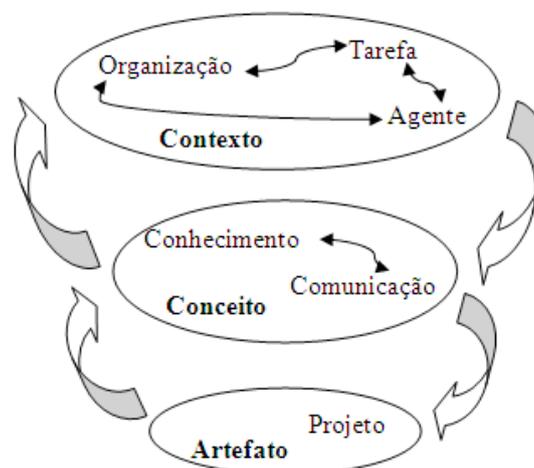


Figura 1 - Metodologia CommonKADS

Fonte: adaptado de SCHREIBER *et al.*, (2002)

O Quadro 1 apresenta uma descrição mais detalhada dos níveis do CommonKADS para a construção de Sistemas Baseado em Conhecimento, bem como sobre os modelos a serem construídos em cada um dos níveis.

Nível	Modelo	Objetivo
Contexto	Organização	Identificar e analisar o contexto organizacional onde o sistema estará inserido.
	Tarefa	Detalhar todas as tarefas envolvidas no sistema, com foco nas atividades intensivas em conhecimento.
	Agente	Identificar os agentes humanos e não humanos e suas responsabilidades no Sistema de Conhecimento.
Conceito	Conhecimento	Apresentar detalhadamente o modelo conceitual de conhecimento do sistema, destacando as tarefas, as inferências e o domínio para sua posterior implementação.
	Comunicação	Modelar a comunicação entre os agentes envolvidos no funcionamento do sistema.
Artefato	Projeto	Definir as especificações técnicas necessárias para a implementação e funcionamento do sistema, com base nas etapas anteriores.

Quadro 1 - Descrição dos modelos do CommonKADS

No trabalho de Lopes *et al.* (2011), um Sistema de Conhecimento foi construído para diagnóstico em Acupuntura. Neste trabalho, foram apresentados o modelo de contexto (organização, tarefas e agentes) e o modelo de conceito (conhecimento e comunicação). A metodologia CommonKADS forneceu uma visão de contexto sistêmico sobre a organização e o problema a ser tratado. Isto a torna menos focada na construção de um artefato, mas sim na elaboração de um modelo de conhecimento que possa gerar resultados de forma a minimizar as incertezas e riscos quanto à implantação de um Sistema de Conhecimento para a organização.

Dentre as planilhas que fazem parte desta metodologia e que compõe o nível de contexto, por exemplo, é apresentada, no quadro 2, a planilha denominada OTA1 - Checklist para Decisão sobre Impactos e Melhorias; ela resume um conjunto de informações que auxiliam o processo decisório de implantação de um SBC para uma clínica que pratica Acupuntura, tornando-o mais seguro, eficaz e abrangente às necessidades de quem irá utilizá-lo.

Ao nível de conceito, o modelo de conhecimento da metodologia CommonKADS é próprio para especificar tarefas intensivas de conhecimento e explica em detalhes o tipo e a estrutura do conhecimento usado na execução de uma tarefa. O modelo é composto por três componentes, em que cada um deles captura uma parte da estrutura do conhecimento:

- conhecimento de tarefa: define o objetivo da tarefa intensiva em conhecimento (meta), podendo separar em pequenas tarefas (decomposição) e controlá-las (controle). Responde o que será feito e como será feito (Diagnóstico);
- conhecimento de inferência: contém as inferências básicas responsáveis pelas tarefas de raciocínio (Cobrir, Especificar, Comparar, entre outras), os papéis de conhecimento que conectam o conhecimento de domínio às inferências (Reclamação, Hipótese, Esperado, Encontrado, Resultado, entre outras) e às funções de transferência (Obter, entre outras);
- conhecimento de domínio: formado pela estrutura de domínio envolvendo os conceitos, as relações e as regras em conjunto com a base de conhecimento (Sintomas, Exames, entre outras).

Modelo de Organização, Tarefa e Agente	OTA1 – Checklist para Decisão sobre Impactos e Melhorias
Impactos e mudanças na organização	<p>Estrutura: não há impactos e mudanças necessárias na estrutura da organização.</p> <p>Processos: diagnóstico será realizado apoiado por um Sistema de Conhecimento, proporcionando maior eficácia e qualidade no processo.</p> <p>Recursos: será necessário a utilização de um Sistema de Conhecimento disponível em um computador.</p> <p>Pessoas: o acupunturista terá a seu dispor um Sistema de Conhecimento, não existindo outros impactos ou mudanças para as pessoas na organização.</p> <p>Conhecimento: o conhecimento, que anteriormente estava apenas no domínio do especialista, passa a estar formalizado e explicitado através de um Sistema de Conhecimento e pode ser compartilhado entre especialistas, profissionais e estudantes da área.</p> <p>Cultura e poder: o Sistema de Conhecimento pode impactar positivamente a organização na medida em que ele possibilitará gerar novos conhecimentos sobre os pacientes e sobre os processos de diagnóstico e tratamento. Permitirá maior compartilhamento de informações, possibilitando que um paciente possa ser acompanhado por mais de um profissional.</p>
Impactos e mudanças na tarefa e nos agentes	<p>O acupunturista, de posse de um Sistema de Conhecimento que auxilie no diagnóstico, poderá tomar decisões mais rapidamente e com maior precisão, reduzindo o grau de incerteza e o tempo de execução da tarefa. Com isso, com o mesmo tempo que ele tem disponível para atendimento na situação atual, ele poderá atender um maior número de pacientes a partir da implantação da solução proposta. O sistema também pode ser utilizado como fonte de conhecimento e formação para novos profissionais que desejam atuar na área. Para os pacientes, espera-se que o tratamento torne-se mais eficaz, melhorando mais rapidamente sua qualidade de vida.</p>
Atitudes e compromissos	<p>O acupunturista deve se comprometer a utilizar o Sistema de Conhecimento como fonte primordial para o seu processo decisório, fazendo sugestões e críticas para revisão e aperfeiçoamento do mesmo.</p>
Ações propostas	<p>Melhoria: implantar um Sistema de Conhecimento que auxilie o diagnóstico para o tratamento de um paciente de acupuntura.</p> <p>Acompanhamento: para a implantação adequada do sistema, é fundamental a validação do sistema com especialistas e treinamento para utilização.</p> <p>Resultados esperados: espera-se agilizar o processo de tratamento, gerando, com isto, ganhos para a organização já destacados anteriormente. Os custos de implantação são relativamente baixos em função da disponibilidade técnica e tecnológica que viabilizam a construção do Sistema de Conhecimento proposto.</p>

Quadro 2 - OTA1: Checklist para Decisão sobre Impactos e Melhorias

A partir dos *templates* apresentados por Schreiber *et al.* (2002) para as tarefas intensivas em conhecimento, foi selecionado o *template* para o diagnóstico. A Figura 2 apresenta o modelo de tarefa completo e anotado para o diagnóstico a partir de geração e teste. Através da figura, é possível ter uma compreensão global da execução da tarefa, considerando as inferências propostas.

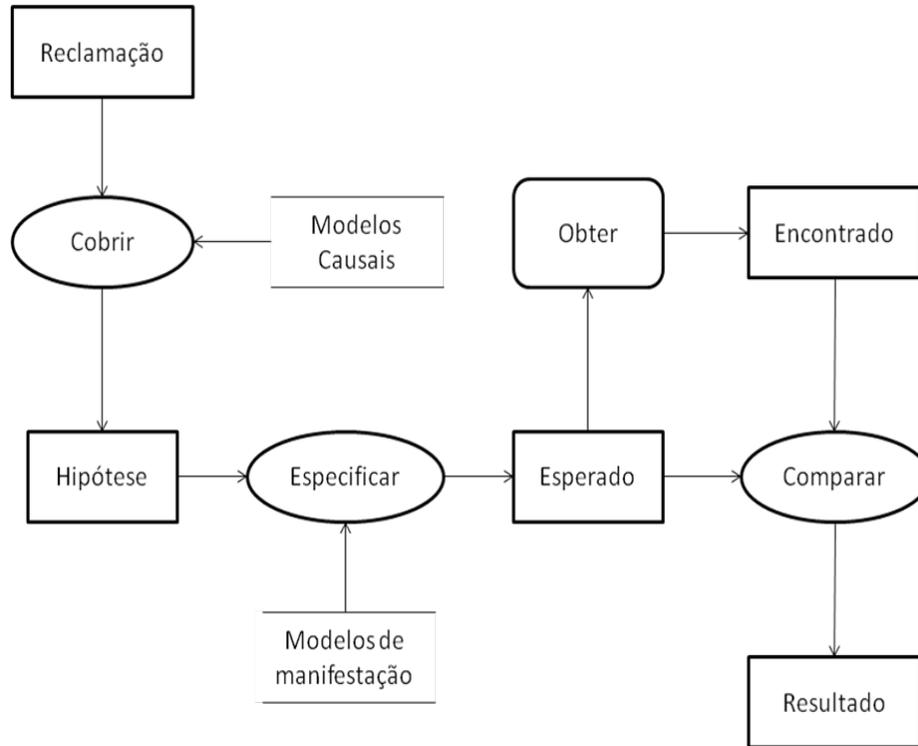


Figura 2 – Modelo de Tarefa para o Diagnóstico
 Fonte: adaptado de SCHREIBER et al. (2002)

Diante deste *template*, um modelo de engenharia do conhecimento será proposto tendo em vista o estudo de viabilidade e os impactos e melhorias que o mesmo poderá proporcionar.

4. MODELO PROPOSTO

O modelo de engenharia do conhecimento proposto, apresentado na figura 3, baseado no *template* do modelo de tarefa para o diagnóstico, procura, a partir de adaptações feitas, apoiar o processo de diagnóstico fornecendo resultados probabilísticos, bem como destacar informações relevantes com a finalidade de gerar novos conhecimentos.

Este modelo tem como elementos fundamentais a integração entre:

- Ontologia - estrutura para a base de conhecimento;
- Cálculo probabilístico - realiza as relações de causa e efeito; e
- Sistema de descoberta baseado na literatura - análise de novos conhecimentos.

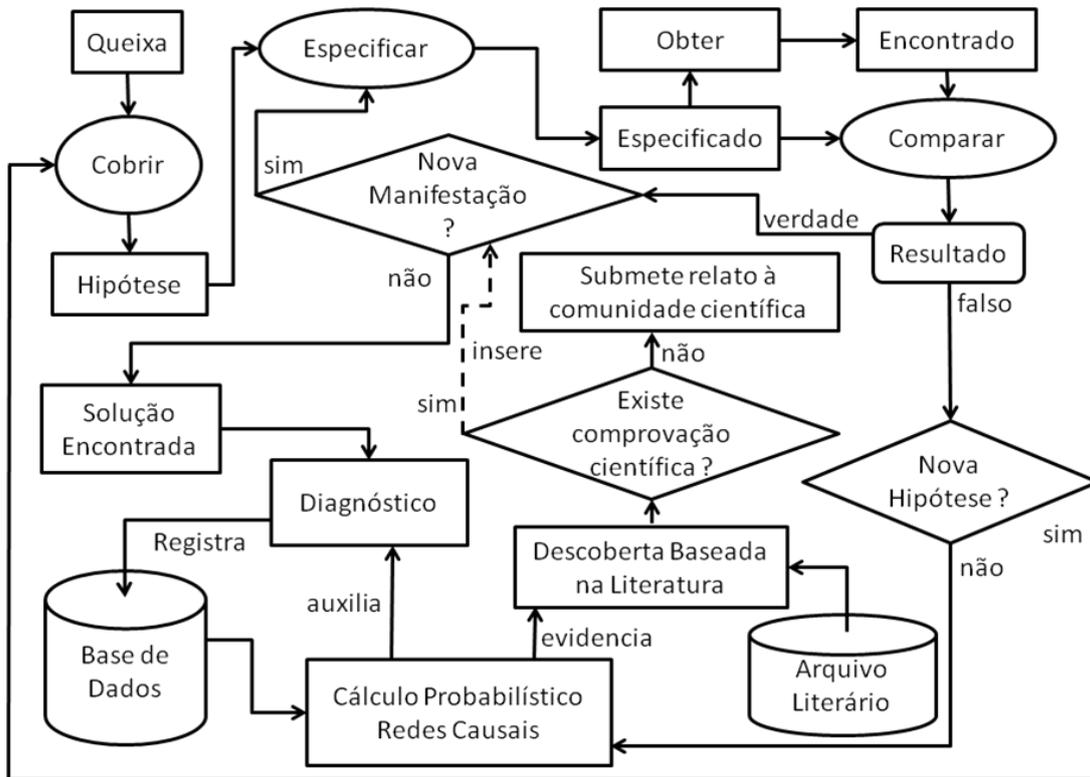


Figura 3 – Modelo proposto para Tarefa de Diagnóstico

Fonte: adaptado de SCHREIBER et al. (2002)

4.1 Arquitetura do modelo proposto

A arquitetura do Modelo proposto é apresentada pela figura 4, sendo as etapas (tarefa, base de conhecimento, descoberta do conhecimento e fonte de dados/informações) descritas na sequência.

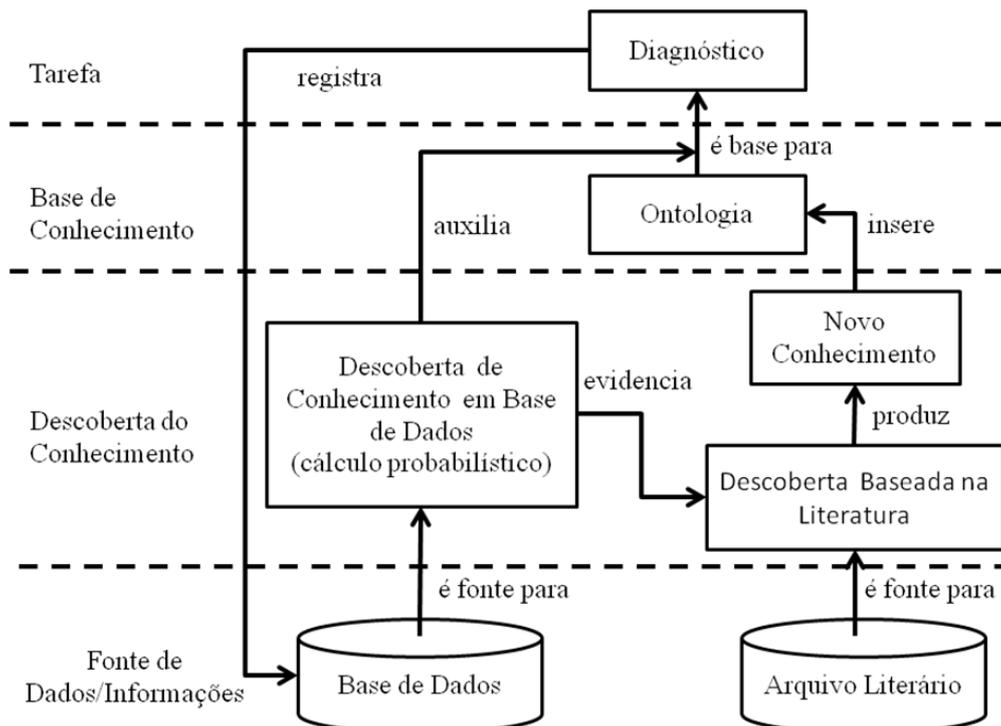


Figura 4 – Arquitetura do modelo proposto

- **Tarefa**

A tarefa realizada é o diagnóstico. Segundo Schreiber *et al.* (2002), trata-se de uma tarefa analítica cujo objetivo é encontrar a resposta ao mau funcionamento do sistema. Recebe as informações provenientes da base de conhecimento e, através das inferências, procura chegar ao resultado. Na sequência, os dados e informações obtidos nos diagnósticos realizados são transferidos para “alimentar” uma base de dados.

- **Base de Conhecimento**

A Base de Conhecimento tem como estrutura a ontologia de um domínio específico. Esta ontologia é construída sob a visão da modelagem CISM (*Composition – Environment – Structure – Mechanism*) Bunge (2003). De acordo com esta proposta, tem-se uma estrutura sistêmica descritiva e abrangente para formalizar a base de conhecimento.

Portanto, o conhecimento bem estruturado na ontologia provê as informações para que a tarefa de diagnóstico possa ser executada. Permite também a inserção de novos conhecimentos em sua estrutura com a finalidade de aprimorar o processo de diagnóstico.

- **Descoberta do Conhecimento**

A Descoberta do Conhecimento é responsável pelos métodos e técnicas para a extração do conhecimento. O cálculo probabilístico tem duas funções, uma é auxiliar a realização do diagnóstico através das inferências e a outra é destacar evidências que serão analisadas pelo processo de Descoberta Baseada na Literatura (LBD) através de predicções semânticas e regras de associação (HRISTOVSKI *et al.*, 2006) como um possível novo conhecimento a ser anexado à Base de Conhecimento.

- **Fonte de Dados/Informações**

A Base de Dados e o Arquivo Literário compõem a fonte de pesquisa para o sistema probabilístico e o sistema de descoberta baseado na literatura, respectivamente. A Base de Dados recebe e armazena os dados/informações dos diagnósticos realizados, servindo como fonte para a realização do cálculo probabilístico. Já o Arquivo Literário fornece as informações científicas necessárias para que um novo conhecimento possa ser identificado.

5. EXEMPLO DIDÁTICO

Nessa seção, apresenta-se um exemplo didático, na área da saúde, que aplica o modelo proposto à tarefa de diagnóstico. Possui como base o uso de uma ontologia para representar o conhecimento de domínio e o raciocínio sobre essa estrutura, assim como análises probabilísticas para auxiliar no processo de tomada de decisão.

Uma evidência, realçada pelo refinamento probabilístico, serve de insumo às análises em literatura científica e, dependendo do resultado obtido, pode gerar um novo conhecimento a ser agregado à base de conhecimento.

5.1 Metodologia para a Construção do Exemplo Didático

A metodologia para a construção do exemplo é como segue:

- 1) Os especialistas da área definem a ontologia de consenso para obter as classes, subclasse, relações, axiomas e instâncias para um determinado domínio específico que possibilitem o processo de raciocínio automático;
- 2) Uma estrutura probabilística formada pelas classes e subclasses da ontologia e alimentada pela base de dados de diagnósticos realizados auxilia na tomada de decisão bem como indica possíveis evidências como candidatas à geração de conhecimento novo;

3) O sistema de descoberta baseado na literatura analisa as evidências realçadas pelo refinamento probabilístico, de forma que tenha o respaldo científico para ingressar na estrutura de inferência da Base de Conhecimento.

Para a realização destas tarefas, são utilizadas as ferramentas Protégé (PROTÉGÉ, 2007), para a estruturação da ontologia; o software Netica (NORSYS, 2008), que desenvolve a rede Bayesiana para o cálculo probabilístico e a ferramenta Bitola, que providencia a pesquisa literária através das predicções semânticas e regras de associação.

5.2 Construção da Ontologia

Os procedimentos de Aquisição do Conhecimento, adquiridos através de consulta a especialistas, manuais e livros e a Representação do Conhecimento permitem a formalização de uma base de conhecimento formada pelas classes e subclasses da ontologia que compõem o modelo hierárquico da mesma.

A ontologia foi expandida através do modelo CESM, o qual proporcionou, dentro de uma visão sistêmica, a explicitação do conhecimento através de uma modelagem mais descritiva e abrangente para os elementos que compõem um domínio específico. A figura 5 ilustra o diagrama CESM para o sistema de diagnóstico, destacando:

- 1) os componentes – coleção de todas as partes do sistema – é composto pelo paciente e seus sintomas e sinais, os resultados dos exames realizados em laboratório e o profissional que executa o diagnóstico;
- 2) o ambiente – coleção dos itens que não pertencem ao sistema, mas atuam ou sofrem a ação por algum ou todos os componentes do sistema - está representado, por exemplo, pelas atividades profissionais, as atividades físicas, os tipos de alimentação e os sentimentos ou emoções, sendo estes alguns dos elementos escolhidos que podem interferir diretamente no resultado do diagnóstico;
- 3) a estrutura no modelo CESM é definida como a coleção de ligações/relações entre componentes e entre esses e itens do ambiente. Como exemplo, pode-se citar as ligações da forma “O paciente tem sintoma” ou “O paciente faz exercícios” ou ainda “A ansiedade é uma emoção do paciente”; e
- 4) o mecanismo - coleção de processos que geram mudanças qualitativas no sistema – pode ser interpretado como as restrições que definem as limitações nas relações entre os componentes e o ambiente e são usados para realizar o processo de consulta ou raciocínio sobre o sistema.

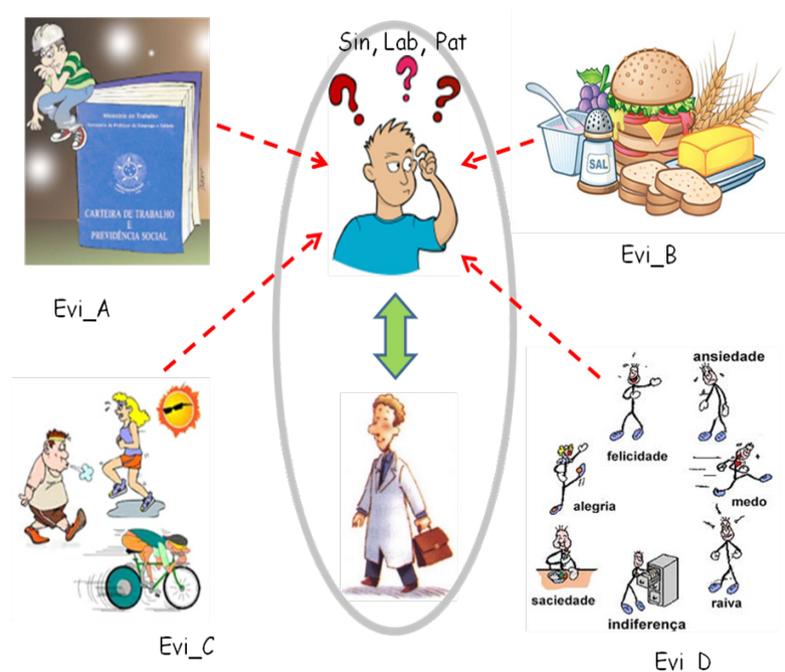


Figura 5- Diagrama CESM para o processo de diagnóstico

A figura 6 apresenta a hierarquia da ontologia proposta, geradas pela ferramenta Protégé em que são consideradas as classes: Paciente, Sintoma, Laboratório e Evidência.

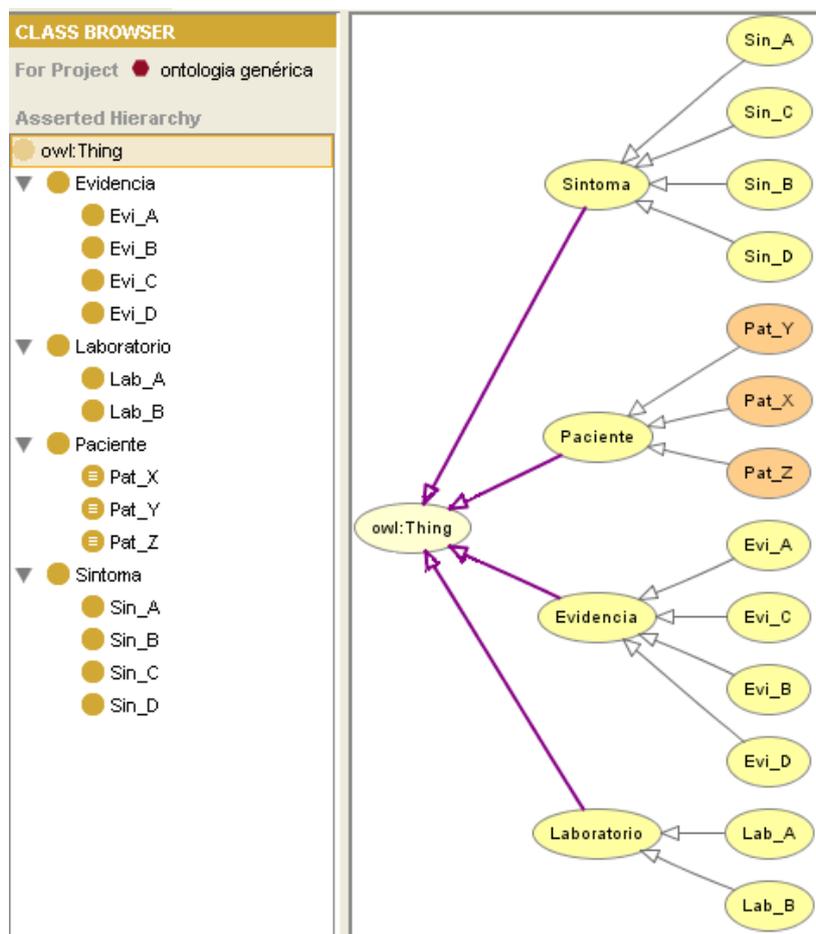


Figura 6 – Classes e subclasses da ontologia

Para a classe Paciente, que está acometido por uma enfermidade, foram definidas as subclasses formadas pelas patologias: Pat_X, Pat_Y e Pat_Z.

Para a classe Sintoma, que representa as manifestações da doença, foram definidas as subclasses: Sin_A, Sin_B, Sin_C e Sin_D.

Para a classe Laboratório, que representa os exames solicitados pelo especialista, foram definidas as subclasses: Lab_A e Lab_B.

Para a classe Evidência, que representa algum elemento externo ao paciente e que pode interferir em sua patologia, foram definidas as subclasses: Evi_A, Evi_B, Evi_C e Evi_D.

5.2.1 Relações entre Classes

As relações existentes na ontologia são necessárias para estabelecer os diferentes relacionamentos entre as classes e subclasses. A figura 7 apresenta as relações existentes entre as classes Paciente, Sintoma, Laboratório e Evidência. A classe Paciente tem um relacionamento com a classe Sintoma do tipo “temSintoma”; assim, pode-se concluir que “Paciente tem Sintoma”. A relação “temSintoma” possui a relação inversa “ehManifestado”, da forma “Sintoma é manifestado no Paciente”.

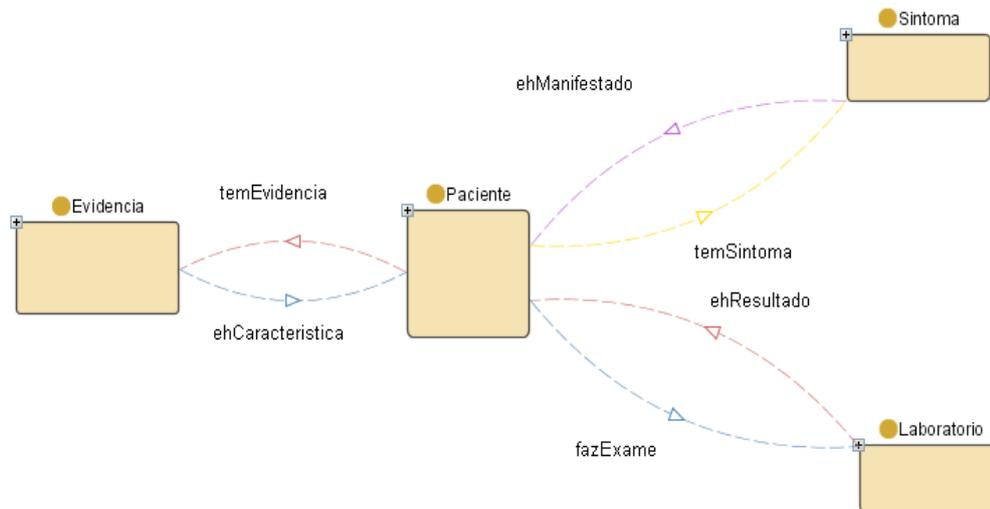


Figura 7 – Relações entre Classes

5.2.2 Regras

Após construir a hierarquia de classes com suas respectivas relações, foram definidas as restrições, que são as regras que restringem os relacionamentos e permitem estabelecer se um indivíduo pertence ou não a uma determinada classe. A figura 8 exibe a restrição formada pelos sintomas Sin_A, Sin_C e exame de laboratório Lab_B, necessários para que um paciente seja diagnosticado com a patologia Pat_Y.

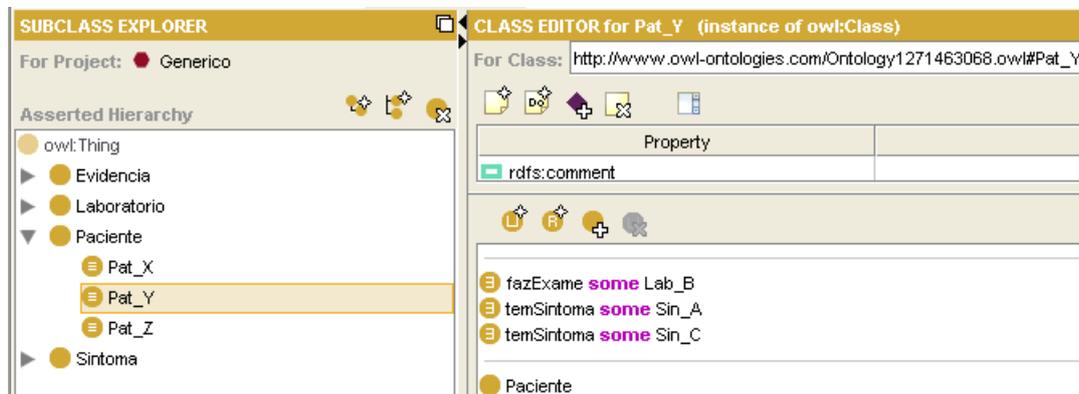


Figura 8 – Restrições de Classe

Para a patologia Pat_X, foram definidas as seguintes restrições: Sin_A, Sin_B, Sin_C e LT_A. Para a patologia Pat_Z os sintomas Sin_B e Sin_D e o exame de laboratório Lab_B compõem as restrições.

5.2.3 Instâncias

A construção das classes torna possível a criação das instâncias da ontologia. As instâncias representam indivíduos específicos de uma determinada classe. Os pacientes (11) denotados por P01 até P11, devidamente registrados clinicamente, formam as instâncias que são submetidas ao processo de inferência.

A figura 9 apresenta a relação dos pacientes e destaca o paciente P01 que tem os sintomas Sin_a, Sin_B e Sin_C, com exame de laboratório Lab_A e as evidências Evi_A, Evi_C e Evi_D.

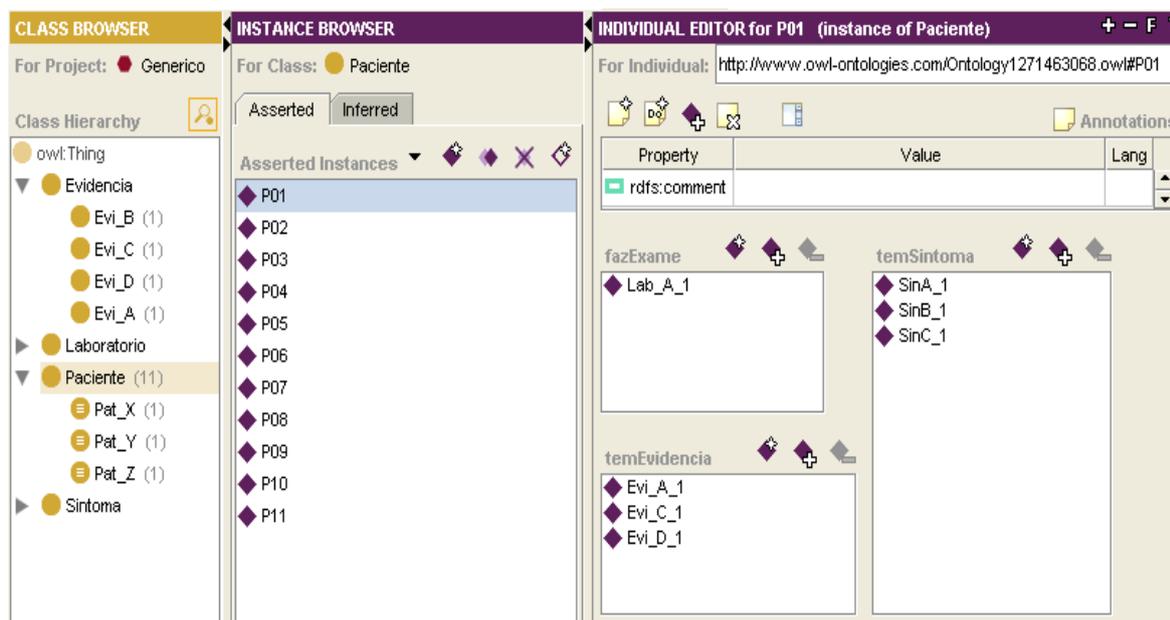


Figura 9 – Exemplo de instância

5.2.4 Inferências

A inferência, ou processo de raciocínio, procura “classificar” o paciente dentro de uma determinada patologia, ou seja, determina a falta que está gerando a reclamação ou o motivo pelo mau funcionamento do sistema.

O resultado do processo de inferência, utilizando o raciocinador pellet 1.4, é apresentado na figura 10. Nesse processo, foram diagnosticados com a patologia Pat_X os pacientes P1 e P9. A patologia Pat_Y está presente nos pacientes P2, P4, P7, P8 e P10 e diagnosticados com a patologia Pat_Z os pacientes P3, P5 e P11.

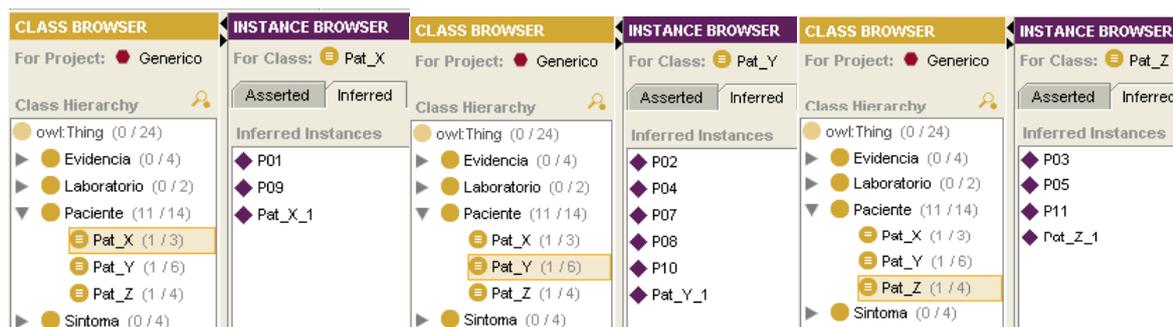


Figura 10 – Inferência no Protégé

Nota-se que, dos onze pacientes, somente P06 não foi diagnosticado. Pelos registros clínicos de P06 apresentados na figura 11, Sin_A, Lab_A, Evi_A, Evi_B e Evi_D, não foi possível indicar um diagnóstico para ele.

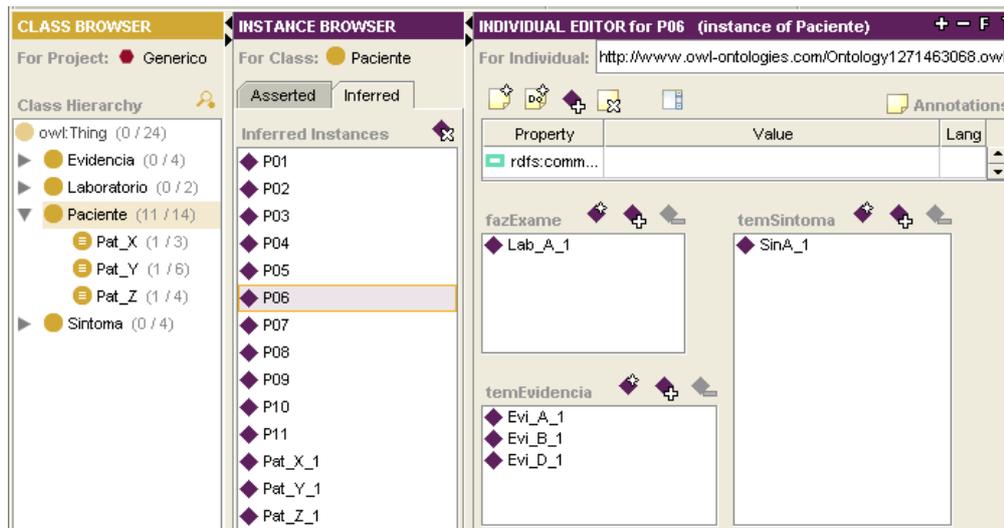


Figura 11 – Registro clínico do Paciente P6

Conclui-se, portanto, que, para P06, os dados informados ao sistema são insuficientes ou incertos, pois a subjetividade nas observações ou a falta de informação adequada ocasionou esta ocorrência. Por outro lado, pode estar havendo a necessidade de uma atualização nas restrições que compõem as patologias de forma que novos conhecimentos possam contribuir para um aprimoramento no processo de diagnóstico.

Diante desta situação, é elaborado para P06 um diagnóstico probabilístico que tem como suporte a base de dados gerada pelos diagnósticos já realizados.

5.3 Base de Dados

Através dos registros clínicos dos dez pacientes que foram diagnosticados e obtiveram sucesso em seus tratamentos, organizou-se a base de dados apresentada na tabela 1, que serve como fonte de dados para o cálculo probabilístico.

Tabela 1 – Base de Dados dos pacientes tratado

Diag_F	D_Sin	D_Evi	Sin_A	Sin_B	Sin_C	Sin_D	Lab_A	Lab_B	Evi_A	Evi_B	Evi_C	Evi_D
Pat_X	Pat_X	Pat_X	sim	Sim	sim	nao	sim	nao	sim	nao	sim	sim
Pat_X	Pat_X	Pat_X	sim	Sim	sim	nao	sim	nao	sim	nao	sim	sim
Pat_Y	Pat_Y	Pat_Y	sim	Nao	sim	nao	nao	sim	nao	sim	sim	nao
Pat_Y	Pat_Y	Pat_Y	sim	Nao	sim	nao	sim	sim	nao	sim	sim	nao
Pat_Y	Pat_Y	Pat_Y	sim	Nao	sim	nao	sim	sim	sim	sim	nao	sim
Pat_Y	Pat_Y	Pat_Y	sim	Nao	sim	nao	nao	sim	nao	sim	nao	sim
Pat_Y	Pat_Y	Pat_Y	sim	Nao	sim	sim	sim	sim	nao	sim	nao	sim
Pat_Z	Pat_Z	Pat_Z	nao	Sim	nao	sim	nao	sim	nao	sim	sim	nao
Pat_Z	Pat_Z	Pat_Z	nao	Sim								
Pat_Z	Pat_Z	Pat_Z	nao	Sim	nao	sim	nao	sim	sim	nao	sim	sim

Nesta tabela, estão sendo registrados os sintomas (Sin_A, Sin_B, Sin_C e Sin_D), os exames de laboratório (Lab_A e Lab_B), as evidências (Evi_A, Evi_B, Evi_C e Evi_D) e o resultado do diagnóstico (Pat_X, Pat_Y e Pat_Z) de cada paciente tratado.

6. RESULTADOS

Para realizar a análise probabilística, foi construído um protótipo simplificado no software Netica (NORSYS, 2008), utilizando redes Bayesianas que são representações que mapeiam conhecimento a partir de um modelo probabilístico, expressando relações de causas e efeitos através da associação de probabilidade em um determinado domínio do conhecimento (RUSSEL e NORVIG, 1995).

A figura 12 apresenta a rede Bayesiana construída a partir da base de dados. Os sintomas (Sin_A, Sin_B, Sin_C, Sin_D) e os testes de laboratório (Lab_A, Lab_B) possibilitam a formação do resultado referente ao diagnóstico D_Sin. As evidências (Evi_A, Evi_B, Evi_C, Evi_D) possibilitam a formação do resultado referente ao diagnóstico D_Evi.

O diagnóstico final (Diag_F) é uma composição entre o D_Sin e D_Evi, que leva em consideração o grau de significância de cada um para a formação do resultado final.

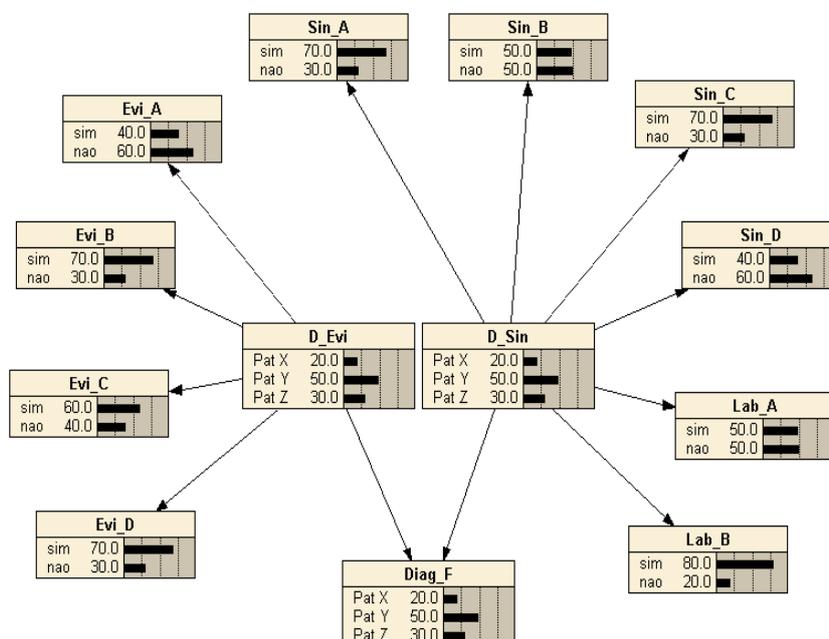


Figura 12 – Rede Bayesiana

A tabela 2 apresenta as porcentagens atribuídas a cada uma das partes acima citadas. Ao diagnóstico proveniente dos sintomas e teste de laboratório foi atribuído, aleatoriamente, um peso de 95% enquanto que ao diagnóstico proveniente das evidências foi atribuído, também aleatoriamente, um peso 5%.

Tabela 2 – Porcentagens atribuídas ao sistema em relação às evidências e aos sintomas

D_Evi	D_Sin	Pat X	Pat Y	Pat Z
Pat X	Pat X	100.00	0.000	0.000
Pat X	Pat Y	5.000	95.000	0.000
Pat X	Pat Z	5.000	0.000	95.000
Pat Y	Pat X	95.000	5.000	0.000
Pat Y	Pat Y	0.000	100.00	0.000
Pat Y	Pat Z	0.000	5.000	95.000
Pat Z	Pat X	95.000	0.000	5.000
Pat Z	Pat Y	0.000	95.000	5.000
Pat Z	Pat Z	0.000	0.000	100.00

6.1 Diagnóstico Probabilístico

Depois de concluída a estruturação da rede Bayesiana, foi possível inserir os dados referentes ao paciente P06. A figura 13 exibe a rede instanciada.

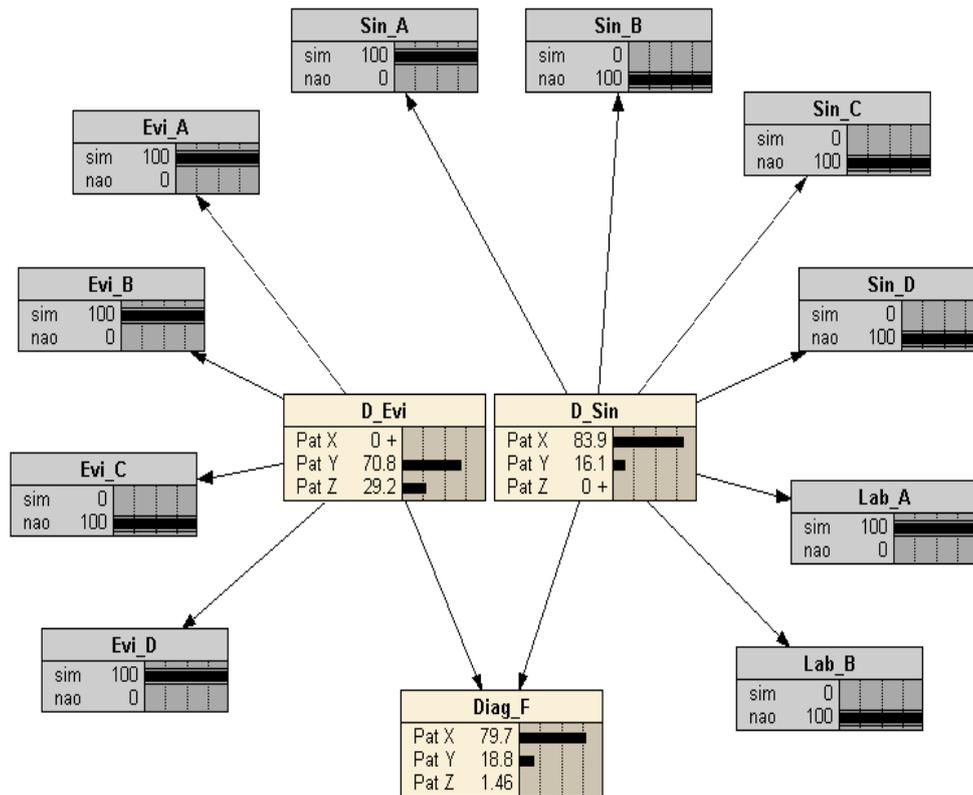


Figura 13 – Rede Bayesiana instanciada para P6

De acordo com os resultados obtidos, tem-se o diagnóstico probabilístico para P06 que atribui um valor de 79,7% para a patologia Pat_X, 18,8% para a patologia Pat_Y e 1,46% para a patologia Pat_Z.

Portanto, com 79,7% de probabilidade de acerto, a Pat_X é a melhor indicação para o tratamento do paciente P06.

6.2 Novo Conhecimento

A base de dados armazena os registros clínicos dos pacientes tratados. Muitas informações relevantes podem surgir com o auxílio do cálculo probabilístico na manipulação destes dados. Uma evidência realçada pode tornar-se um novo conhecimento quando confrontada com o arquivo literário através do sistema de descoberta baseada na literatura.

Por exemplo: na figura 14, está sendo instanciada a patologia Pat_Y. Em destaque, os sintomas Sin_A e Sin_C e teste de laboratório Lab_B referem-se às restrições impostas à patologia. Observa-se também que a evidência “Evi_B” aparece em 100% dos pacientes com esta patologia. É uma informação que, ao ser confrontado com a literatura científica, pode originar um novo conhecimento.

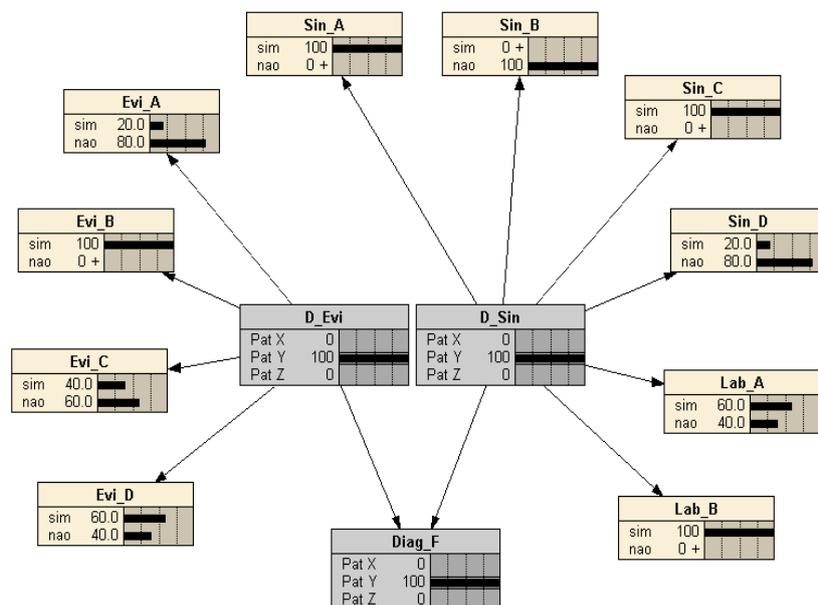


Figura 14 – Rede Bayesiana instanciada para Pat_Y

Diversas ferramentas de pesquisa apoiam este processo de descoberta. Entre elas, o software Bitola (HRISTOVSKI *et al.*, 2001), que é um sistema interativo e funciona como suporte de descoberta biomédica baseado na literatura. A proposta do sistema é ajudar os pesquisadores biomédicos a fazer novas descobertas através das predicações semânticas e regras de associação.

Utilizando dois sistemas de processamento de linguagem natural, BioMedLEE e SemRep, as predicações são produzidas através de aplicações combinadas (HRISTOVSKI *et al.*, 2006). A figura 15 exemplifica uma consulta feita no software Bitola procurando relacionar a patologia “Pat_Y” como tipo semântico “doença” com a evidência “Evi_B” como sendo um tipo de alimento.

BITOLA - Biomedical Discovery Support System: Closed Discovery

(Program authors: Dimitar Hristovski, Borut Peterlin)

Find Starting Concept X

Concept: Pat_Y

Semantic Types: doença

Select a starting concept X from the list by clicking on its name:

Find End Concept Z

Concept: Evi_B

Semantic Types: alimento

Figura 15 – Consulta ao software BITOLA

O resultado desta busca é dado através da relação de vários artigos científicos que relacionam a patologia “Pat_Y” com a evidência “Evi_B”. Estes artigos selecionados são apresentados ao profissional para análise e, diante da confirmação de um novo conhecimento, pode-se propor uma atualização na base de conhecimento do domínio específico.

7. DISCUSSÃO

Este trabalho teve como objetivo principal o desenvolvimento de um modelo de engenharia do conhecimento capaz de contribuir para a sistematização do processo de diagnóstico, que envolve incertezas e subjetividades. Assim, questionam-se e discutem-se quais contribuições de fato foram estas.

Os modelos baseados em ontologias são muito importantes para representação sistemática e formal do conhecimento, com completude e consistência suficientes para cobrir determinado domínio. Este conhecimento representado, quando aliado a outras técnicas, como as redes Bayesianas e a descoberta baseada em literatura, pode ajudar o profissional no processo de diagnóstico, apresentando melhores opções de conclusão diagnóstica e orientando nas perguntas que podem ser feitas durante o processo.

Além disto, cabe salientar a importância do modelo para a descoberta de novos conhecimentos, pois os profissionais, muitas vezes, ficam restritos a determinadas estratégias de raciocínio e não conseguem visualizar novas alternativas para a solução de um caso. Desta forma, a existência de um modelo sistêmico que vislumbre novas estratégias de ação para se chegar a um diagnóstico e que gere novos conhecimentos e o apresente ao profissional pode ser de grande importância inclusive para capacitação e atualização do mesmo. Ainda, ao descobrir novos conhecimentos, o modelo proposto pode melhorar e favorecer a habilidade de diagnóstico de um profissional que não seja tão especialista ou experiente no domínio em questão.

O uso de associação e o estabelecimento de correlações entre um novo diagnóstico e aqueles já conhecidos é semelhante à forma de raciocínio de muitos especialistas, comparando algo novo com algo aprendido ou com casos antigos. Assim, a descoberta baseada na literatura fornece a capacidade de se comparar um caso novo até mesmo com casos ou informações que não se teve a oportunidade de vivenciar em sua rotina, aumentando o conhecimento e melhorando a capacidade de decisão do profissional.

Outra importante contribuição do modelo está na possibilidade de sistematização e registro das atividades de diagnóstico de um profissional. A rotina faz com que se mantenham poucos registros da atividade cotidiana; isto corrobora a necessidade de se extrair e arquivar dados sobre diagnóstico de forma sistemática e acessível para contribuir com novos diagnósticos e gerar novos conhecimentos. Desta forma, possibilitará não somente o “re-uso” das informações, mas também o seu compartilhamento.

8. CONCLUSÕES E SUGESTÕES PARA TRABALHOS FUTUROS

O artigo apresenta um modelo de Engenharia do Conhecimento baseado em ontologia e no cálculo probabilístico que possibilita, além de diagnósticos probabilísticos, a criação de novos conhecimentos para apoio ao diagnóstico, ajudando a decisão humana.

Com a análise da problemática envolvendo o processo de diagnóstico, observam-se oportunidades de aprimorar os modelos correlatos, recentemente desenvolvidos, que procuram auxiliar esta tarefa intensiva em conhecimento. Estes trabalhos analisados, fundamentados em ontologias, mostraram eficiência na manipulação das incertezas envolvidas no processo de diagnóstico aplicando técnicas apropriadas (lógica fuzzy, redes neurais, raciocínio baseado em casos). Entretanto, verificou-se a deficiência na geração de novos conhecimentos, a fim de atualizar a base de conhecimento.

Assim, o problema enfrentado foi como trabalhar de forma adequada com as informações e dados obtidos pelos diagnósticos realizados, de forma que novos conhecimentos possam ser gerados e contribuam para o aprimoramento do diagnóstico.

O trabalho inicial foi identificar na literatura, dentro da visão da nova Engenharia, Conhecimento, metodologias, métodos e técnicas para a elaboração de um modelo que atenda à deficiência na geração de novos conhecimentos. Os elementos encontrados e utilizados foram: metodologia CommonKADS, Ontologia, Cálculo Probabilístico, modelagem CEM e Sistemas de Descoberta baseados na Literatura.

A metodologia CommonKADS forneceu as etapas para a construção de um modelo de conhecimento. Esta metodologia demonstrou ser adequada na medida em que apresenta uma visão de contexto sistêmico sobre a organização e o problema a ser tratado, tornando-a focada não tanto na construção de um artefato, mas sim na construção de um modelo de conhecimento onde o ser humano é o ator principal.

Tem-se, no uso de Ontologia aliado à modelagem CESM, uma integração que permite a formalização e sistematização do conhecimento com completude e consistência suficientes para cobrir determinado domínio possibilitando ainda envolver elementos evocados pela experiência e observação do profissional.

O Cálculo Probabilístico, no trato às incertezas da base de conhecimento, atua de forma complementar ao processo de diagnóstico, fornecendo diagnósticos probabilísticos bem como destaca informações relevantes para a geração de possíveis novos conhecimentos. O diagnóstico, na forma probabilística, parece ser a maneira mais adequada de apresentar resultados em domínios onde a subjetividade e incertezas fazem parte do processo.

Os Sistemas de Descoberta Baseados na Literatura trazem como inovação as predicções semânticas e as regras de associação. Por intermédio destas técnicas, artigos científicos envolvendo os elementos de pesquisa são apresentados para análise. A utilização de tais sistemas tem a importância de levar ao profissional um material útil e seletivo, proveniente de diversas especialidades, evitando leituras desnecessárias e desgastantes.

Assim, tem-se um modelo de Engenharia do Conhecimento que emprega metodologia, métodos e técnicas disponíveis e possibilita, através da integração entre eles, construir um Sistema de Conhecimento que minimiza os riscos de implantação e, com sua utilização, fornece diagnósticos probabilísticos bem como propõe possíveis novos conhecimentos para apoio ao diagnóstico.

Como sugestões para trabalhos futuros de pesquisa que possam ser embasados no modelo de Engenharia do Conhecimento proposto, indicam-se as seguintes oportunidades de aprimoramento:

- Transformação da base de dados categóricos para uma base de dados contínuos, de forma que a precisão das informações também seja considerada (sistema híbrido difuso – probabilístico) ou ainda o uso da lógica paraconsistente no tratamento das incertezas envolvendo contradições;
- A base de dados também pode ser substituída por uma base de casos possibilitando, através de correlações de semelhança, a construção de um modelo mais robusto;
- Utilização dos dados e informações de diagnósticos realizados que não foram bem sucedidos; estes elementos podem e devem ser aproveitados para a geração de novos conhecimentos;
- Construção de uma rede social para compartilhar informações úteis dentro de um domínio específico, envolvendo profissionais de diferentes especialidades e a comunidade em geral;
- Implementação deste modelo, com interface amigável, utilizando técnicas de raciocínio incremental (PARSIA, HALASCHEK-WIENER e SIRIN, 2006; GRAU *et al.* 2007) e extensões à linguagem OWL, tais como a OntoBayes (YANG e CALMET, 2005), PR-OWL (COSTA, LASKEY e LASKEY, 2005), que facilitem a integração automática das ferramentas necessárias para a realização do diagnóstico, bem como a pesquisa em novos conhecimentos;
- Realização de estudo de caso para outras áreas do conhecimento, por exemplo, área jurídica tendo como arquivo literário as jurisprudências existentes, entre outras.

9. REFERÊNCIA

BUNGE, M. **Emergence and convergence: Qualitative novelty and the unity of knowledge**. [S.l.]: University of Toronto Press, 330 p., 2003.

CALHOUN, M. A.; STARBUCK, W. H. Barriers do creating knowledge. In. EASTERBY-SMITH, M.; LYLES, M.. **Handbook of organizational learning and knowledge management**, Edited by: Easterby

M, Lyles MA. p. 473-492, 2003.

COSTA, P.C.G.; LASKEY, K.B.; CHANG, K. PROGNOS: Applying Probabilistic Ontologies to Distributed Predictive Situation Assessment in Naval Operations. **Proceedings of the Fourteenth International Command and Control Research and Technology Conference (ICCRTS 2009)**. Washington, D.C., USA, p. 15-17, 2009.

COSTA, P.C.G.; LASKEY, K.B.; LASKEY, K.J. PR-OWL: a Bayesian ontology language for the Semantic Web, **Proceedings URSW-2005**, Galway, Ireland, 2005, p. 23–33.

FENZA, G.; LOIA, V.; SENATORE, S. Fuzzy modeling of medical diagnosis through a service-oriented architecture. IEEE, **Annual Conference of the North American Fuzzy Information Processing Society – NAFIPS**, p. 1-6, 2008.

GARCÍA-CRESPO, A. *et al.* ODDIN: Ontology-driven differential diagnosis based on logical inference and probabilistic refinements. **Expert Systems with Applications. Madrid**, p. 2621-2628, 2010.

GÓMEZ-PÉREZ, A. Ontological engineering: a state of the art. **British Computer Society**, v. 2, p. 33-43, 1999.

GRAU, B. C.; HALASCHEK-WIENER, C.; KAZAKOV, Y. History matters: Incremental ontology reasoning using modules. In. **Proceedings of the 6th International Semantic Web Conference (ISWC 2007)**, Busan, South Korea, Springer, v. 4825, p. 183-196, 2007.

HRISTOVSKI, D. *et al.* Supporting discovery in medicine by association rule mining in Medline and UMLS. In **Proceedings of MedInfo Conference**, London, England, v.10, n. 2, p.1344–1348, 2001.

HRISTOVSKI, D. *et al.* Exploiting Semantic Relations for Literature-Based Discovery. **AMIA 2006 Symposium Proceedings**. p. 349-353. 2006.

KOLLER, D.; PFEFFER, A. Probabilistic frame-based systems. In **Proceedings of the 15th. National Conference on Artificial Intelligence (AAAI'98)**. Madison: AAAI Press. p. 580 - 587. 1998.

LOPES, L. F. *et al.* Sistema de Conhecimento para diagnóstico em acupuntura: uma modelagem usando o CommonKADS. **Gestão e Produção**, (UFSCAR. Impresso). v. 18, n. 2, p. 351-366, 2011.

MAHONEY, S. M.; LASKEY, K. B. Network engineering for complex belief networks. In **Proceedings of the 12th Annual Conference on Uncertainty in Artificial Intelligence (UAI'96)**. San Francisco: Morgan Kaufmann, p. 389 – 396, 1996.

MINCHIN, R. *et al.* Syntoms Ontology for Mapping Diagnostic Knowledge Systems. **Proceedings of the 19th IEEE Symposium on Computer-Based medical Systems**. 2006.

NORSYS, Netica Bayesian Networks Software from Norsys. Disponível na Web em: <<http://www.norsys.com/netica.html>>. Acesso em: 22 set. 2008.

PARSIA, B., HALASCHEK-WIENER, C.; SIRIN, E. Towards incremental reasoning through updates in OWL-DL. In: **Proceedings of the workshop at 15th international, World Wide Web conference on reasoning on the Web**, 2006.

PROTÉGÉ. (2007). Disponível em: <http://protege.stanford.edu>. Acesso em: 2 de setembro de 2009.

RODRÍGUEZ, A. *et al.* ADONIS: Automated Diagnosis System based on Sound and Precise Logical Descriptions. **Computer-Based Medical Systems. 22nd IEEE International Symposium**, Madrid, p. 1-8, 2009a.

RUSSEL, S.; NORVIG, P. **Artificial intelligence: a modern approach**. New Jersey: Prentice-Hall, 1995.

SCHREIBER, G. *et al.* **Knowledge engineering and management: the CommonKADS methodology**. Massachusetts: MIT Press, 2002.

SHORTLIFFE, E. H. *et al.* **Medical Informatics: Computer Applications in Health Care and Biomedicine**. 2. ed. New York: Springer-Verlag, 2001.

STRUDER, V. R.; BENJAMINS, R.; FENSEL, D. Knowledge engineering: principles and methods, **Data & Knowledge Engineering**, v. 25, p. 161–197, 1998.

YANG, Y. **A Framework for Decision Support Systems Adapted to Uncertain Knowledge**. Tese (Doutorado em Ciência da Computação), Universität Karlsruhe (TH), 2007.

YANG, Y.; CALMET, J. OntoBayes: An ontology-driven uncertainty model, in: **Proceedings IAWTIC-2005**. IEEE Computer Society, p. 457-463, 2005.

ZHAO, W.; YANXIANG, H.; HUI, J. A model of intelligent distributed diagnosis. In **Proceedings of the 8th international conference on high-performance computing in the Asia-Pacific Region**. 2005.

AN ENGINEERING MODEL BASED ON ONTOLOGY AND PROBABILISTIC CALCULATION TO SUPPORT THE DIAGNOSIS

Abstract

The diagnosis, as knowledge-intensive task, is a complex process since there is a wide variety of elements and circumstances to be considered for a decision-making. Uncertainty generated by the subjectivity, vagueness and/or lack of updated information exist in almost all stages and interfere for the safety and efficacy in the outcome. The data and useful information, when collected and treated appropriately, deriving from diagnosis accomplished and which remain latent (unobserved/asleep), can become a valuable source of knowledge if associated with the experience and observation of the individual who uses them. The goal of this article is to propose a model of Knowledge Engineering that allows the creation of new knowledge to support the diagnosis process. The methods and techniques of Knowledge Engineering, used on this model to support the process are: CommonKADS, Ontology, Probabilistic Calculation and Discovery Systems Based on Literature. Through the integration of these elements, the proposed model is applied to a didactic example which allows evidence to be highlighted and analyzed through research literature as potential new knowledge. After the information of a new knowledge, the inference process is updated. It is concluded, therefore, that through this research, the proposed model meets the requirements for the generation of new knowledge, and contributes to the improvement of the diagnostic test.

Keywords: Knowledge Engineering. Knowledge Based Systems. Diagnosis. Ontology. Probabilistic Calculation.