

PROJETO GISSA:
META FÍSICA 3 – Atividade 3.1
Definir Modelo de Inteligência de Gestão na Saúde

Draft 1.0

Novembro 2015

RESERVADO

| | | |
|------------------------------------|--|---------------------|
| Responsável: Mauro Oliveira | | Modelo: B.01 |
| Elaborador(es): | e-mail | Telefone |
| Mauro Oliveira | amaurooliveira@gmail.com | |
| Ronaldo Ramos | ronaldo.ramos@gmail.com | |
| Anilton Sales Garcia | aniltonsallesg@gmail.com | |
| Tiago Prince Sales | tiago_sales@atlantico.com.br | |

ÍNDICE

| | | |
|-----|--|----|
| 1 | Histórico | 4 |
| 2 | Introdução | 4 |
| 2.1 | Sobre deste documento | 4 |
| 2.2 | Escopo do Modelo de Inteligência de Gestão na Saúde do proposto..... | 4 |
| 2.3 | Glossário Resumido | 5 |
| 3 | Referencial Teórico | 6 |
| 3.1 | Gestão do Conhecimento..... | 7 |
| 3.2 | Ontologias..... | 9 |
| 3.3 | Mineração de Dados | 13 |
| 4 | O Framework GISSA..... | 13 |
| 4.1 | Funcionalidades | 20 |
| 4.2 | Componentes..... | 21 |
| 5 | Framework de Gestão do Conhecimento..... | 24 |
| 5.1 | Bloco de Inferência por Regras..... | 24 |
| 5.2 | Bloco de Busca Semântica..... | 27 |
| 5.3 | Bloco de Mineração de Dados | 28 |
| 6 | O GISSA como um Modelo de Inteligência de Gestão de Saúde | 33 |
| 7 | Conclusão | 34 |
| 8 | Referências | 35 |

1 HISTÓRICO

| Data | Versão | Responsável | Alteração |
|------------|--------|----------------|--|
| 06/04/2015 | A.01 | Ronaldo Ramos | - Criação do documento base 2 (Análise de Risco Sistêmico utilizando Mineração de Dados) |
| 11/08/2015 | A.02 | Tiago Prince | - Criação do documento base 1 (Framework de Gestão do Conhecimento) |
| 29/11/2015 | A.03 | Mauro Oliveira | - Criação do Documento Meta Física 3 0 Atividade 3.1 (Modelo de Inteligência de Gestão de Saúde) |
| 30/11/2015 | A.04 | Anilton Garcia | - Revisão do documento |

2 INTRODUÇÃO

2.1 Sobre deste documento

Este documento tem como objetivo apresentar os resultados da META FÍSICA 3 – Atividade 3.1 que definem o Modelo de Inteligência de Gestão na Saúde do Projeto GISSA.

O documento aborda os três componentes que compõem o modelo de inteligência de gestão na saúde do GISSA: Inferência por Regras, Busca Semântica e Mineração de Dados. Nestes três componentes são detalhados o percurso metodológico das propostas apresentadas, bem como as arquiteturas associadas.

2.2 Escopo do Modelo de Inteligência de Gestão na Saúde do proposto

O GISSA fornecerá inteligência de governança na tomada de decisão nos cinco domínios clássicos da área de Gestão em Saúde: sistêmico, normativo, funcional, clínico e de cuidados, integrando os sistemas de informação em saúde no âmbito do Sistema Único de Saúde (SUS) e Cartão Nacional de Saúde (CNS).

Estruturalmente, o GISSA será constituído por um conjunto de componentes habilitados para coletar informações em tempo real de contextos inerentes às situações monitoradas de saúde e alimentar mecanismos inteligentes capazes de produzir informações qualificadas, as quais municiam diversas aplicações voltadas para os mais variados atores envolvidos: usuários do sistema, profissionais de saúde, gestores de instituições de saúde e gestores governamentais, tomadores de decisão do sistema de saúde.

O GISSA envolverá áreas de fronteira do conhecimento técnico-científico, em dois aspectos: de um lado, no campo da saúde coletiva, o planejamento e gestão em

saúde pública e saúde da família; de outro lado, no campo da tecnologia da informação, os mecanismos de inferência, baseados em técnicas de inteligência artificial e modelos de ontologias. Ele será constituído por um conjunto de componentes habilitados para coletar informações em tempo real de contextos inerentes às situações monitoradas de saúde e alimentar mecanismos inteligentes capazes de produzir informações qualificadas, as quais municiam diversas aplicações voltadas para os mais variados atores (desde pacientes, agentes de saúde, médicos, secretários de saúde à presidente da república) tomadores de decisão do sistema de saúde.

2.3 Glossário Resumido

- **Framework:** conjunto de componentes de software e/ou processos combinados para a solução de uma classe de problemas. Um framework fornece pontos de extensão para a construção de soluções customizadas para problemas específicos. O framework possui uma arquitetura que define as relações entre seus componentes, a forma de utilizá-los e como implantar a solução final.
- **Gestão do Conhecimento (GC):** processo cíclico composto principalmente por atividades de criação, integração e disseminação do conhecimento em uma organização (ou um conjunto delas) [1].
- **Ontologia:** modelo que formaliza de forma explícita e inequívoca a visão de mundo de uma comunidade sobre um domínio de interesse. Uma ontologia define um conjunto de conceitos, suas relações e propriedades.
- **Inferência:** conclusão de novas informações a partir de informações já existentes. Inferências podem ser feitas por meio de raciocínio automatizado, utilizando um conjunto de regras formalizado em uma ontologia, ou por meio de algoritmos de mineração de dados e *machine learning*.
- **Data Mining:** conjunto de ferramentas, técnicas e metodologia baseadas em algoritmos de aprendizagem ou em classificação (redes neurais e estatística) capazes de explorar um conjunto de dados, extraindo ou ajudando a evidenciar padrões nestes dados e auxiliando na descoberta de conhecimento (ver conceito no item 3.1).
- **Business Intelligence:** refere-se ao processo de coleta de grande volume de dados, organização, análise, compartilhamento e monitoramento de informações (normalmente contidas em um Data Warehouse/Data Mart), analisando-os e desenvolvendo percepções e entendimentos a seu respeito, tomada de decisão.

- **Data Warehouse:** possibilita a análise de grandes volumes de dados, coletados dos sistemas transacionais (OLTP). São as chamadas séries históricas que possibilitam uma melhor análise de eventos passados, oferecendo suporte às tomadas de decisões presentes e a previsão de eventos futuros.
- **OLTP** (Online Transaction Processing): são sistemas que têm a tarefa de monitorar e processar as funções básicas e rotineiras de uma organização, tais como processamento da folha de pagamento, faturamento, estoque, etc. Os fatores críticos de sucesso para este tipo de sistema são: alto grau de precisão, integridade a nível transacional e produção de documentos em tempo hábil.
- **OLAP** (Online Analytical Processing): trata-se da ferramenta mais popular para exploração de um data warehouse. OLAP fornece para organizações um método de acessar, visualizar, e analisar os dados corporativos com alta flexibilidade e performance, via um modelo de dados natural e intuitivo. DW & OLTP

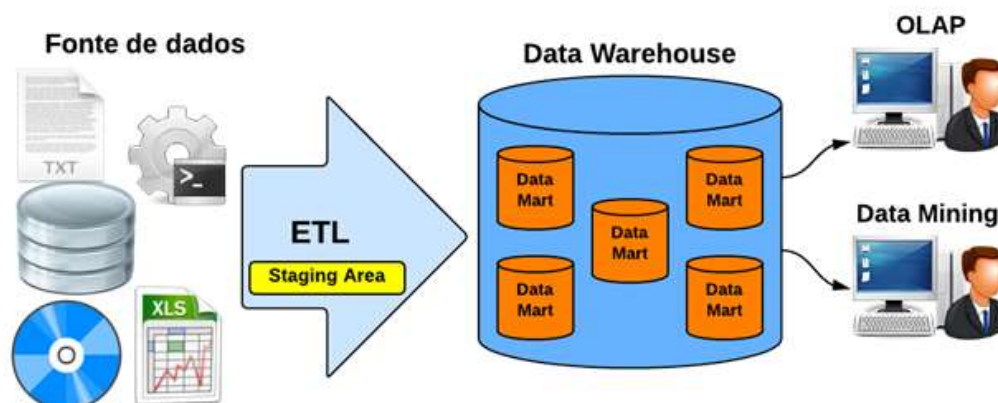


Figura 0. Fluxo de dados: OLTP/ ETL/ Datawarehouse/ DataMining-OLAP

3 REFERENCIAL TEÓRICO

O framework de gestão do conhecimento do GISSA é construído em cima de três pilares: gestão do conhecimento, ontologia e mineração de dados. Essa seção descreve os conceitos fundamentais de cada uma das áreas.

3.1 Gestão do Conhecimento

Antes da definição de um framework de gestão do conhecimento, é necessário estabelecer o que se entende sobre dois conceitos fundamentais: *conhecimento* e *gestão do conhecimento* (GC). Essa discussão é necessária visto que não há consenso, nem mesmo na própria comunidade de GC, sobre o que eles significam. De fato, as diferentes visões sobre o conceito de “conhecimento” e as diferentes aplicações da gestão do conhecimento, contribuem para o surgimento de várias definições.

A definição de *conhecimento* que se aplica melhor ao contexto do GISSA é a construída por Alavi e Leidner em [2]: conhecimento é informação contida na mente de uma pessoa; ela é uma informação “personalizada”, influenciada por interpretações, ideias, observações, julgamentos, que podem ou não ser corretos, úteis, precisos ou estruturados. Para os autores, conhecimento é análogo ao conhecimento tácito e informação ao conhecimento explícito, como definido por Nonaka e Takeuchi [3].

Para exemplificar a relação entre conhecimento e informação, considere um secretário municipal de saúde. Ao receber um relatório indicando que o número de partos prematuros aumentou nos últimos meses, ele está recebendo uma informação (ou um conhecimento explícito). Ao analisar essa informação e combinar com outras para concluir que isso se deve à falha no acompanhamento das gestantes por parte dos agentes comunitários de saúde (ACS), o secretário produz conhecimentos em sua mente (conhecimento tácito). Para mitigar o problema, o secretário decide capacitar seu grupo de agentes. Após alguns meses, ele percebe que o número de partos prematuros voltou ao normal. Percebendo que sua ação deu resultado, ele registra essa ação em um manual operacional. Nesse momento, ele está formalizando seu conhecimento tácito, tornando-o informação (conhecimento explícito).

A partir dessa definição de conhecimento, adota-se no GISSA que *gestão do conhecimento* significa o conjunto de atividades para a criação, transformação, integração e disseminação de conhecimentos em um organização [2]. Os principais benefícios da adoção e incentivo a práticas de GC para uma organização incluem: a melhoria na tomada de decisão, o aprendizado organizacional, o acesso facilitado

ao conhecimento, o desenvolvimento de inovação e o aumento do desempenho nos processos organizacionais.

Os produtos de software que dão suporte a um ou mais processos são chamados de Sistemas de Gestão do Conhecimento (SGC). Existem diversos tipos de SGC que apoiam diferentes etapas da gestão do conhecimento e, portanto, contribuem de diferentes formas para as organizações. Alguns dos principais tipos são:

- **Sistemas de Memória Organizacional:** memória organizacional consiste em um repositório de conhecimentos adquiridos no passado que são trazidos para as atividades do presente, a fim de melhorar o desempenho organizacional [4]. Fazem parte da memória organizacional tanto documentos formais, como planos, manuais e padrões de procedimento, como registros de conhecimento tácito, como justificativas do por que decisões foram tomadas, comunicações informais e resultados de ações [5].
- **Sistemas de Colaboração:** sistemas que apoiam grupos de pessoas trabalhando em conjunto para atingir objetivos em comum [6]. Também conhecidos como Groupware, essa classe de SGC é muito difundida. São exemplos de sistema de colaboração: sistemas de workflow, chat, email, conferência, entre outros.
- **Sistemas de Apoio a Decisão:** sistemas construídos utilizando alguma tecnologia de inferência para apoiar o processo de tomada de decisão [5]. Sistemas especialistas são um tipo particular de sistemas de apoio à decisão que capturam o conhecimento de experts em uma *knowledge base* (base de conhecimento) que, ao ser combinada com um *reasoner* (máquina de inferência), é capaz de simular o processo de solução de problemas realizado por especialistas, gerando novos conhecimentos a partir de entradas de dados. Uma revisão da literatura sobre sistemas de apoio a decisão no contexto de saúde pública pode ser encontrada em [7].
- **Sistemas de Recomendação:** sistemas que auxiliam usuários na seleção de itens de interesse em um grande repositório, apoiando-os na tarefa de lidar com uma vasta fonte de informação [5]. Exemplos de soluções dessa natureza são utilizados comumente em serviços de *stream* de filmes e música, como Netflix e Spotify, mas também em sistemas de e-commerce, com Amazon e Submarino.
- **Enterprise Knowledge Portal:** soluções computacionais que funcionam como grande “guarda-chuva” de SGC em um organização [5]. Pode ser entendido com uma interface amigável para os membros da organização acessarem os diferentes sistemas por um mesmo canal de entrada. Uma iniciativa dessa natureza é reportada em [8], [9].

3.2 Ontologias

Ontologias são representações formais e explícitas de uma conceitualização compartilhada por comunidade sobre um domínio de interesse. Em outras palavras, ontologias são teorias que definem um conjunto de conceitos, com suas interrelações e propriedades, que são utilizados por um grupo de pessoas para entender e classificar uma porção da realidade.

Uma ontologia sobre nascimento, por exemplo, poderia definir que o nascimento é um evento, do qual apenas um ser vivo participa, o nascido, e que ocorre em lugar determinado e de forma concomitante a outro evento, o parto, da qual a mãe do respectivo nascido participa.

Existem diversas classificações para ontologias na literatura. No contexto do GISSA, é importante fazer a distinção entre duas delas: ontologias como *modelos de referência*; e ontologias como *modelos de implementação* para raciocínio automatizado.

Ontologias, no sentido de *modelos de referência* são modelos voltados para o entendimento e compreensão de seres humanos e sua principal aplicação é na interoperabilidade semântica, isto é, na construção de soluções computacionais que provejam uma alta confiabilidade de que as informações compartilhadas sejam entendidas de forma correta [10].

IMPORTANTE: A ontologia de referência provê a visão dos especialistas do domínio sobre os dados manipulados pelos sistemas e não a visão dos profissionais de tecnologia de informação que implementam o sistemas que manipulam esses dados.

OntoUML [11] é o exemplo de linguagem utilizada para construção de tais artefatos. Construída a partir de teorias oriundas da Ontologia Formal, Psicologia, Ciências Cognitivas e Teorias Lógicas, a linguagem provê um conjunto de elementos bem definidos para a construção de ontologias bem fundamentadas. OntoUML é apoiada por um conjunto de ferramentas de software que apoiam a construção [12], validação [13] e implementação [14] de tais modelos.

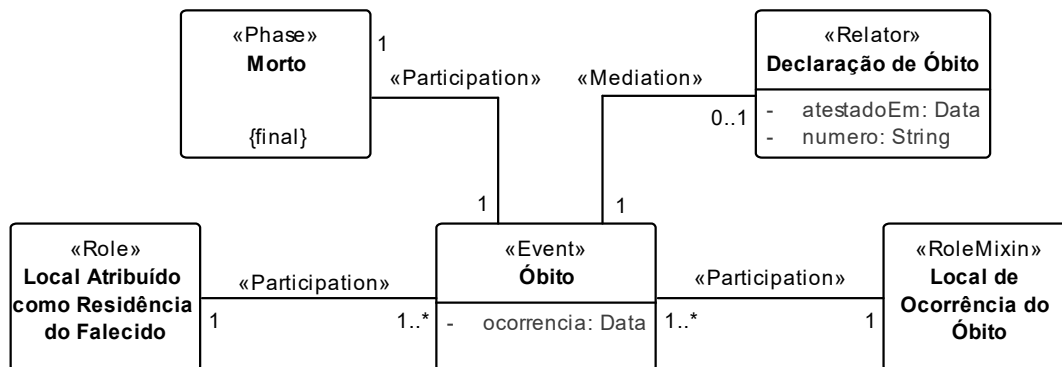


Figura 1. Algumas propriedades de óbito.

O diagrama apresentado na Figura 1 faz parte da ontologia do domínio clínico epidemiológico do projeto GISSA, e descreve algumas propriedades de óbito, como entendido pelos especialistas de saúde do projeto e a legislação oficial vigente. Nesse modelo, um morto é definido como a pessoa que participou de um evento de Óbito, o qual ocorre em um determinado local, e para o qual uma Declaração de Óbito pode ser escrita.

A Figura 2 apresenta um outro segmento da ontologia, que contém uma parte da hierarquia de tipos de óbito, com ênfase nos óbitos fetais e infantis. A ontologia captura, por exemplo, que óbitos neonatais são aqueles que ocorrem até 30 dias após o nascimento, enquanto óbitos pós neonatais são aqueles que ocorrem entre o 31º e o 364º dia após o nascimento.

Ontologias de referência podem dar origem a implementações nas mais diversas tecnologias. Este processo é normalmente realizado de forma automática, por meio de ferramentas orientadas a modelo (Model-driven Architecture - MDA) [15]).

O Menthor Editor¹ é um exemplo de ferramenta que provê tais transformações, permitindo a geração automática de modelos de implementação em RDF (*Resource Description Framework*) [16] e OWL (*WEB Ontology Language*) [17], linguagens padronizadas pela W3C (*WWW Consortium*) para implementação de ontologias na web semântica.

As ontologias implementadas permitem a construção de ferramentas com raciocínio automatizado, se populadas e alimentadas a máquinas de inferência (chamadas de *reasoners*).

São exemplos de *reasoners*: Pellet, Fact++, Racer, Hermit, Snorocket, entre outros.

¹ <http://www.menthor.net>

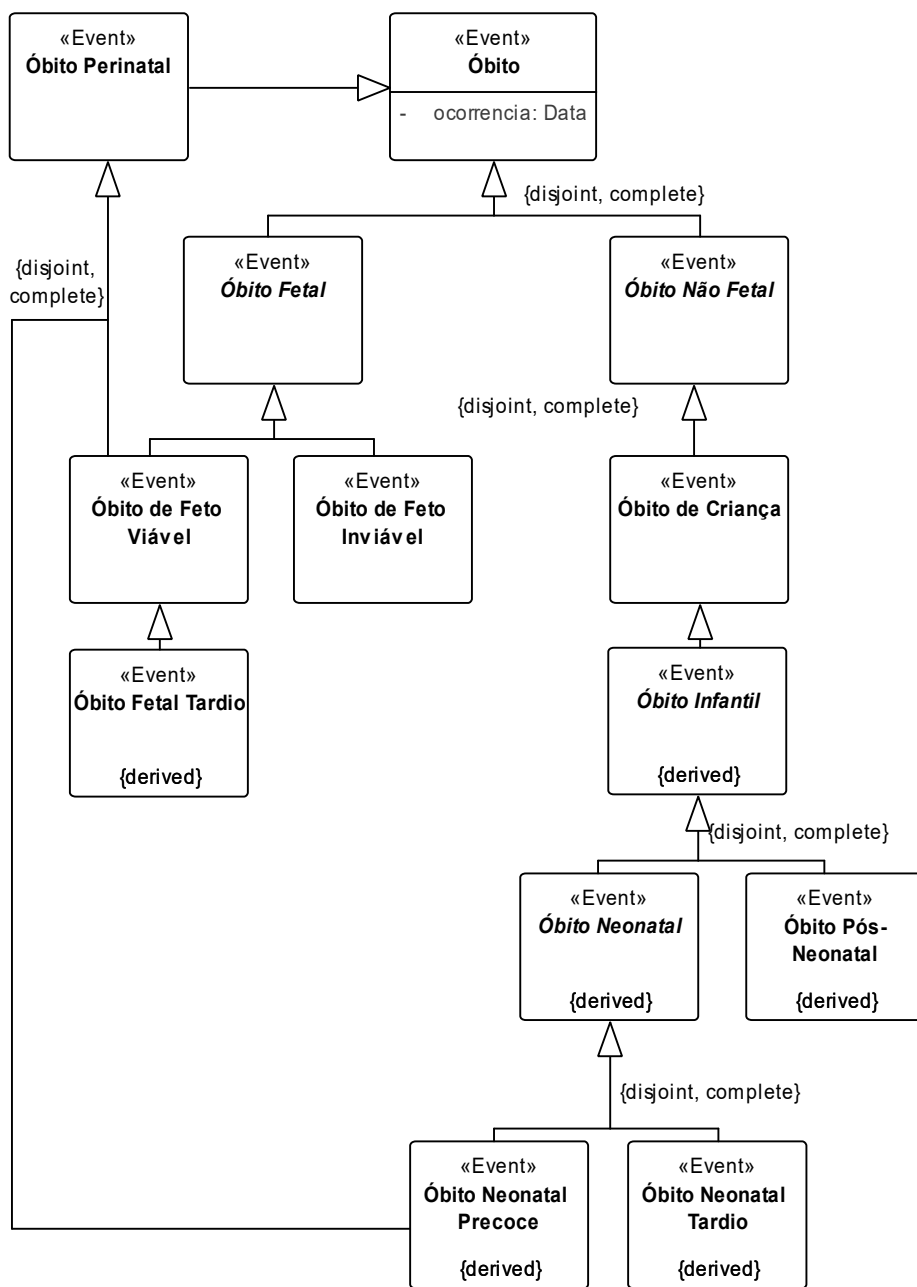


Figura 2. Classificação de óbitos fetais e infantis formalizada na ontologia.

A regra a seguir exemplifica, no contexto do domínio clínico-epidemiológico, o processo de inferência realizado pelos *reasoners*:

SE ((gestante tem idade igual ou inferior a 15 anos) **E** (gestante possui histórico de abortamento) **E** (gestante é membro de uma família de baixa renda))
ENTÃO { gestante é de alto risco } **E** {gestante requer acompanhamento constante de seu agente de saúde}

Regras de inferência são sempre especificadas no formato SE-ENTÃO. Se determinado conjunto de condições for verdade, então uma conclusão pode ser obtida. Caso a regra descrita anteriormente fosse avaliada para uma gestante de 14 anos, que já abortou no passado e cuja mãe recebe bolsa-família, o sistema iria classificá-la como sendo uma gestante de alto risco e apontaria a necessidade de acompanhamento próximo.

Sistemas que manipulam ontologias e suas instâncias armazenam essas informações em sistemas gerenciadores de banco de dados especiais, conhecidos como *triple stores*. São exemplos de *triple stores*: Stardog², Virtuoso³, Oracle Spatial and Graph⁴, GraphDB⁵, AllegroGraph⁶, Neo4J⁷, entre outros.

Ressalta-se que muitos desses servidores de bancos de dados provêm os próprios mecanismos de inferência, propiciando um desempenho muito melhor para aplicações que lidam com grandes massas de dados.

Dentro das *triple stores* são armazenadas as *bases de conhecimento*. Essas bases são a versão os bancos de dados relacionais do ambiente da web semântica. As bases de conhecimento contém dois componentes principais:

- T-Box: a implementação da estrutura da ontologia (em OWL, por exemplo), enriquecida com regras de inferência (em SWRL-*Semantic Web Rule Language*) [18], por exemplo); e
- A-Box: as instâncias da ontologia

Em uma analogia com o mundo relacional, um T-Box é equivalente a um esquema do banco de dados, enquanto a A-Box se equivale ao conjunto de todos os registros armazenados nas tabelas da base. A principal diferença do mundo de triplas para o mundo relacional é que as *triple stores* tratam a estrutura e as instâncias da ontologia da mesma forma, tornando essa solução mais adequada a evolução e integração de diferentes repositórios.

A recuperação dos dados armazenados em *bases do conhecimento* é feita utilizando uma linguagem chamada SPARQL [19]. Fazendo novamente uma comparação com o mundo relacional, SPARQL seria o SQL para consultar bases de conhecimento.

² <http://www.stardog.com>

³ <http://virtuoso.openlinksw.com/>

⁴ <http://www.oracle.com/us/products/database/038407.htm>

⁵ <http://ontotext.com/products/ontotext-graphdb/>

⁶ <http://franz.com/agraph/allegrograph/>

⁷ <http://neo4j.com/>

Finalmente, são destacadas a seguir as principais aplicações das tecnologias de web semântica:

- **Sistemas de Apoio à Decisão:** como explicado na seção anterior sobre Gestão do Conhecimento, essa classe de sistema faz uso intensivo dos mecanismos de inferência.
- **Sistemas de Anotação e Busca Semântica:** essa classe de sistemas busca auxiliar o processo de armazenamento e recuperação de informação não estruturada, isto é, informações que não são organizadas em bases de dados mas no formato de textos, imagens, vídeos e similares. O papel da ontologia nesses sistemas é servir de vocabulário para anotar tais artefatos não estruturados e mapear relações entre eles, que servirão de input para construção de ferramentas inteligentes de recuperação dos conteúdos anotados. Esse tipo de solução é muito utilizado por corporações de mídia digital, como Globo [20] e BBC [21].
- **Repositórios de Dados Abertos:** soluções cujo objetivo principal é a publicação de dados em formatos legíveis por máquina. O papel das ontologias nessa classe de produtos é prover a estrutura que será utilizada para caracterizar os dados que serão disponibilizados. Diversas iniciativas privadas e governamentais estão em funcionamento, como MusicBrainz [22], um repositório público produção musical; Dbpedia [23], uma grande enciclopédia em formato de dados abertos, e Data.gov [24], uma iniciativa Norte Americana para publicar dados governamentais.

3.3 Mineração de Dados

Desde os primórdios, o ser humano sempre aprendeu pela observação. Com o advento de metodologias científicas ele passou a observar padrões formulando hipóteses e testando-os para descobrir regras. No entanto, a modernidade da era digital tem criado um volume enorme extraordinário de dados que impossibilita a identificação de padrões de forma manual em um prazo razoável. Não é sem razão que “Big Data” tem se tornado um tema da atualidade tanto no âmbito acadêmico quanto de mercado.

Mineração de Dados é uma tecnologia computacional construída por um conjunto de ferramentas, técnicas e metodologia que usam algoritmos de aprendizagem ou classificação, baseados em redes neurais e estatística, capazes de explorar um

conjunto de dados, extraindo ou ajudando a evidenciar padrões nestes dados e auxiliando na descoberta de conhecimento tácito (ver item 3.1).

Pode-se, assim, diferenciar business intelligence (BI) e mineração de dados (MD). O BI busca um conhecimento novo e útil acerca do seu meio ambiente e funciona no plano estratégico (“aquele conhecimento que o indivíduo obtém da informação explícita interpretada”). A MD visa obter, a partir dos dados operativos brutos, informação útil para subsidiar a tomada de decisão no plano tácito (“aquele conhecimento que o indivíduo adquiriu ao longo da vida, pela experiência”).

O data mining, diferentemente do OLAP, fornece informações de dados corporativos ocultos em grandes bancos de dados, podendo prever comportamentos futuros, se tornando uma importante ferramenta para tomada de decisão dos gestores. Os tipos de informações obtidos com o data mining abrangem associações, sequências, classificações, aglomerações e prognósticos

A seguir é mostrada uma metodologia que usa mineração de dados no contexto do GISSA com o objetivo de prever a situação final da paciente após certo período de acompanhamento pelo sistema. A ideia é que se possa intervir junto a paciente caso a mesma apresente alta similaridade de casos em que há fracasso, ou seja, houve perda da vida da mãe ou de alguma criança ou ainda a presença de sequelas.

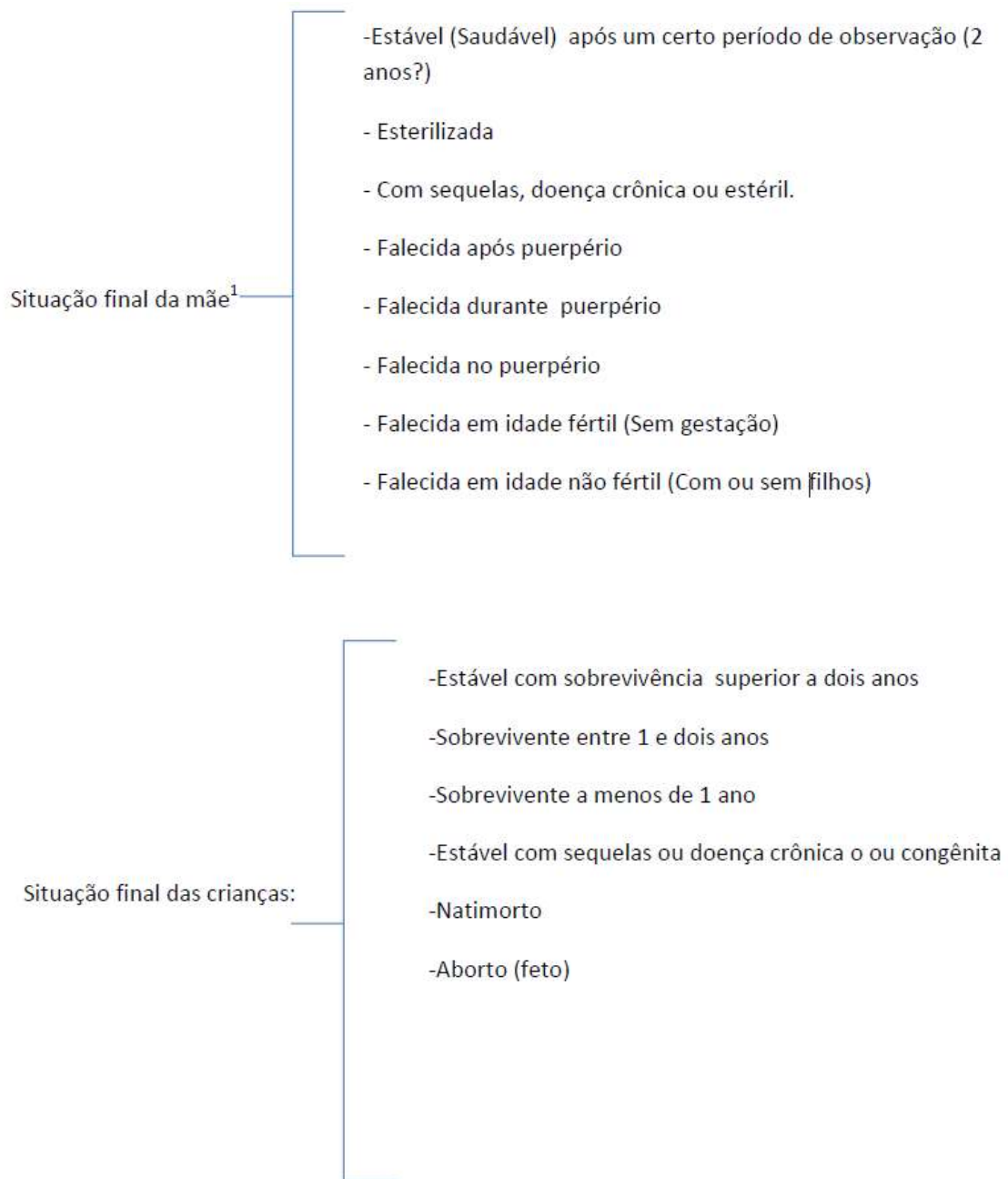
Metodologia

Inicialmente, são definidas as classes ou categorias alcançáveis em um certo período de acompanhamento pelo sistema. Em seguida são levantados todos os atributos a partir dos diversos bancos de dados dos sistemas de acompanhamento. É montado o modelo de inferência e são selecionados os atributos mais relevantes com o intuito de afinar o modelo até chegar a uma taxa de acerto (ver o tema métrica do classificador) aceitável.

Classes (Categorias) e Atributos

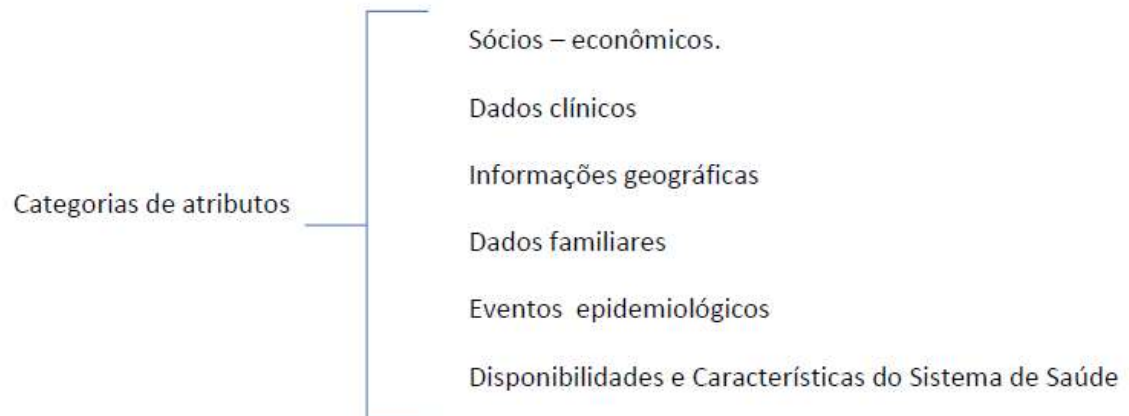
Inicialmente é feito o levantamento das classes que serão as metas dos classificadores a serem utilizados. As categorias se definem pelo resultado final do processo que tem a ver com a situação final da mãe e/ou das crianças.

O levantamento inicial leva a seguinte categorização:



Algumas combinações destes elementos formam as classes “META”. Desde o início do atendimento poder-se-á calcular a probabilidade da acompanhada terminar seu processo em uma das classes existentes.

Após a definição das classes são levantados os atributos que serão analisados. Pode-se, inicialmente, definir os atributos nas categorias apresentadas abaixo:



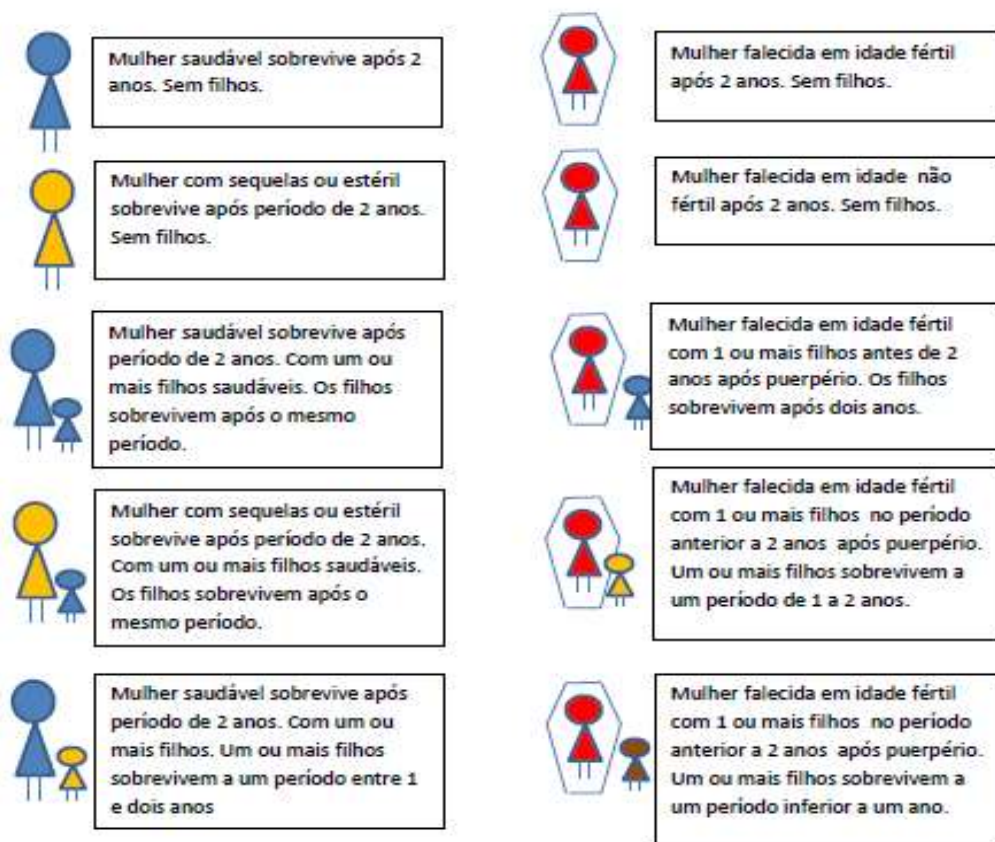
| | | |
|-------------------------|---------------|------------------|
| Projeto: G125A | Status: DRAFT | Versao 1.0 |
| Autor: Ronaldo F. Ramos | Edição: | Data: 06/04/2015 |
| | | |

ENTRADA DO PROCESSO (REDE CEGONHA)



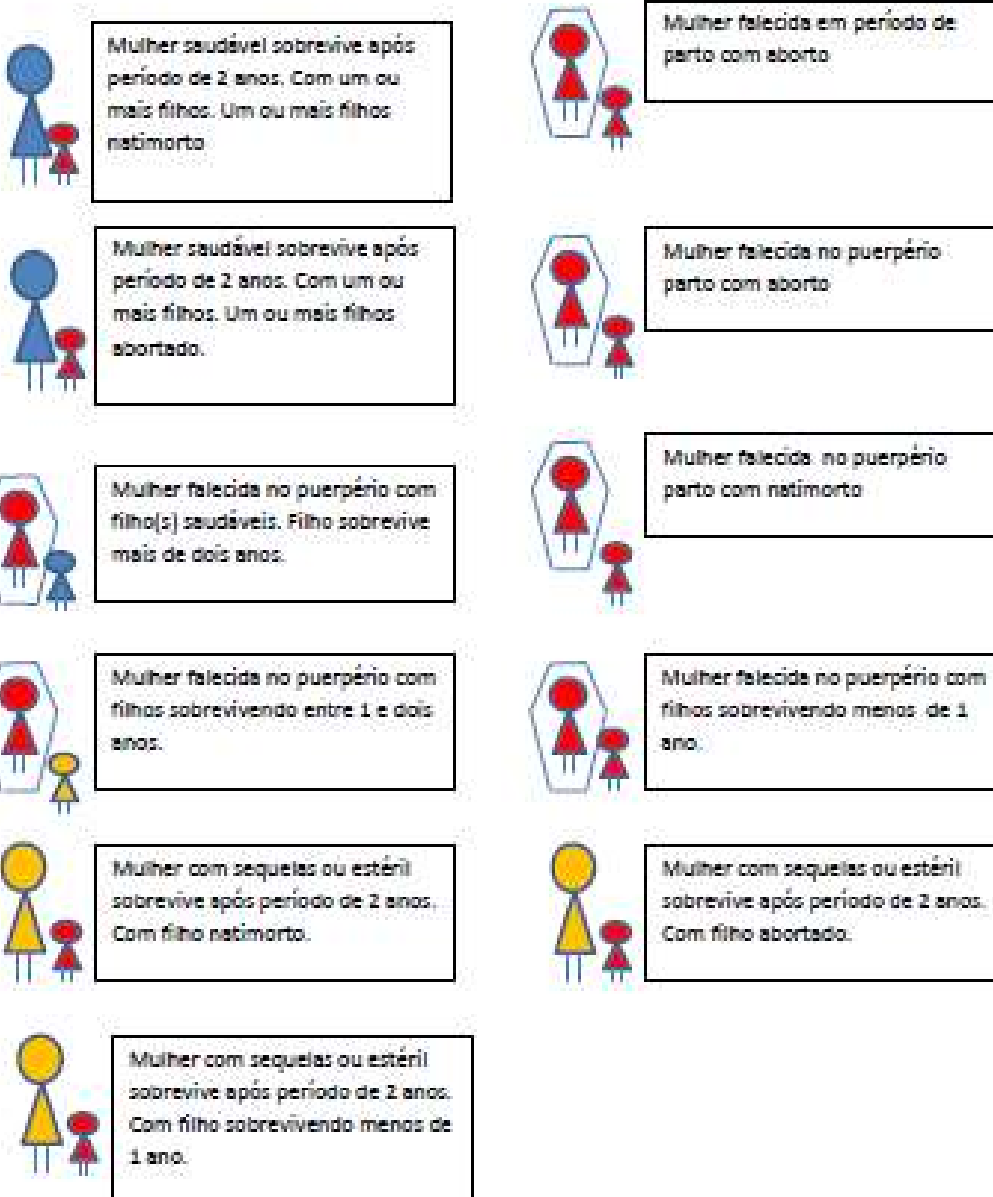
Mulher (Atributos: idade, no de filhos, doenças preexistentes, etc)

SAÍDAS DO PROCESSO (Classes)



| | | |
|-------------------------|---------------|------------------|
| Projeto: GI25A | Status: DRAFT | Versão: 1.0 |
| Autor: Ronaldo F. Ramos | Edição: | Data: 06/04/2015 |

SAÍDAS DO PROCESSO (Classes) (Continuação)



Após consolidado o conjunto de atributos obtidos a partir das diversas bases de dados, deve-se utilizar um método de seleção de atributos (ainda a definir) e um modelo de inferência a utilizar. Algumas tecnologias a testar seriam as redes neurais artificiais e árvores de decisão (Algoritmos da família do ID3/ C4.5)

A próximas etapas a serem executadas são:

1. Concluir o levantamento das classes.
2. Definir os atributos e como obtê-los dos sistemas legados (Grafos de Localização e Composição de informação)
3. Propor elementos a equipe de bancos de dados do GISSA para contemplar atributos ausentes.
4. Montar as tabelas (Data Marts)
5. Aplicar os algoritmos de seleção de atributos
6. Aplicar algoritmos de inferência
7. Selecionar tecnologia a ser incorporada ao GISSA
8. Implementar a tecnologia como um serviço do GISSA.

4 O FRAMEWORK GISSA

O framework GISSA (de Governança Inteligente dos Sistemas de Saúde) é uma solução genérica para construção de sistemas de informação que apoiem o processo de tomada de decisão no contexto da saúde pública. Como prova de Conceito (PoC) do framework será implementada uma solução para apoiar a rede cegonha do município de Tauá – CE. O framework é formado por um conjunto de componentes para permitir coleta, integração, transformação, inferência e visualização de informações de forma a fornecer aos usuários finais fatos e dados necessários às diversas decisões relacionadas a saúde pública.

4.1 Funcionalidades

O framework prevê cinco principais classes de funcionalidades, descritas a seguir.

Relatórios são funcionalidades que listam um conjunto de eventos, pessoas ou itens e suas características relevantes, como, por exemplo: óbitos, nascimentos, gestantes, crianças, etc. Relatórios são as formas de apresentação de informações mais granulares do GISSA. Sua função é fornecer aos usuários a capacidade de realizar análises específicas sobre itens em particular. Para exemplificar um cenário de uso da função Relatório, considere uma listagem das gestantes de risco, disponibilizada para um secretário de saúde municipal. Com ela, um gestor de saúde, por exemplo, é capaz de acessar as características de uma gestante de risco em particular, permitindo-o entender os fatos que levaram o sistema a classificá-la como tal. Com esse conhecimento, o gestor de saúde pode solicitar ao agente de saúde responsável que realize um acompanhamento mais próximo.

Indicadores são funções que fornecem informação agregada sobre um determinado fato. Indicadores podem ser taxas (ex: taxa de mortalidade infantil), porcentagens (ex: cobertura vacinal de sarampo), totalizadores (ex: número de gestantes com data provável de parto para o próximo mês), gráficos (ex: um gráfico de barras com as divisões dos óbitos por causas básicas), entre outros. Ao mostrar informações agregadas de forma simples e direta, e por possuírem parâmetros definidos, os indicadores permitem identificar rapidamente se existe ou não um problema. Como cenário do uso dessa classe de funcionalidade, considere um indicador de taxa de partos cesáreos no último mês, disponibilizado para um secretário de saúde. Como existe um incentivo governamental para partos vaginais, uma taxa de 80% de partos cesáreos mostra que o município não está cumprindo sua meta, evidenciando a necessidade de ações de retificação do problema.

Dashboards são funcionalidades de apresentação de informações onde se combinam múltiplos indicadores e relatório, permitindo a construção de painéis que

mostrem, de forma agregada, a situação de determinado evento. No GISSA, os dashboards não provêm novas informações, e sim apresentam de forma simultânea as funcionalidades de relatórios e indicadores. O GISSA prevê pelo menos dois níveis de dashboard. O primeiro nível orientado para os domínios de gestão pública de saúde: clínico-epidemiológico, técnico-administrativo, normativo, gestão compartilhada e gestão do conhecimento. E um segundo, orientado para determinados tópicos dentro de cada domínio. Exemplo de tópicos no domínio clínico-epidemiológico são: pré-natal, puericultura, nascimento e óbito.

São classificadas como **alertas** as funcionalidade orientadas a geração de mensagens para informar aos usuários da ocorrência de um evento em particular (ex: nascimento da pessoa A, entrada no hospital da pessoa B), ou da configuração de uma determinada situação (ex: estoque baixo do medicamento X, cobertura vacinal baixa). Alertas podem ser gerados automaticamente pelo sistema ou a partir de entradas manuais de dados. Alertas também são enviados para diferentes tipos de usuários de acordo com sua gravidade. Por exemplo, um alerta de óbito infantil acidental é enviado apenas para agentes de saúde, enquanto um causado por desnutrição também é enviado para o governador. O objetivo desse tipo de funcionalidade é agilizar a disponibilização de informação aos atores da saúde pública, chamando atenção para necessidade de ações em particular.

A última macro funcionalidade que o GISSA apoia é a de **buscas**. Essa classe de solução visa auxiliar os usuários na recuperação de artefatos que contém informação não estruturada, isto é, texto, imagens e vídeos. As buscas são voltadas principalmente para três dos cinco domínios do GISSA:

- *normativo*, no qual se deseja recuperar conhecimentos contidos em portarias, leis, decretos e resoluções;
- *gestão compartilhada*, no qual se deseja recuperar documentos produzidos nas comissões intergestores, como atas e deliberações; e
- *gestão do conhecimento*, no qual a recuperação é orientada para protocolos de saúde, os quais contém conhecimentos clínicos consolidados para orientar procedimentos médicos em situações específicas, como pré-natal de alto risco.

4.2 Componentes do framework GISSA

Para dar suporte ao desenvolvimento das funcionalidades descritas o GISSA possui quatro grandes componentes: coleta, persistência, inteligência e visualização. O agrupamento das atividades do GISSA nesses componentes é representado na Figura 3. Nela é mostrada a visão lógica do framework, destacando os componentes do framework de gestão do conhecimento (Inteligência) em azul.

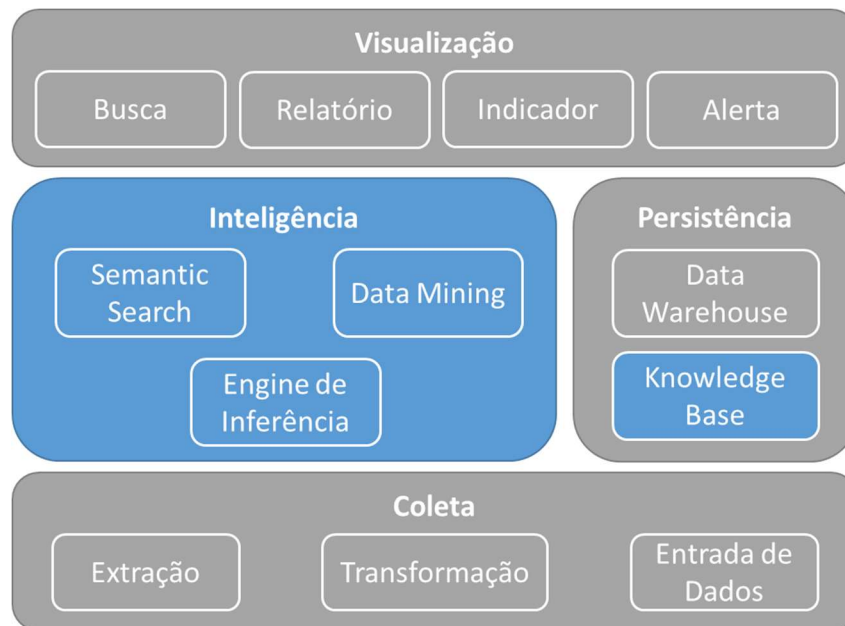


Figura 3. Visão lógica do GISSA framework, com ênfase no framework de GC

- **Coleta:** componente responsável por recuperar dados das diversas fontes produtoras e transformá-los no formato do GISSA. O componente auxilia a extração de dados das mais diversas fontes que incluem: sistemas locais, como e-SUS e SINASC; sistemas web, como SISPRENATAL e HORUS; barramento de *web services*, como o CADSUS; e fontes de dados públicas disponíveis na internet, como o repositório do bolsa família. O componente inclui scripts de ETL genéricos para a transformação dos dados para o formato padrão do GISSA.
- **Visualização:** componente responsável por prover bibliotecas para a construção das interfaces com usuário, sendo otimizado para a construção das cinco classes de funcionalidades descritos na seção anterior. São exemplos de partes desse componente: componente de dashboard genérico, componentes de formulário (ex: *date picker*), componente de fachada de serviço REST.
- **Inteligência:** parte integrante do framework de gestão do conhecimento, o componente de inteligência contém mecanismos de geração de novos conhecimentos, como o *engine* de inferência e o componente de mineração de dados, além do *engine* de busca semântica, que permite o uso do conhecimento formalizado na recuperação dos artefatos com informação não estruturada.
- **Persistência:** componente de armazenamento de informações que contém os diversos repositórios de dados utilizados pelo GISSA, como o

Datawarehouse, a *Knowledge Base* (como descrita na Seção 3) e a Base para as entradas de dados auxiliares.

Para o framework de Gestão de Conhecimento, a seguir descrito, deverão ser explicitamente definidos quais os relatórios, alertas, consultas e indicadores que serão baseados em no conceito de inteligência computacional. Devem também estar definidos as tecnologias a serem usadas.

Assim, ao modelo de inteligência do GISSA, através de uma saída extremamente refinada, é incumbida a tarefa de:

- Aplicar de Mineração de Dados e Inferência das Informações
- Gerar de Alarmes relacionados à Morte Materna e Morte Infantil. A especificação de quais alertas farão parte da *engine* de inteligência tem como base os requisitos que serão implementados e priorizados no backlog.
- Realizar de Buscas Textuais e Mineração de Textos.
- Possuir a capacidade de adaptação
- Possuir algum tipo de feedback, característica imprescindível para um sistema inteligente.

5 FRAMEWORK DE GESTÃO DO CONHECIMENTO

Como mostrado na figura 3 (Visão lógica do framework GISSA), o framework de gestão do conhecimento (destacado em azul) é o componente do GISSA responsável por torná-lo uma fundação para construção de “sistemas inteligentes”. Este componente é constituído por 3 blocos:

- Inferências por Regras, (Engine de Inferência)
- Buscas Semânticas (Semantic Search)
- Mineração de dados (Data Mining)

A inteligência, nesse contexto, refere-se à capacidade de um agente computacional gerar novas informações a partir de outras contidas no sistema, bem como a capacidade de utilizar o conhecimento formalizado para melhorar as funções do sistema.

O framework de GC compõe os componentes de *backend* do sistema. Isso quer dizer que ele provê novas capacidades para os sistemas desenvolvidos utilizando o framework GISSA, ou ainda melhora as capacidades existentes. Do ponto de vista das principais funcionalidades do sistema (relatórios, indicadores, dashboards, buscas, e alertas) o framework não acrescenta uma nova classe mas permite torná-las inteligentes. Para exemplificar essa característica, considere que o sistema GISSA proverá a função de identificar crianças vulneráveis. Para fornecer essa informação, o sistema utilizará um raciocínio usando a *knowledge base* e o engine de inferência. O usuário acessará um relatório de crianças vulneráveis, por exemplo, que será construído utilizando componente de visualização, alimentado pelo componente de inteligência do framework de GC.

A seguir são detalhados os três blocos funcionais do framework de Gestão do Conhecimento, os quais contém recursos de inteligência para a construção de relatórios, indicadores, alertas e buscas:

5.1 Bloco Inferência por Regras

Como discutido na Seção 2.3, um agente (humano ou computacional) faz uma inferência quando conclui novas informações a partir de outras que ele já possui. Como esse processo pode ser feito de maneiras complementares, este documento separa inferências feitas por meio de regras das realizadas por meio de mineração de dados. Esta subseção aborda apenas o primeiro caso.

As inferências por regras (ou simbólicas) capturam o processo de raciocínio dos especialistas por meio de modelos computacionais (as ontologias) que, ao serem alimentadas a um engine de inferência (ou reasoner), são capazes de reproduzir como os especialistas resolvem um problema.

O framework de GC possuirá uma ontologia que captura formalmente os conhecimentos e regras dos cinco domínios de gestão pública: clínico-epidemiológico, técnico-administrativo, normativo, gestão compartilhada e gestão do conhecimento. Essas ontologias são especificadas em OntoUML e implementadas em OWL e SWRL (Seção 3.2).

As ontologias em OWL serão inseridas em *triple stores*, compondo a T-Box da *knowledge base* (KB) do framework. A população das ontologias é feita por meio de scripts ETL, os quais ligam a KB ao *datawarehouse* e demais fontes externas. O papel desses scripts é recuperar as informações, transformá-las em triplas (o formato da ontologia) e carregá-las na base, formando assim o A-Box da KB. Assim que a KB contém as regras e indivíduos, ela já é capaz de realizar inferências, que são geradas a partir de um *reasoner*.

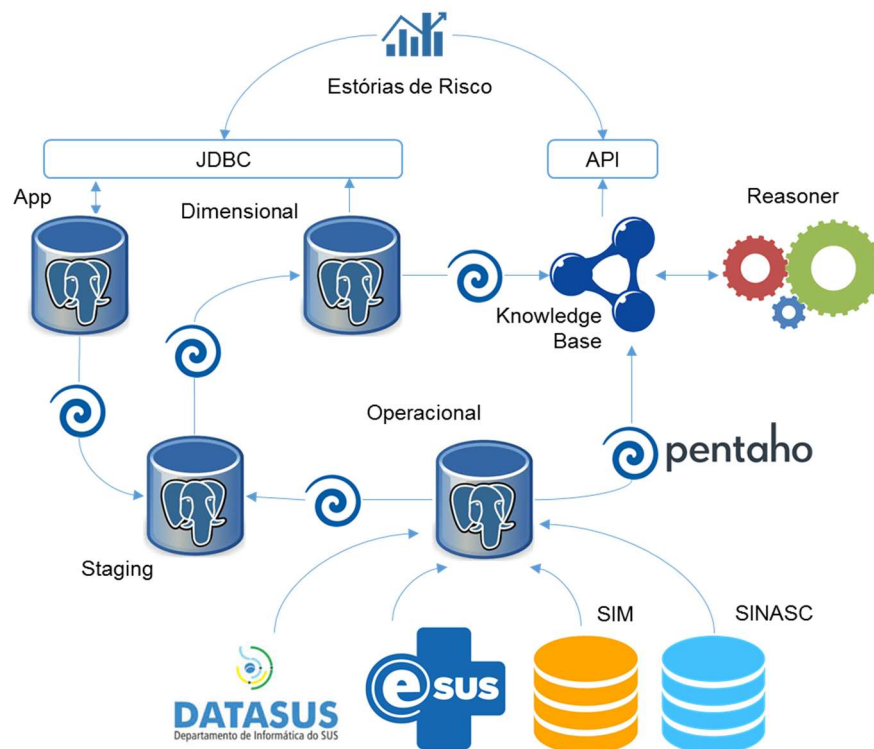


Figura 4. Componentes para a realização de inferências.

Os componentes supracitados, bem como suas relações com os demais componentes do GISSA são apresentados na Figura 4.

Uma funcionalidade inteligente que usa a inferência por regras é o relatório de risco de óbito materno, cuja interface é apresentada na Figura 5. O objetivo da estória é fornecer aos gestores de saúde municipais uma visão do risco de uma morte materna ocorrer, identificando as potenciais causas. Isso é feito por meio da análise de um conjunto de características clínicas, sociais e do sistema de saúde que afetam uma gestante ou puérpera.

Entende-se por fator *clínico* de risco aqueles relacionados a características biológicas da gestante, bem como seu quadro clínico histórico e atual. O fato da mãe ter menos de 15 anos é uma propriedade biológica. O fato dela já ter tido abortos no passado é um fator histórico clínico e ela ter adquirido hipertensão na atual gestação corresponde ao seu quadro clínico atual.

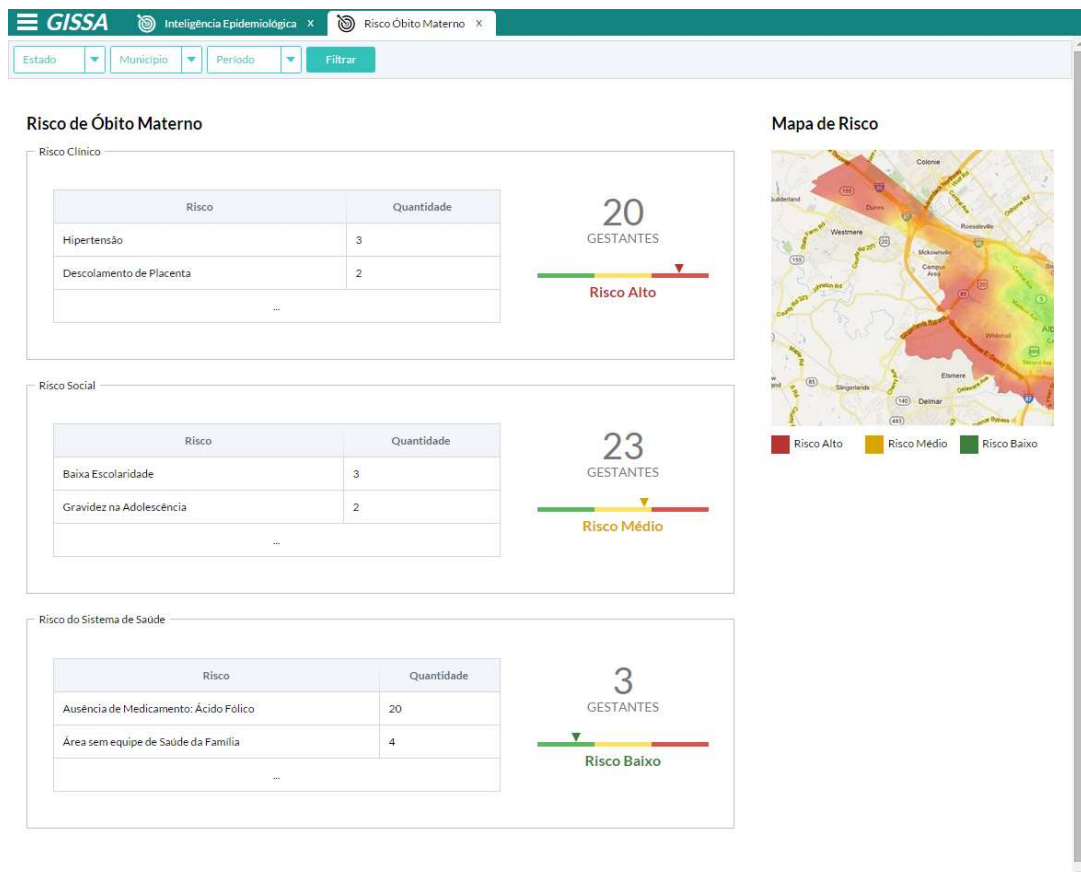


Figura 5. Protótipo de alta fidelidade para a User Story Exibir Detalhes do Risco de Óbito Materno [25].

Por sua vez, fatores de risco *social* são aqueles relacionados ao contexto social da gestante e de sua família. A baixa escolaridade e cultural de uma gestante é considerada um risco por que ela pode não ter conhecimentos básico da gravidez e tomar ações que a põe em risco, como consumir bebidas alcólicas. A baixa renda da família da qual a gestante faz parte também é um risco social porque isso pode implicar na ausência de recursos para uma alimentação adequada.

Finalmente, os fatores de risco do *sistema de saúde* são aqueles advindos da ausência ou precariedade dos serviços providos pela saúde pública brasileira. Em outras palavras, eles correspondem a limitações do sistema que podem colocar a vida de gestantes em risco. Alguns exemplos são a ausência de leitos

especializados, baixa cobertura vacinal de gestantes e estoque de medicamentos insuficiente.

No protótipo da Figura 5 o gestor de saúde poderia visualizar o número de gestantes por grupo de risco (social, clínico e do sistema de saúde), o nível desse risco (baixo, médio ou alto), bem como o detalhamento do número de gestantes por cada fator de risco específico. No lado direito da o relatório prevê a distribuição geográfica do risco de óbito materno, destacando em vermelho as regiões com maior risco e em verde as de menor risco.

5.2 Bloco Busca Semântica

O GISSA trabalha com uma grande massa de dados estruturados contidos nos diversos sistemas de saúde pública, mas também requer o acesso à uma vasta gama de artefatos contenedores de informações não estruturadas, principalmente em formatos textuais, como protocolos de saúde, atas, deliberações, leis e portarias.

Busca semântica é um nome genérico dado a uma classe de soluções computacionais voltada justamente para auxiliar a gestão de informações não estruturadas. No contexto do GISSA, a busca semântica pode ser dividida basicamente em dois processos: anotação e recuperação de artefatos.

Anotação

Anotar um artefato significa utilizar uma *knowledge base* para registrar que seu conteúdo ou parte dele possui uma determinada relação com uma entidade (ex: classe, indivíduo, propriedade...). Essa anotação pode ter ou não uma semântica definida. Por exemplo, pode-se anotar uma lei apenas com uma relação de citação com as entidades do domínio. Nesse caso, utilizando o código civil como exemplo, poderia se dizer que ele cita a Lei 12.010 e a Lei 11.107.

Caso houvessem diferentes tipos de anotação, poderia se dizer que o código civil possui um inciso *revogado* pela Lei 12.010 e outro inciso *alterado* pela Lei 11.107. Ressalta-se ainda que o processo de anotação pode ser feito de forma automática ou manual, ou ainda uma combinação desses. Anotações manuais são feitas por usuários do sistema, enquanto processos automatizados utilizam técnicas de processamento de linguagem natural (no caso de artefatos textuais) para, juntamente com a ontologia, detectar automaticamente as entidades e relações expressas no artefato analisado.

Recuperação de Artefatos

Após anotados, os artefatos podem ser recuperados utilizando as entidades da *knowledge base*. Nesse momento, as relações definidas na ontologia contribuem muito na qualidade da busca e navegação pelos documentos.

Para exemplificar esse valor, considere que um gestor de saúde queira recuperar todas as deliberações da Comissão Intergestores Tripartite (CIT) que impliquem seu município. Isso quer dizer: ele quer todas as deliberações que afetem seu município, a região de saúde da qual seu município é parte, sua macrorregião de saúde, seu estado e região geográfica. Uma busca semântica utilizaria as relações de composição entre as divisões territoriais e combinação com as relações de citação entre deliberações e unidades territoriais para fornecer a resposta.

Os componentes do framework de GC responsáveis pela busca semântica e descritos nos parágrafos anteriores estão representados na Figura 6.

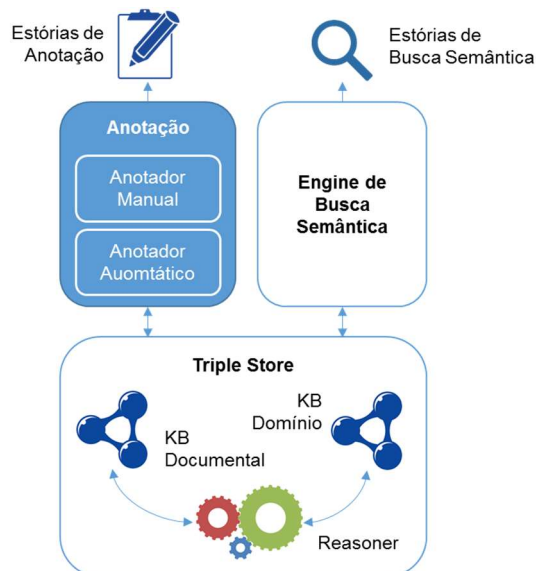


Figura 6. Componentes para realização de buscas semânticas.

5.3 Bloco Mineração de Dados

A modelo Inteligência de Gestão do GISSA compreende a adequação do Bloco Mineração de Dados ao Componente de Gestão do Conhecimento (Inteligência) apresentado na figura 3 (Visão lógica do GISSA framework).

A figura 7 mostra a arquitetura do modelo de inteligência do bloco Mineração de Dados proposta para o framework GISSA.

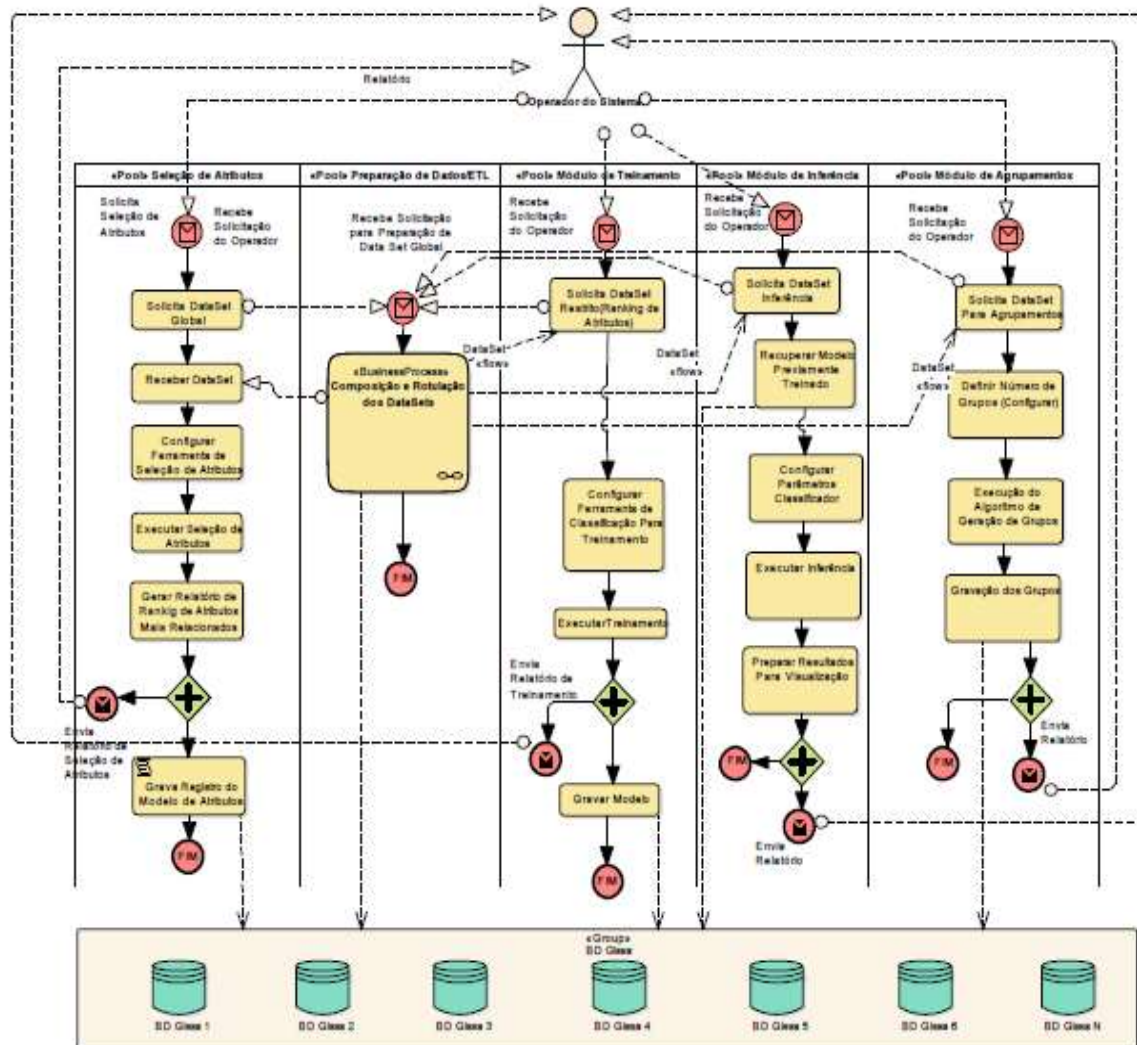


Figura 7. Modelo de Inteligência de Mineração de Dados do GISSA

Este modelo de inteligência é constituído pelas seguintes “pools”:

- Seleção de Atributos
- Preparação de dados / ETL
- Módulo de Treinamento
- Módulo de Inferência
- Módulo de Agrupamentos

O modelo de inteligência do bloco Mineração de Dados proposta para GISSA é resultado de uma metodologia descrita a seguir, especialmente desenvolvida pelo grupo de Reconhecimento de Padrões e Aprendizagem de Máquina, do laboratório Centauro da Universidade Federal do Ceará (fig. 8).

Para a utilização do modelo será realizada a análise para aquisição, formatação e estruturação dos dados para a base de dados de data mining. Este passo possibilitará a construção dos modelos de inteligência computacional e a realização dos experimentos de data mining. Nesta atividade, também serão definidos as categorias a que cada padrão poderá pertencer.

O desempenho de métodos de data mining depende, além das diferenças entre os algoritmos e os parâmetros utilizados, da seleção dos atributos a serem utilizados pelo classificador. A seleção dos atributos será realizada a partir de experimentos e análises de avaliação de desempenho de algoritmos de seleção de atributos a fim de obter um conjunto de características que melhor descrevem os dados do modelo.

O passo seguinte compreende a análise comparativa dos algoritmos de reconhecimento de padrões da literatura. Compreende-se nesta etapa a definição dos modelos de aprendizagem de máquina que serão utilizados bem como as abordagens de aprendizagem a serem adotadas (supervisionada, não supervisionada, semisupervisionada, etc). Além disso, será analisado se e como o sistema deverá ou não aprender após a inclusão de novas informações.

O framework open source Weka (FRANK; HALL; TRIGG, 2000) contém uma extensiva coleção de algoritmos de aprendizagem de máquina e será inicialmente utilizado para a avaliação e testes dos algoritmos de inteligência computacional. Para a realização de predições/classificações, é necessário que o tempo de resposta seja viável e em tempo real.

Será realizada a avaliação das técnicas de ML selecionadas quanto a viabilidade de classificação em tempo real. Em seguida, será realizado o estudo de viabilidade de tecnologias de machine learning para incorporação no GIS (frameworks e bibliotecas).

Nesta etapa, será verificado a possibilidade de reuso de ferramentas de data mining / inteligência computacional ou utilização de implementações específicas de algoritmos de Inteligência computacional.

Os entregáveis deste projeto são a lista de atributos mais relevantes e os algoritmos de aprendizagem de máquina que deverão ser utilizados para cada indicador/alerta/relatório do GISSA.

A figura 8 mostra as diversas etapas da metodologia desenvolvida pelo grupo de Reconhecimento de Padrões e Aprendizagem de Máquina da UFC, da qual resultou o Modelo de Inteligência de Mineração de Dados proposto para o GISSA (fig. 7):



Figura 8. Metodologia de Reconhecimentos de Padrões desenvolvida na UFC

1. NIVELAMENTO: Nesta fase é realizado um treinamento com todos os envolvidos. Trata-se de um treinamento básico sobre reconhecimento de padrões e aprendizagem de máquina. Ao fim do treinamento toda a equipe estará falando o mesmo "idioma".
2. PROSPECCAO INICIAL: Os representantes da empresa (cliente) apresentam claramente o problema. Esta apresentação deve ser repetida até que todos estejam de acordo sobre os desafios.
3. SELECAO INICIAL: Nesta fase são selecionados ferramentas e algoritmos a serem testados. Podem ser selecionados classificadores elementares, redes neurais, comitês de classificadores, etc.
4. PREPARAÇÃO DOS DADOS: Dependendo como os dados estão disponíveis é possível que se tenha que voltar ao passo 3.
5. TRANSFORMAÇÃO DOS DADOS: Modificação nos dados para submissão à ferramentas de análise.
6. EXPERIMENTOS: Primeiros experimentos com as ferramentas selecionadas;
7. GANHADORA: Identificação do algoritmo de Machine Learning que comporá a tecnologia solução.

8. **CONSTRUÇÃO:** Nesta fase é construída uma ferramenta de software contendo uma solução fechada para o problema proposto. Aqui usa-se Engenharia de Software convencional.
9. **EXPERIMENTOS:** Realização de experimentos com a ferramenta desenvolvida.
10. **TESTES:** Realização de testes de integração e testes unitários
11. **ENTREGA DO PRODUTO DE SOFTWARE:** O módulo Bloco Mineração de Dados estará pronto para ser incorporado ao componente de Gestão do Conhecimento do framework GISSA (fig 3).

O modelo a ser proposto deverá especificar claramente como os dados serão preparados e formatados. A partir destes dados, a avaliação de desempenho dos algoritmos de seleção de atributos e classificação (machine learning/reconhecimento de padrões) poderá ser realizada. Inicialmente, será criado um grafo que mostre os atributos dos bancos de dados e onde estes estão. Esta etapa tem uma interface com o grupo de Ontologias. Para esta etapa, é necessário estudar estas tabelas e compreender o significado dos campos desta tabela.

6 O GISSA COMO UM SISTEMA DE GESTÃO DO CONHECIMENTO

Embora este documento defina uma parte do GISSA como o framework de gestão do conhecimento, o framework como um todo pode ser classificado como um *Enterprise Knowledge Portal*. Essa classificação se deve ao fato do GISSA ser projetado como um grande portal de entrada para os usuários acessarem as diferentes soluções de gestão do conhecimento presentes no framework.

Como mostrado na Visão lógica do framework GISSA (fig 3), o framework de gestão do conhecimento proposto do GISSA é responsável por torná-lo uma fundação para construção de “sistemas inteligentes”, constituído por 3 blocos: Inferências por Regras, Buscas Semânticas e Mineração de dados. A inteligência, nesse contexto, refere-se à capacidade de um agente computacional gerar novas informações a partir de outras contidas no sistema, bem como a capacidade de utilizar o conhecimento formalizado para melhorar as funções do sistema.

O componente de datawarehouse é um grande repositório de memória organizacional. Ao armazenar uma grande quantidade de informações relevantes para os gestores públicos de saúde, o GISSA permite que os usuários analisem o impacto de ações passadas e recuperem conhecimentos de forma ágil e em tempo oportuno.

Os relatórios, indicadores e dashboards que são construídos a partir de técnicas de inferência por regras classificam o GISSA como um sistema de apoio a decisão, em particular, um sistema especialista. A construção das ontologias que capturam o conhecimento dos especialistas, o framework contribui para a externalização do conhecimento. Ao aplicar esse conhecimento em novos cenários (como o de risco de óbito materno), o sistema alerta usuários sem esse conhecimento que problemas existem e aponta como solucioná-los.

No caso dos relatórios produzidos a partir de inferências realizadas por mineração de dados, eles também se enquadram na categoria de sistemas de apoio a decisão.

Finalmente, as funcionalidades de anotação e busca semântica se enquadram no grupo dos sistemas de gestão de conteúdo. Ao permitir a recuperação de protocolos de saúde, o GISSA auxilia no processo de disseminação do conhecimento no sistema público de saúde, uma vez que esses contêm conhecimentos clínicos sobre como proceder em determinadas situações. Ao contribuir na recuperação de leis, portarias e deliberação, a contribuição é a mesma da anterior, dessa vez orientada ao conhecimento normativo e de gestão compartilhada.

7 CONCLUSÃO

O GISSA é um framework para construção de soluções computacionais que ajudem os usuários do sistema de saúde a tomarem melhores decisões, evitando o desperdício de recursos públicos e diminuindo o número de óbitos maternos e infantis. O framework de gestão do conhecimento é a parte dessa solução que adiciona mecanismos de inteligência ao produto, utilizando técnicas de inteligência artificial e inteligência computacional para tal.

Neste documento foram abordados, em especial, a Inferências por Regras, a Buscas Semânticas e a Mineração de dados, os três componentes que compõem o Modelo de Inteligência da Gestão de saúde do framework GISSA. Foram discutido suas relações com os demais componentes e apresentando as funcionalidades que eles permitem adicionar à solução final do projetoB.

Embora o GISSA tenha sido iniciado como um projeto de dois anos, já se prevê um longo caminho a percorrer após esse período. Nessa direção, já são consideradas diversas novas funcionalidades, inclusive do framework de gestão do conhecimento:.

Uma delas é o suporte a colaboração entre os usuários no contexto de alertas. Isso significa que ao ser gerado um alerta de nascimento, por exemplo, um gestor poderia enviar uma mensagem contextualizada ao agente de saúde responsável pelo recém-nascido e pela puérpera, perguntando sobre a previsão de visita, a qual poderia ser respondida com informações sobre a criança. Essa nova funcionalidade agregaria uma nova característica de gestão do conhecimento, o suporte à colaboração (ou a classificação como sistema *Groupware*).

A segunda projeção refere-se a um componente do framework de GC que torne o GISSA também um sistema de recomendação. Isso funcionaria da seguinte forma: ao identificar que algum indicador está com problemas, como alto número de partos prematuros ou baixa cobertura de uma determinada vacina, o GISSA recomendaria ao gestor de saúde ações (entre as muitas possíveis) para mitigar tais problemas.

Por último, ressalta-se que pelo fato do GISSA ser um projeto de inovação, o framework está em constante evolução. Ao se descobrir novos problemas e anseios da saúde pública brasileira, novas soluções são incluídas ao framework. O objetivo, no entanto, não muda: tornar o sistema de saúde mais eficiente, ajudando os usuários a tomarem decisões melhores.

8 REFERÊNCIAS

- [1] G. Fischer and J. Ostwald, “Knowledge Management: Problems, Promises, Realities, and Challenges,” *Intell. Syst.*, vol. 16, no. 1, pp. 60–72, 2001.
- [2] M. Alavi and D. E. Leidner, “Knowledge Management Systems: Issues, Challenges, and Benefits,” *Commun. Assoc. Inf. Syst.*, vol. 1, no. 2es, 1999.
- [3] I. Nonaka and H. Takeuchi, *The Knowledge Creating Company: How Japanese Companies Create the Dynamics of Innovation*. New York, USA: Oxford University Press, 1995.
- [4] J. Q. Chen, T. E. Lee, R. Zhang, and Y. J. Zhang, “Actualizing Organizational Memory with Information Systems,” *Commun. ACM*, vol. 46, no. 12, pp. 73–78, 2003.
- [5] R. S. S. Guizzardi, “Agent-oriented Constructivist Knowledge Management,” University of Twente, 2006.
- [6] C. A. Ellis, S. J. Gibbs, and G. L. Rein, “Groupware: Some issues and experiences,” *Commun. ACM*, vol. 34, no. 1, pp. 39–58, 1991.
- [7] B. E. Dixon, R. E. Gamache, and S. J. Grannis, “Towards public health decision support: a systematic review of bidirectional communication approaches,” *J. Am. Med. Inform. Assoc.*, vol. 20, no. 3, pp. 577–583, Mar. 2013.
- [8] C. A. M. Bastos, L. Rezende, M. Caldas, A. S. Garcia, S. Mecena Filho, M. L. Sanchez, and J. de L. P. Castro Junior, “Building up a Model for Management Information and Knowledge: The Case-Study for a Brazilian Regulatory Agency,” in *International Workshop on Software Knowledge (SKY)*, 2011.
- [9] C. A. M. Bastos, A. C. M. Bruno, A. Garcia, L. Rezende, M. A. F. Caldas, M. L. Sanchez, and S. J. Mecena Filho, “Managing Information and Knowledge: a proposal methodology for building an integrated model based on information assets identification,” in *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval and the International Conference on Knowledge Management and Information Sharing*, 2013, pp. 520–525.
- [10] G. Guizzardi, “Theoretical foundations and engineering tools for building ontologies as reference conceptual models,” *Semant. Web J.*, vol. 1, no. 1,2, pp. 3–10, 2010.
- [11] G. Guizzardi, “Ontological Foundations for Structural Conceptual Models,” Centre for Telematics and Information Technology, University of Twente, Enschede, The Netherlands, 2005.
- [12] G. Guizzardi, A. P. das Graças, R. S. S. Guizzardi, and A. Pinheiro, “Design Patterns and Inductive Modeling Rules to Support the Construction of Ontologically Well-Founded Conceptual Models in OntoUML,” in *International Workshop on Ontology-Driven Information Systems (ODISE)*, 2011, vol. 83, pp. 402–413.
- [13] G. Guizzardi and T. P. Sales, “Detection, Simulation and Elimination of Semantic Anti- patterns in Ontology-Driven Conceptual Models,” in *International Conference on Conceptual Modeling (ER)*, 2014.
- [14] P. P. F. Barcelos, V. A. dos Santos, F. B. Silva, M. Monteiro, and A. S. Garcia, “An Automated Transformation from OntoUML to OWL and SWRL,” in

- Seminário de Pesquisa em Ontologias no Brasil (ONTOBRAS)*, 2013, pp. 130–141.
- [15] OMG, “Model Driven Architecture (MDA): MDA Guide Rev. 2.0,” 2014.
 - [16] W3C, “RDF/XML Syntax Specification,” 2014.
 - [17] W3C, “OWL 2 Web Ontology Language Document Overview,” 2012.
 - [18] W3C, “SWRL: Semantic Web Rule Language - Combining OWL and RuleML,” 2004.
 - [19] W3C, “SPARQL 1.1 Overview,” 2013.
 - [20] F. Carolo and L. Burlamaqui, “Improving Web Content Management with Semantic Technologies,” in *SemTech*, 2011.
 - [21] Y. Raimond, T. Scott, S. Oliver, P. Sinclair, and M. Smethurst, “Use of Semantic Web technologies on the BBC Web Sites,” in *Linking Enterprise Data SE - 13*, D. Wood, Ed. Springer US, 2010, pp. 263–283.
 - [22] A. Swartz, “MusicBrainz: a semantic Web service,” *Intelligent Systems, IEEE*, vol. 17, no. 1. pp. 76–77, 2002.
 - [23] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives, “DBpedia: A Nucleus for a Web of Open Data,” in *The Semantic Web SE - 52*, vol. 4825, K. Aberer, K.-S. Choi, N. Noy, D. Allemang, K.-I. Lee, L. Nixon, J. Golbeck, P. Mika, D. Maynard, R. Mizoguchi, G. Schreiber, and P. Cudré-Mauroux, Eds. Springer Berlin Heidelberg, 2007, pp. 722–735.
 - [24] J. Hendler, “US Government Linked Open Data: Semantic.data.gov,” *IEEE Intell. Syst.*, vol. 27, no. 3, pp. 25–31, Apr. 2012.
 - [25] GISSA, “User Story - Exibir Detalhes de Risco de Óbito Materno,” 2015.

ANEXO 1: ESPECIFICAÇÕES DO MODELO DE INTELIGÊNCIA DO GISSA (Data Mining)

| DOCUMENTOS RELACIONADOS | SEÇÃO/SUBSEÇÃO |
|--|---------------------|
| ESPECIFICAÇÃO DE REQUISITOS NÃO FUNCIONAIS (GIS_RNF) | 4.2.1, 4.3.1, 5.1.1 |
| ESPECIFICAÇÃO DA ARQUITETURA (GIS_EAR) | 4.3, 6.1, 6.3 |

1.1 ESPECIFICAÇÃO DE REQUISITOS NÃO FUNCIONAIS (GIS_RNF)

- ✓ Subseção 4.2.1: **Acurácia da informação**
 - **Comentários:** Para a eficácia dos algoritmos de *machine learning* (aprendizagem de máquina) os dados devem ser desprovidos de erros. Os procedimentos para garantir a total acurácia da informação deverão ser analisados para verificação de corretude das mesmas.
- ✓ Subseção 4.3.1: **Tempo de Resposta Web**
 - **Comentários:** Para a realização de predições/classificações, é necessário que o tempo de resposta seja viável e em tempo real. Deverão ser investigadas abordagens de *machine learning* que atendam estes quesitos para os dados utilizados no GISS.
- ✓ Subseção 5.1.1: **Acurácia dos Dados Externos**
 - **Comentários:** Embora seja assumido que os dados estão devidamente filtrados e livres de anomalias, os procedimentos que levaram a esta suposição deverão ser devidamente analisados.

1.2 ESPECIFICAÇÃO DA ARQUITETURA (GIS_EAR)

- ✓ Seção 4.3: **Análise de opções de tecnologias e arquitetura - Tecnologia/Arquitetura: Machine Learning / Data Mining**
 - **Comentários:**
 - As ferramentas citadas (Linguagem R, Weka, etc.) devem ser avaliadas para a adequação à tarefa de classificação para o GISS.
 - O tempo de resposta e precisão média dos algoritmos de Inteligência Computacional devem ser avaliados para o conjunto de dados do GISS. Além disso, o modelo de aprendizado de máquina deve possibilitar a classificação/predição em tempo real.
 - Ainda não foi descrito como o modelo de classificação/predição, construído a partir de um algoritmo de machine learning, será atualizado (a partir de novos dados) e reavaliado.
 - É necessário um estudo preliminar para a seleção de técnicas de classificação baseadas em machine learning bem como algoritmos de seleção de atributos para os dados a serem utilizados no projeto GISS.
 - É necessário o conhecimento da estrutura das outras bases de dados para poder compor a estrutura de dados da base de dados de *datamining*.
- ✓ Seção 6.1: **Item 4: "aplica mineração de dados e inferência das informações"**
- ✓ Seção 6.3: **Data Warehouse - datamining**

- o **Comentário:** Com que frequência esta base será atualizada com novos dados? Os dados desta base serão utilizados para construir diretamente os modelos de Inteligência Computacional ou necessitarão de alguma ~~processação~~ processamento?
- ✓ Seção 6.4: " O Componente de **Inteligência** deverá suportar diversas algoritmos de mineração de dados, ontologia, estatística e outras mecanismos de inferência."

2. REQUISITOS DO BACKLOG (ASSOCIADOS À INTELIGÊNCIA COMPUTACIONAL)

| | |
|-------|---------------------------------------|
| INFRA | Montagem de Datawarehouse |
| ETL | Extração de dados <u>DataSUS</u> |
| ETL | Extração de dados de Sistemas Locais |
| ETL | Extração de dados de contexto externo |
| ETL | Motor de Transformação de dados |

Observação: Não há no backlog um requisito que esteja diretamente vinculado à incorporação dos algoritmos de inteligência computacional.