

### 1.8.3 Metodologia 101

Este método foi proposto por Natalya Noy e Deborah McGuinness e se apresenta como um processo voltado para iniciantes, com o objetivo de simplificar as etapas formadoras de uma ontologia. Nele os usuários poderão ter um facilitador a mais na complexa tarefa de elaboração ontológica. É chamado de 101, pois este número é usado nas disciplinas iniciais na Universidade (Roque, 2009). Segundo (Roque, 2009) o método 101 faz a proposta de que a construção da ontologia siga uma sequência de passos não lineares, pois se trabalha uma modelagem voltada para a prática, “...sendo realizadas inúmeras iterações até alcançar o modelo adequado, para então ser validado” (Roque, 2009). Os passos sugeridos por Noy e McGuinness são os seguintes:

Passo 1: Determinar o domínio e o escopo da ontologia: É a fase onde ocorre o planejamento geral sobre o tema que o estudo tratará. Basicamente se procurará responder a 5 perguntas: i) Qual o domínio que a ontologia cuidará?; ii) Para que ela será criada?; iii) Para quais indagações a ontologia deverá fornecer respostas?; iv) Quais serão os seus usuários? e v) Quem executará a sua manutenção?.

Passo 2 : Verificar a possibilidade de reuso de outras ontologias : Se existirem ontologias pertencentes a área de domínio que será estudada, deverá se levar em consideração o seu reuso nesta nova construção. Há repositórios de ontologias disponíveis na *Web* entre eles pode-se citar: *Daml Ontology*, *Damoz*, *Ontolingua Server*, *RosettaNet*, *Universal Repository*, *University Stanford (Protégé)*, *UNSPSC* entre outros.

Passo 3 : Enumerar os termos relevantes da ontologia : Inicia-se uma análise do domínio que se pretende mapear, através da coleta de termos que se relacionam com o tema central. Uma lista deles deve ser elaborada agrupando-se os mais significativos. Para que se mantenha fidedignidade em relação ao que ocorre no mundo real, estes termos devem ser selecionados por especialistas da área, ou por fontes balizadoras do conhecimento que se deseja transformar em ontologia. Ainda neste item, deve-se começar a investigar possíveis classes, superclasses, relações e propriedades entre os termos presentes na lista.

Passo 4 : Definir classes e suas hierarquias : Existem 3 maneiras de se realizar esta fase: Por *top-down* (de cima para baixo), *bottom up* (de baixo para cima) ou usar ambas combinando-

as. Após decidir qual o sentido adotado para se efetuar a definição pode-se subdividir o restante do passo em 3 etapas:

- Termos que funcionam como objetos e possuem existência independente em relação aos demais: Estes podem se tornar classes do processo de construção da ontologia.
- Termos que não se enquadram no item imediatamente anterior, poderão ser entendidos como instâncias dos objetos já mencionados ou mesmo como propriedades deles.
- Poderá ocorrer descarte de termos, por constatação de redundância no uso do vocabulário.

Passo 5 : Definir as propriedades das classes : Todos os termos da primeira lista que não foram descartados por redundância, nem se tornaram classes, serão aproveitados como propriedades. Poderão estar representadas como atributos das classes ou explicitando relacionamentos entre elas.

Passo 6 : Definir as características de cada propriedade : Neste item as propriedades serão preenchidas com valores diversos em conformidade com sua representação semântica dentro do domínio. As propriedades poderão ser preenchidas com *strings*, números (*integer*), booleano, enumerados e instâncias.

Passo 7 : Criação de instâncias : Na derradeira etapa do método, será a vez de se preencher tanto classes como subclasses com instâncias ou indivíduos. É o “povoamento” da ontologia. Para que um indivíduo passe a figurar como membro da ontologia é necessário cumprir três condições essenciais. São elas:

- Ser instância de alguma classe ou subclasse;
- Ser distinto de todos os outros indivíduos e
- Ter suas propriedades devidamente preenchidas.

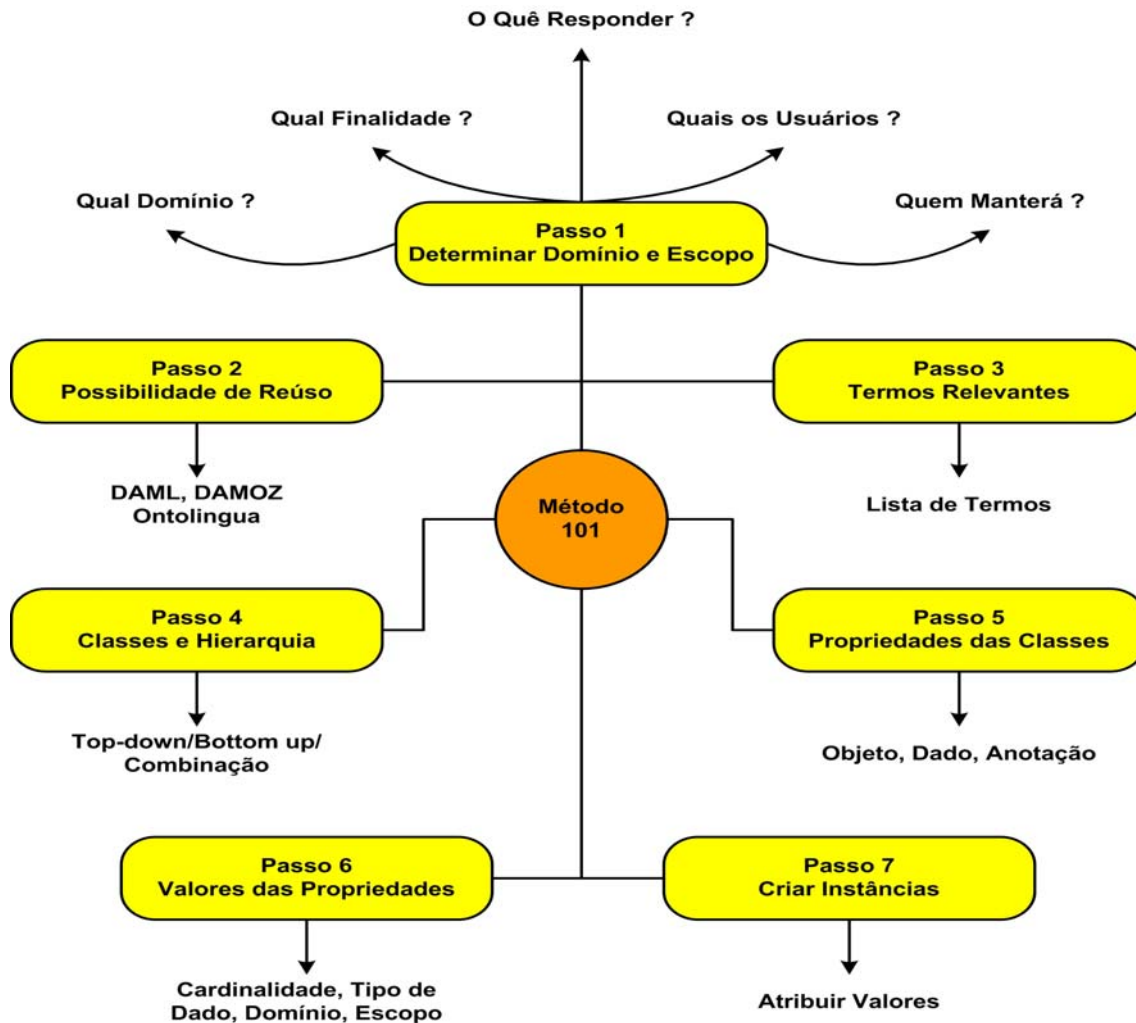


Figura 5 – Processamento do Método 101

## 1.9 Editores de Ontologia

Uma das características mais desejáveis em um editor é a sua usabilidade e a clareza na manipulação por parte do usuário, dos conjuntos de classes e propriedades que são as partes principais de uma ontologia. Nas diversas etapas que compõem a elaboração e montagem de uma ontologia são usados itens variados como: linguagens de construção ontológica, motores de inferência e os ambientes de desenvolvimento. O uso destes ambientes é importante para que ocorra a “disponibilização de ontologias para reuso e extensão, integrando diferentes grupos de pesquisa, amiúde distantes geograficamente, que pesquisam sobre as mesmas áreas ou áreas afins” (Freitas, 2003). Dentre as principais ferramentas para edição de ontologias, encontra-se o framework Protégé, que será usada para representar a ontologia sobre chuvas e suas consequências, utilizada no exemplo apresentado no capítulo 3. Originalmente, o Protégé foi desenvolvido pelo setor de informática da Universidade de Stanford, para ser uma

ferramenta de aquisição de conhecimento voltada para a área da oncologia. Com o passar dos anos o ambiente revelou potencial para ser usado em outras áreas que não a médica. Atualmente possui código aberto e uma comunidade com quase 100.000 usuários no mundo todo. Isto se deve ao fato de estar constantemente sendo modernizado e melhorado para atender os avanços dos sistemas baseados em conhecimento. Possui sua construção nativa na linguagem Java e diversos *plugins* integráveis que enriquecem o seu uso e agregam funcionalidades ao framework. Um dos mais utilizados é o Jambalaya, que permite a representação das classes em um diagrama baseado em herança (classe-pai, classe-filha), que atua em conexão com o GraphViz. Este plugin foi utilizado nesta dissertação como será visto no capítulo 3. Protégé tem a capacidade de gerar ontologias em diversos formatos., sendo os mais comuns o RDF, o XML e o OWL. Outras características do Protégé são:

- Utiliza a *Protégé Axiomatic Language* (PAL). Ela admite a colocação de sentenças axiomáticas e restrições, para as classes, as instâncias e para os seus atributos.
- A consistência da ontologia pode ser testada pelo próprio *software*, que possui o Pellet como um dos motores de inferência que podem ser adicionados ao programa, entre outros capazes de serem usados para este mesmo fim.
- Possui uma interface, acoplada a um gerador de formulários para as classes existentes, facilitando ao usuário, realizar a entrada do conhecimento.
- Ambiente open source e interativo, podendo ser baixado pelo site da Universidade de Stanford: <http://protege.stanford.edu>.
- Por ser uma ferramenta gráfica e visual, contribui para que atenção do usuário esteja focada sobre os relacionamentos entre as classes, especificações do domínio e restrições dos atributos, e não nos códigos a serem digitados.
- Permite que sejam criadas superclasses, classes e subclasses. As classes representam os conceitos do domínio que se está estudando.
- Apresenta a visualização da ontologia, disposta pelo sistema de herança de classes, utilizando o plugin Jambalaya.
- Admite a criação de slots, também chamados de propriedades. Eles são os atributos dos objetos da ontologia.
- As classes podem ser instanciadas (exceção para as classes abstratas). Oferece um ambiente de formulação e gravação de consultas.

Considerando esta fundamentação teórica, foram pesquisados na literatura disponível, trabalhos que tratassem da extensão dos serviços atuais de *Feed* e/ou que propusessem imprimir semântica às buscas na *Web*, usando ontologias.

### 1.10 Trabalhos Relacionados

Atualmente, o grande problema que decorre da falta de estruturação de dados contidos na *Web* é o de como prover meios que forneçam respostas relevantes às buscas efetuadas pelos usuários, principalmente as que ocorrem mediante a utilização dos chamados filtros de busca como Google, Yahoo e Excite. Ainda que esta dificuldade pudesse ser completamente solucionada com o uso de programas, métodos, e agentes inteligentes haveria uma outra apresentada em Teevan et al (2004), que aponta que nem sempre os usuários *Web* expressam de maneira adequada o que desejam achar em suas páginas. Freitas (2002) menciona a sobrecarga de informação na Internet como ingrediente que gera outros obstáculos para o usuário. Entre eles, cita as dificuldades que os mecanismos de busca enfrentam diante da polissemia (uma palavra que possua diversos significados); a página-resultado, oferecida ao usuário apresenta links heterogêneos (que não versam sobre o mesmo assunto) e a escolha de palavras que embora tenham a ver com o que se está buscando não são as mais descritivas dentro do tema procurado.

Freitas (2002) aponta a inexistência de uma extração integrada de informações via *Web*, especificando que os sistemas de extração ainda atuam sobre domínios muito restritos. O autor propõe um mecanismo que opere esta integração na extração de informações, que sejam relevantes para o usuário. Assim são acopladas três etapas chamadas complementares e constituídas como recuperação, categorização e extração. Para este autor (2002, p.9-10):

Sistemas de recuperação podem fornecer o acesso a um conjunto inicial de páginas, que possui alta cobertura e baixa precisão. Após isto, sistemas de categorização deveriam selecionar quais páginas pertenceriam às classes a serem processadas, e então extratores poderiam capturar a informação requerida. Durante o processo de extração, poderiam ainda ser encontrados, dentre as âncoras das páginas processadas, endereços de outras páginas que pertenceriam a outras classes também processadas. Estas sugestões são encontradas dentro de um contexto muito mais semântico e seguro, e, assim, o processo de extração estaria auxiliando a recuperação. A extração poderia ainda refinar a categorização, uma vez que páginas que não contivessem os dados que caracterizam a classe processada seriam desprezadas.

O autor oferece como solução ao problema da falta de relevância a construção de uma arquitetura de sistemas multiagentes cognitivos chamada *MASTER-Web* (Multi-Agent

System for Text Extraction, Classification and Retrieval over the *Web*) baseado em ontologias e que se propõe a extrair informações de um grupo de páginas específico, que pertencem a uma classe, como a classe do conhecimento científico, por exemplo.

#### 1.10.1 Classify

Faria e Lima (2008) apresentam *Classify*, um programa que tem como objetivo principal extrair informações a partir da análise de RSS. Sua proposta é classificar automaticamente as notícias oriundas da *Web*. Para isto, o trabalho utiliza técnicas de inteligência artificial, com o uso de redes neurais e de mineração de dados. O projeto para realizar o treinamento de seus índices usa a ferramenta Weka (Waikato Environment Knowledge Analysis), produzida em Java. O sistema é composto por dois módulos: o *classifybot* e o *classify*. O primeiro é responsável pelo recebimento das informações via RSS e pelo consequente treinamento do conjunto. Já o segundo funciona via *Web* atendendo aos colaboradores permitindo interação entre os usuários inclusive dando-lhes condições de realizarem os procedimentos específicos no módulo *Classifybot*.

#### 1.10.2 ePaper

Tenenboim, Shapira e Shoval (2008) apresentam o *ePaper*, que é um jornal eletrônico para dispositivos móveis. Seu objetivo é trazer ao usuário notícias de diversas fontes (como jornais digitais, serviços via *Web* e revistas), de maneira que possibilite ao usuário construir o seu próprio jornal agregando as notícias como desejar. O *ePaper* é uma aplicação cliente-servidor e seu funcionamento se estabelece destes dois lados da seguinte forma: i) no lado do servidor, as notícias são recebidas e agregadas ininterruptamente, utilizando os provedores de notícias via RSS. Quando as notícias chegam ao servidor, o programa classifica cada item de acordo com os conceitos, baseados em uma ontologia de notícias. A partir daí um algoritmo aplica os critérios de relevância sobre elas, distribuindo-as (enviando-as para os dispositivos móveis) conforme as pré-definições de cada usuário. Esta etapa também utiliza conteúdo colaborativo e filtros para a informação; ii) já no lado do cliente, o programa se desdobra em um layout intuitivo que proporciona facilidades de navegação e rápida configuração do ambiente.

A estrutura do servidor é composta por cinco camadas, especificadas pelos seguintes itens: Agregador, Gerenciador de Conteúdos, Personalizador, Serviços de Entrega de

Conteúdo e Ferramentas de Gerenciamento do Sistema. A arquitetura em questão pode ser vista na figura 6.

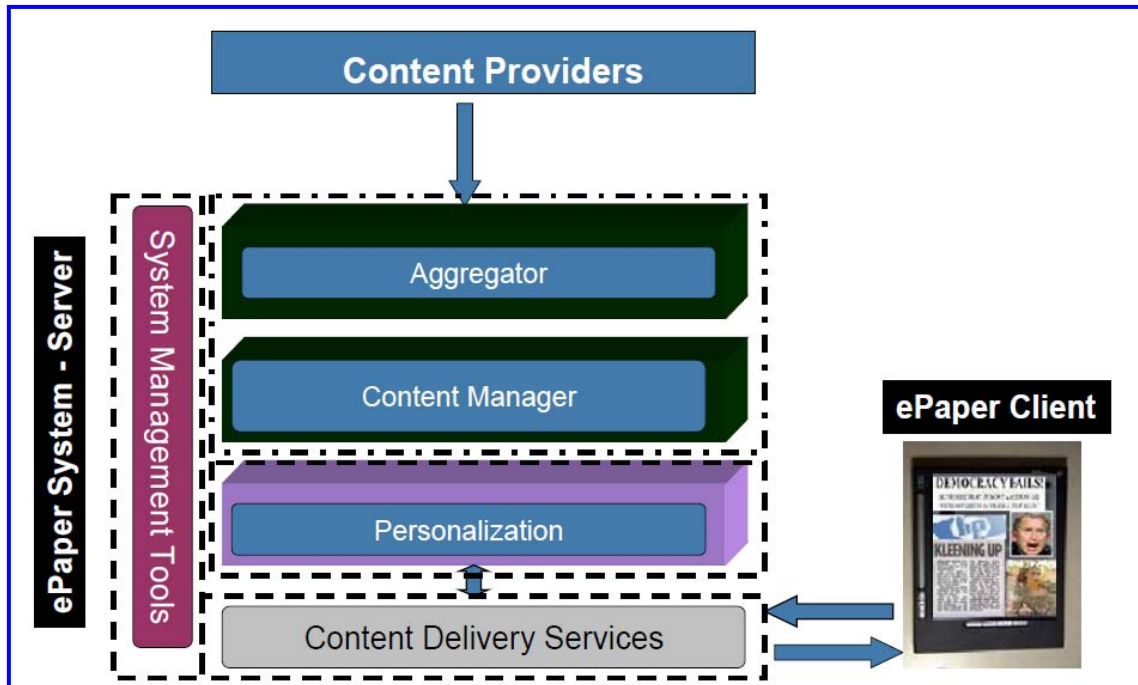


Figura 6 – Sistema ePaper

O Agregador trabalha recebendo as notícias via RSS e as estocando no repositório do ePaper. Os novos itens ou novas notícias armazenadas são resultados da interação do Agregador junto aos provedores de informação. O Gerenciador de Conteúdos é o alvo do Agregador, recebendo dele todos os novos itens recém chegados. Seu principal objetivo é operar a classificação dos itens, através de conceitos que possam melhor representá-los. Nesta etapa, o Gerenciador faz uso de um algoritmo que utiliza como apoio uma ontologia composta de 1400 conceitos e subdividida em três níveis de hierarquia além de códigos taxonômicos do IPTC ([www.iptc.org](http://www.iptc.org)). Depois de devidamente classificada, a notícia é entregue para a Personalização, que aplica sobre ela algoritmos que medirão relevância e similaridade de sua classificação no ePaper, com as classificações de importância, feitas por cada usuário do sistema. A camada composta pelos Serviços de Entrega de Conteúdos é responsável pela interação entre a Personalização e o usuário. É nela que os pedidos dos clientes chegam e são submetidos à Personalização para então serem devolvidos e entregues aos usuários registrados. É também aqui que ocorre o monitoramento das ações do usuário. Estes dados são enviados à Personalização que se incumbe de atualizar o perfil do leitor para futuros envios de notícias, que mais se adequem aos seus usos do aplicativo mais recentes.

Na última camada, o programa apresenta algumas ferramentas voltadas para administração do projeto dentre elas destacam-se uma seletora dos conteúdos jornalísticos a serem recebidos e uma que edita as ontologias, permitindo sua manutenção e conseqüente evolução.

### 2.10.3 Hermes

Desenvolvido por Borsje et al (2007), o sistema tem como objetivo fornecer ao usuário condições de estabelecer relações com as notícias, que chegam ao programa via RSS, ligando-as aos conceitos que desejam encontrar. Todo este mecanismo é sustentado por uma ontologia de domínio, que versa sobre 19 empresas componentes da bolsa eletrônica NASDAQ. Caracteriza-se por trabalhar relacionamentos e conceitos nas notícias para que itens relevantes aos usuários sejam recuperados, após serem semanticamente analisados. As principais tecnologias utilizadas no projeto são:

- Linguagens RDF e OWL, além do dialeto de XML, RSS, para o recebimento das notícias.
- SPARQL para operar consultas, no modo ARK, específico para a API Jena.
- Linguagem Java para construção do programa.
- Aplicativo OntoSem para analisar as notícias e o SemNews para ser o motor de processamento de linguagem natural.
- O Artequakt é usado para se extrair conhecimento das páginas *Web*. Segundo Alani et al (2002), é um sistema especialista voltado para o universo artístico e é responsável pela compilação automática de biografias.
- O Protégé foi o editor escolhido para se montar a ontologia do domínio NASDAQ.
- Biblioteca Java – Informa agrega as notícias em um objeto. Em seguida a API Jena carrega, edita e grava a ontologia no sistema.
- O projeto WNWA, é o responsável por fornecer os sinônimos para o Hermes, através do envio de um arquivo XML.
- Para operar-se a visualização dos termos e dos conceitos selecionados, foi usada a biblioteca gráfica Java Prefuse.

A ontologia das 19 empresas está montada em três distintos níveis: i) base do conhecimento (conceitos relevantes dentro do domínio); ii) classificação de notícias e iii)



relações entre os conceitos e as notícias. O programa possui uma interface que permite ao usuário escolher as notícias por períodos de tempo e pelos conceitos, conforme desejar segundo uma lista apresentada a ele. Após o recebimento dos feeds, o classificador de notícias será iniciado e somente acrescentará as notícias que possuírem pelo menos dois conceitos relacionados com a ontologia. O sistema proporciona a visualização dos itens e permite que o usuário altere configurações de como são apresentados no gráfico, e sobre quais conceitos deseja realizar sua busca, conforme mostra a figura 7:

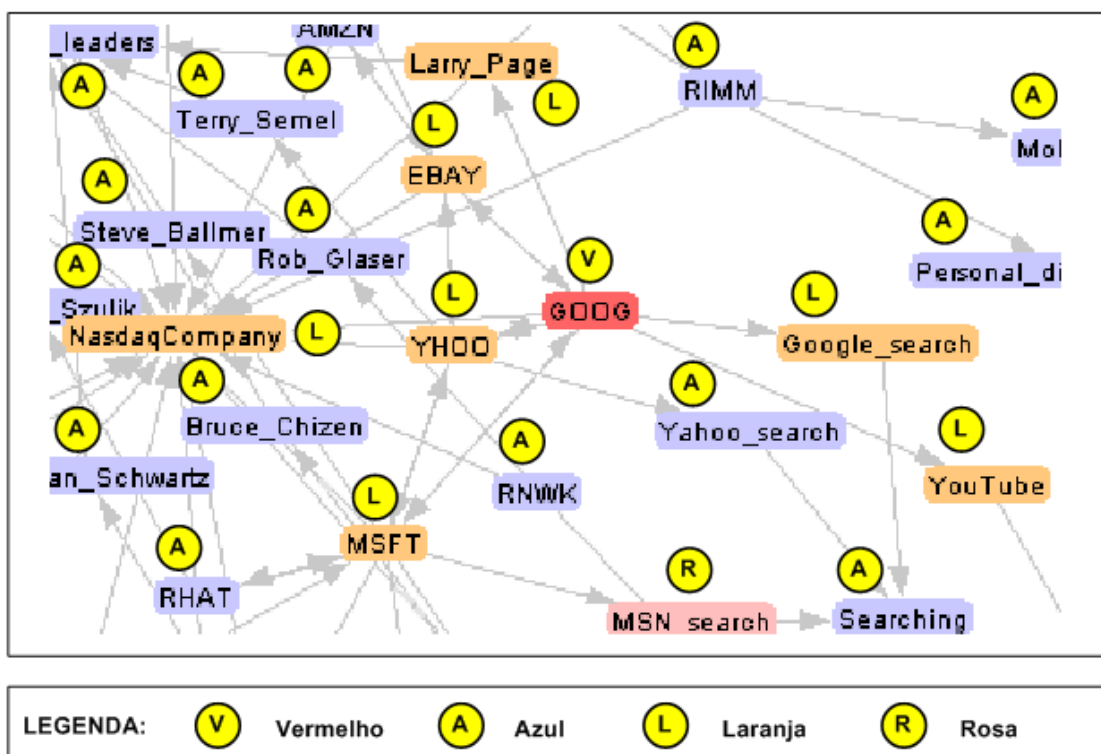


Figura 7 – Funcionamento do Sistema Hermes

As diferentes cores usadas nos nós da visualização indicam nó selecionado (vermelha), os nós relacionados (cor laranja), os que aparecem como resultados de uma busca por palavras-chave (cor rosa) e os demais são nós que não apresentam nenhum significado especial dentro do domínio (cor azul). O escopo de visualização funciona interrelacionando duas notícias e seus conceitos dentro da ontologia, como exemplo Borsje et al (2007), informa sobre notícias do item A, que versam sobre Microsoft e Google, enquanto as notícias do item B, descrevem Google e o Youtube. Assim estas duas notícias estão ligadas pela presença do conceito Google. Seu alvo final é fornecer ao usuário a notícia com os conceitos previamente especificados pelo usuário.

#### 1.10.4 Plataforma KMAI®

A tecnologia KMAI® (*Knowledge Management with Artificial Intelligence - Gestão do Conhecimento com Inteligência Artificial*) é um modelo de sistema baseado em conhecimento que utiliza ontologias como forma de representar adequadamente a informação. Seu funcionamento ocorre em atividades de inteligência, recuperação de informações de cunho estratégico com o objetivo de embasar os processos voltados para tomadas de decisão. Um de seus clientes é o Governo Federal do Brasil, através da SAEI (Secretaria de Acompanhamento de Estudos Institucionais do Gabinete de Segurança Institucional da Presidência da República).

Sua concepção estrutura-se inicialmente pela reunião de três profissionais: o engenheiro do conhecimento, o analista de sistemas e o especialista do domínio. É nesta etapa que se define o escopo do sistema, sua condição de viabilidade, delimitação do domínio. Em seguida, inicia-se a etapa de engenharia de ontologias e de acordo com Ribeiro (2003, p.78) “[...]nessa fase é criada, com total participação dos especialistas-usuários, a ontologia da área com a construção de vocabulários controlados e dicionários de sinônimos e a rede de ontologias propriamente dita”.

Seu processo constituinte compreende quatro fases distintas (coleta, armazenamento, análise e difusão), precedidas pela etapa da delimitação do escopo.

A documentação e o detalhamento das funcionalidades do KMAI® são extensos e somente serão alvos de descrição o Sistema de Coleta de Informações (SC-Info), o Editor de Ontologias e a análise textual. SC-Info é o módulo encarregado de coletar documentos via *Web*. Para cada coleta cria-se um observador, através do preenchimento de um arquivo com extensão XML. Assim cada fonte de documentos terá seu próprio observador. Conforme assevera Ribeiro (2003, p.83): “Os observadores são agentes inteligentes que incorporarão atributos necessários para coletas produtivas em seus alvos digitais, considerando tempo de atualização, relevâncias dos assuntos, estrutura da fonte e armazenamento diferenciado”. O processo continua com a indexação dos documentos obtidos por meio do uso de ontologias. Estas são inseridas e organizadas dentro da tecnologia, através do Editor de Ontologias assim descrito por Bueno (2005, p.115):

O Editor de Ontologias é um sistema que foi concebido para auxiliar o trabalho da equipe de Especialistas na construção de ontologias. Consiste em uma estrutura que relaciona termos complexos, considerando seus conceitos no domínio de conhecimento específico do aplicativo, permitindo que este reconheça o contexto dos documentos em análise.

A fase seguinte é onde se dá a análise das informações ou análise textual. KMAI® dispõe de duas ferramentas: a PCE (Pesquisa Contextual Estruturada) e a RC2D (Representação do Conhecimento Contextualizada Dinamicamente) que se baseiam em um componente de IA chamado RBC (Raciocínio Baseado em Casos). HOESCHEL (2002, p.35-56) assim os explica:

A PCE é pesquisa no sentido de configurar um sistema que pesquisa informações, ou as busca, ou as recupera. É contextual porque o conhecimento por ela utilizado é representado contextualmente, e a pesquisa é realizada na base levando em consideração o contexto apresentado quando da solicitação da pesquisa. É estruturada em função de assim analisar tanto o contexto da solicitação quanto aquele dos documentos nos quais a pesquisa é realizada. Para que esta técnica funcione adequadamente, é fundamental que a representação do conhecimento seja contextualizada, e que isso ocorra dinamicamente [...] A RC2D consiste no processo de representação do conhecimento e respectivo ajuste, de forma dinâmica, a fim de que o funcionamento da pesquisa seja potencializado. Na construção desses sistemas fato importante é a definição das expressões relevantes em termos de recuperação da informação. Nesse sentido é construído um vocabulário controlado, com base em expressões consideradas relevantes pelos usuários do sistema. Para tanto é construída uma base de conhecimento, estruturada sobre um conjunto de expressões, utilizando diferentes referenciais, de forma a tratar com o máximo possível de personalização essas expressões. Esse processo de construção é chamado RC2D.

De acordo com Nicolini (2006), tanto o funcionamento da PCE quanto o da RC2D ocorrem mediante o uso de ontologias, construídas utilizando-se relações com sinônimos, conexos, 'parte de' e 'tipo de' dos vocábulos ou expressões usados. Todas estas inserções são feitas no sistema do Editor de Ontologias. Após municiar-se o KMAI® com os itens anteriores, o usuário poderá implementar pesquisas e análises segundo NICOLINI (2006, p.52):

A pesquisa e análise das informações são efetuadas no Módulo de Análise. Esse módulo permite busca em linguagem aberta e recupera os documentos mais semelhantes com o texto de entrada, armazenados na base de conhecimento, nas fontes selecionadas pelo cliente, definidas previamente na fase de Engenharia do Conhecimento.

Figura 8 – Tela de Entrada para Busca - Ferramenta Análise - KMAI

Uma outra aplicação da tecnologia KMAI®, ocorre no Ontoweb. Ele é um filtro de busca *Web* focado no cenário do Governo Eletrônico. Opera mediante ontologias, que se alicerçam em algoritmos capazes de realizar buscas mais precisas e devolver respostas com elevado grau de relevância aos usuários. Entre suas características destaca-se a possibilidade de entrada de até 15.000 caracteres para serem analisados pelo sistema. De acordo com Monteiro (2009) “o campo de pesquisa não é limitado a palavras-chave ou a simples expressões de busca [...]”. Seu cerne de operações repousa no uso de ontologias contextualizadas. Através delas ocorre a expansão dos conceitos (base do conhecimento) pesquisados, tendo como desdobramento o encontro de documentos mais pertinentes em relação à pesquisa inicial. A ferramenta também possui um pacote gráfico que avalia a evolução dos assuntos no decorrer do tempo. Os detalhamentos dos resultados das buscas podem ser verificados na figura 9 a seguir:



### Expressão em busca

[Análise Avançada](#)

Exemplo de consulta: "lavagem de dinheiro".  
 Na sua busca você terá os resultados desde 01/03/2004.  
 Para consultar períodos maiores, clique em [Análise Avançada](#).

**Dirceu diz que governo não vai se envolver com decisão sobre CPMI**  
 Agência Brasil - 22/05/2005 00:00:00  
 (ECT). "Os partidos têm que decidir sobre qual o caráter dela (da CPMI), qual o objetivo dela e se realiza CPMI ou não". Dirceu destacou que nunca se combateu no país a **lavagem de dinheiro**, como se combate hoje, o **narcotráfico** e o **crime organizado**. Segundo o ministro, há intenções eleitorais para a instalação da CPMI. "Não se trata de fazer CPMI sobre os crimes. O que se trata é de fazer, antes da hora, a disputa sobre o governo", afirmou. Dirceu também disse estar "profundamente indignado, perplexo" com uma matéria veiculada pela revista Veja ...

Em Arquivo...

Figura 9 – Tela de Busca do OntoWeb

Após digitar-se “lavagem de dinheiro”, um dos retornos liga esta expressão à palavra destacada pelos retângulos “narcotráfico” e à expressão “crime organizado”. Também dispõe de uma graduação tanto os documentos quanto para os textos mais significativos relacionados com o tema. Também apresenta um gráfico estatístico de graduação de respostas através dos anos, como mostrado na figura 10.

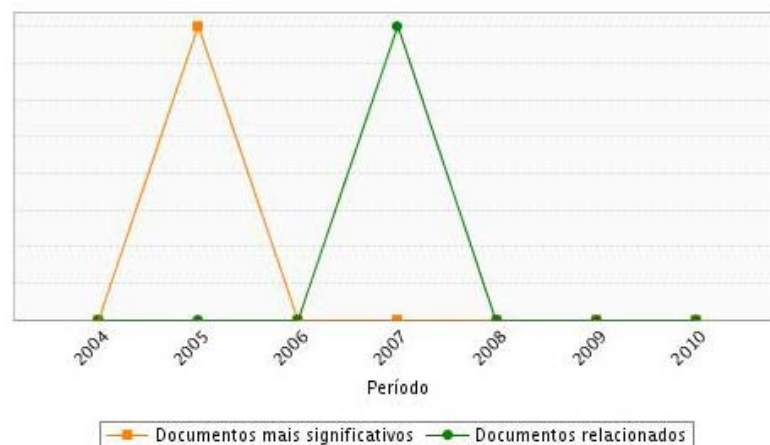


Figura 10 – Graduação de Respostas Emitida pelo OntoWeb

Foi realizada uma análise comparativa entre as propostas apresentadas nesta seção e OntoFeed (Ver Quadro Comparativo entre os Sistemas Estudados e OntoFeed, página 67 desta dissertação), no tocante à proposta do sistema, às tecnologias usadas, à adoção ou não de RSS, à forma de acesso, às possibilidades do usuário por configurar a interface, à plataforma utilizada, às possibilidades de localização da Notícia, à forma de busca (perguntas pré-definidas), ao objetivo do sistema e aos componentes do sistema (Módulos / Camadas / Fases). Verifica-se que todos os sistemas apresentados, assim como OntoFeed, como será visto no próximo capítulo, são aplicações para a Web, porém nenhum deles recupera informação contextualizada em agregadores de notícias.

### 1.11 Conclusões do Capítulo

Neste capítulo, foram discutidas as tecnologias que dão sustentação teórica à dissertação e foi oferecido ainda um breve estado da arte na área da pesquisa. Verifica-se que as ontologias se sobressaem como uma tecnologia versátil para imprimir semântica e representar conhecimento para diferentes tipos de aplicações e sistemas, como os motores de busca, os sistemas de suporte à decisão, entre outros.

A utilização de ontologias para apoiar a recuperação de informação em diversos domínios e com diferentes propósitos é um tópico recorrente na pesquisa e desenvolvimento na área de ciência da computação. Contudo, na literatura consultada foram encontradas propostas robustas, desenvolvidas por empresas ou grandes centros de pesquisa, como por exemplo, o *Master-Web* em desenvolvimento na Fiocruz. Mas, não foram localizados trabalhos que propusessem alguma extensão aos agregadores de notícia, visando os dotar de contexto ou semântica.

No próximo capítulo, é apresentado OntoFeed.

## 2 ONTOFFED: UM LEITOR LEITOR DE *FEEDS* COM EXTENSÃO ONTOLÓGICA

Este capítulo apresenta uma proposta de solução para o problema discutido ao longo da dissertação. A primeira seção oferece uma descrição do sistema OntoFeed e seu cenário de uso. As seções seguintes descrevem o protótipo desenvolvido; a ontologia para o domínio de chuvas e suas consequências, construída como exemplo; o funcionamento do protótipo; e, as conclusões do capítulo.

### 2.1 Sistema Ontofeed

O sistema proposto é uma aplicação para a *Web* capaz de buscar notícias sobre um domínio específico em portais de informação na *Web*. As tecnologias empregadas foram as linguagens de programação PHP, XML, JavaScript com AJAX e HTML com CSS e o banco de dados relacional MySQL. A aplicação proposta utiliza, ainda, a API<sup>1</sup> do Google Maps V1 para a localização georreferenciada da notícia, sempre que esta informação estiver disponível.

O sistema é acessado por uma URL e seu cenário de uso prevê a existência de uma ontologia para o domínio de interesse das consultas pré-determinadas pelo administrador apresentadas na interface com o usuário, por meio de perguntas que as orientam. OntoFeed é alimentado pelos sites de notícias cadastrados, que disponibilizam arquivos XML (RSS feed) com um conjunto de notícias individuais. Estas notícias individuais são filtradas uma a uma e, se contiverem as palavras-chave do pré-filtro (chuva, por exemplo), são armazenadas em uma tabela de RSS válidos no banco de dados para pesquisa por ontologia. Este pré-filtro é realizado via programação PHP. O sistema realiza a pesquisa por ontologia em três etapas:

- a primeira faz uma pré-seleção na tabela com os RSS válidos via consulta SQL procurando se alguma palavra chave da ontologia está no título ou no texto da notícia.
- a segunda etapa é um filtro mais refinado via programação PHP que busca combinações de palavras-chave no texto ou em frases, conforme cadastrado.

---

<sup>1</sup> API (*Application Programming Interface* ou Interface de Programação de Aplicativos) é um conjunto de rotinas e padrões estabelecidos por um software para a utilização das suas funcionalidades por aplicativos que não pretendem envolver-se em detalhes da implementação do software, mas apenas usar seus serviços. (Wikipedia)

- a terceira etapa é a localização do bairro da notícia via programação PHP. Nem sempre a localização é possível. Também podem ocorrer localizações erradas no caso de mais de um bairro na notícia ou mais de um bairro com o mesmo nome.

Após estas etapas, são exibidas as notícias filtradas e as localizações geográficas encontradas em um mapa do Google. Para compreender a proposta, é necessário entender o funcionamento dos feeds. Os Feeds chegam aos usuários como RSS, que são agregadores de conteúdo XML. Um RSS tem o formato interno como mostrado na figura 11.

```

<item>
  <title>Fernão Dias será fechada hoje em SP para detonação de rocha</title> Título do Item
  <link>http://redir.folha.com.br/redir/online/cotidiano/rss091/*http://www1.folha.uol.com.br/cotidiano/865568-fernao-dias-sera-fechada-hoje-em-sp-para-detonacao-de-rocha.shtml</link> Link do Item - Portal UOL Cotidiano
  <description>A rodovia Fernão Dias será totalmente interditada no sentido São Paulo na manhã desta terça-feira no km 13, altura de Bragança Paulista (85 km de São Paulo), para a detonação de uma rocha. <a href="http://www1.folha.uol.com.br/cotidiano/865526-erosoes-provocadas-pelas-chuvas-interditam-duas-estradas-de-sp.shtml">Erosões provocadas pelas chuvas interditam estradas de SP</a> Segundo a concessionária responsável pela estrada, o bloqueio terá início às 11h, e deve durar cerca de 30 minutos. Nesse período deve ocorrer a explosão de uma rocha e o trabalho de limpeza da pista, se necessário. <a href="http://redir.folha.com.br/redir/online/cotidiano/rss091/*http://www1.folha.uol.com.br/cotidiano/865568-fernao-dias-sera-fechada-hoje-em-sp-para-detonacao-de-rocha.shtml">Leia mais</a> (25/01/2011 - 09h41)</description> Descrição do Item
  <pubDate>25 Jan 2011 09:41:00 -0300</pubDate> Data
</item>

```

Figura 11 – Detalhes do item *description* do XML

No caso específico do domínio das chuvas do exemplo desenvolvido, o algoritmo desenvolvido para Ontofeed filtra as notícias para o evento “chuva”, descartando as que não forem relevantes para o tema. Após filtrá-las, procede à análise do item descrição do RSS (*description* na figura 20), extraíndo dele as palavras-chave do contexto da ontologia, apresentando ao usuário elementos novos (que não estão presentes na notícia). Esta ação corresponde a produzir inferências ligando a notícia ao arcabouço ontológico presente no algoritmo desenvolvido. O algoritmo utilizado no exemplo desenvolvido é o seguinte:

```

(arquivo buscaFontesRSSAjax.php)
Algoritmo principal da busca e carga dos RSS
  SE ConexaoComBanco=OK
  ENTÃO
    SE CarregaPalavrasChave=OK
    ENTÃO
      SE CarregaLinksRSS=OK
      ENTÃO
        PARA CADA linkRSS
        FAÇA
          SE BuscaRSS=OK
            LimpaRSSTemporarios
            InsereRSSTemporarios
            AtualizaParametroUltimaAtualizacao
          AtualizaCoordenadasRSSValidosDaFonteRSS
        ENTÃO

```



```

FIM-SE
SE InserirRSSValidos=OK
ENTÃO
    Inclusão OK
SENÃO
    Mensagem de erro na inclusão dos RSS validos
FIM-SE
FIM-FAÇA
SENÃO
    Mensagem de erro na carga dos links RSS
FIM-SE
SENÃO
    Mensagem de erro na carga das palavras-chave
FIM-SE
DesconectaBanco
SENÃO
    Mensagem de erro na conexão com o banco de dados
FIM-SE
-----
function BuscaRSS
SE ArquivoXML=OK
ENTÃO
    PARA CADA Noticia
    FAÇA
        PARA CADA PalavraChave
        FAÇA
            SE AchouPalavraChave no TituloNoticia ou AchouPalavraChave na
            DescricaoNoticia
            ENTÃO
                NoticiaValida=SIM
            SENÃO
                NoticiaValida=NAO
            FIM-SE
        FIM-FAÇA
    FIM-FAÇA
SENÃO
    Erro no arquivo XML com notícias
FIM-SE
=====
(arquivo FiltroCarregaRSSAjax.php)
function carregaRespostasOntologias
SE BuscaPalavrasFiltro=OK
ENTÃO
    PARA CADA PalavraDoFiltro
    FAÇA
        SE BuscaRSSQueTenhaQualquerPalavraDaLista //Case Insensitive
        SE TipoBusca=1 //Qualquer palavra no título ou descrição do RSS
        ENTÃO
            SE FiltraRSSEncontrados
            ENTÃO
                ArmazenaRssFiltrados
            FIM-SE
        SENÃO SE TipoBusca=2 //Combinação de palavras nas frases do título ou
        descrição do RSS
        ENTÃO
            SE FiltraFrasesRSSEncontrados
            ENTÃO

```

```

                                ArmazenaRssFiltrados
                                FIM-SE
                                FIM-SE
                                ENTAO
                                Exibe mensagem de erro na carga dos RSS
                                FIM-SE
                                FIM-FAÇA

                                OrdenaResultados
                                RemoveRssRepetidos
                                SENAO
                                Exibe mensagem de erro na carga das ontologias
                                FIM-SE

```

```

-----
function filtraTextoRSS
  PARA CADA TextoDoRss //Faz para título e descrição
  FAÇA
    SE AchouPalavraNoTexto
    ENTAO
      DestacaPalavraEncontrada
      ArmazenaRssParaExibicao
    FIM-SE
  FIM-FAÇA

```

```

-----
function filtraFrasesRSS
  QuebraTextoRssEmFrases //Faz para título e descrição
  PARA CADA Frase
  FAÇA
    SE AchouPalavraNaFrase
    ENTAO
      DestacaPalavraEncontrada
      ArmazenaRssParaExibicao
    FIM-SE
  FIM-FAÇA

```

```

=====
(FuncoesPHP.php)
function procuraPalavraNoTexto
  SE PalavraNoTexto=OK
  ENTÃO
    SE CaracterAnteriorAPalavra não é letra
    ENTAO
      EncontrouPalavra=SIM
    SENAO
      //Palavra é parte de outra
      EncontrouPalavra=NAO
    FIM-SE
  SENÃO
    EncontrouPalavra=NAO
  FIM-SE

```

```

-----
function atualizaCoordenadasRssValidos

  CarregaNomesDosBairros
  CarregaRSSSemCoordenadas

  PARA CADA Rss
  FAÇA
    PARA CADA Bairro
    FAÇA

```

SE AchouBairroNoTextoDoRss ou AchouBairroNoTituloDoRss  
 ENTAO  
     AtualizaCoordenadasRSS  
     FIM FAÇA-BAIRRO  
 FIM-SE  
 FIM-FAÇA-BAIRRO  
 FIM-FAÇA

O sistema proposto é o responsável pela aplicação do contexto ontológico às notícias vindas dos portais, com o objetivo de executar uma classificação automática das informações irrelevantes (atuando como um leitor comum de *Feeds*) e das relevantes (provendo material para as perguntas da ontologia). Esta solução computacional não tem características do tipo *desktop* por causa da própria natureza do trabalho na qual está envolvida. Assim, pensou-se numa aplicação *Web*, onde quaisquer usuários pudessem acessá-la, a qualquer momento, bastando para isto estarem conectados à Internet. As principais características do Ontofeed são:

1. Receptor de Feeds: ele recebe Feeds de diversos portais de notícias como G1, R7, UOL cotidiano e etc. Nesta fase, ele atua como um leitor comum de Feeds.
2. Opera como um filtro de notícias: As notícias que são relevantes no contexto da busca são armazenadas em um banco de dados e ficam disponíveis para que as consultas dos usuários via *Web* (através das perguntas da ontologia) aconteçam. As demais notícias são expostas do lado esquerdo da página *Web* do sistema. (vide figura 16)
3. Analisador do Item Descrição: Título e Descrição dos RSS sofrem análise do algoritmo do Ontofeed, com a finalidade de se localizar e extrair palavras-chave do contexto da ontologia e apresentá-las aos usuários, como respostas às perguntas previamente determinadas da ontologia. Esta análise e extração obedecem aos conceitos e definições usados na elaboração da ontologia. No caso do exemplo desenvolvido, se usou o previsto na doutrina da Defesa Civil (Glossário, Manuais de Desastre e o CODAR - Codificação de Desastres, Ameaças e Riscos).
4. Possibilidade de Expansão: Novas perguntas da ontologia podem ser adicionadas àquelas que já existem, ampliando a abrangência das pesquisas no banco de dados.

A figura 12 apresenta a concepção do sistema. Na fase 1, que abrange o cadastro dos portais de notícias, são cadastrados no ambiente do OntoFeed diversos sites emissores de Feeds. Na fase 2, se dá a recepção dos RSS enviados pelos portais. Nesta etapa o OntoFeed atua como um coletor de notícias. Na fase 3, o sistema começa a filtrar os Feeds coletados de

acordo com um algoritmo que engloba o domínio da ontologia e o resultado alcançado é duplo pois os Feeds não relevantes são expostos para leitura (OntoFeed atuando como leitor comum de RSS) e posteriormente descartados, enquanto que os Feeds que contiverem palavras-chave do contexto da ontologia são armazenados em um banco de dados para posteriores consultas.

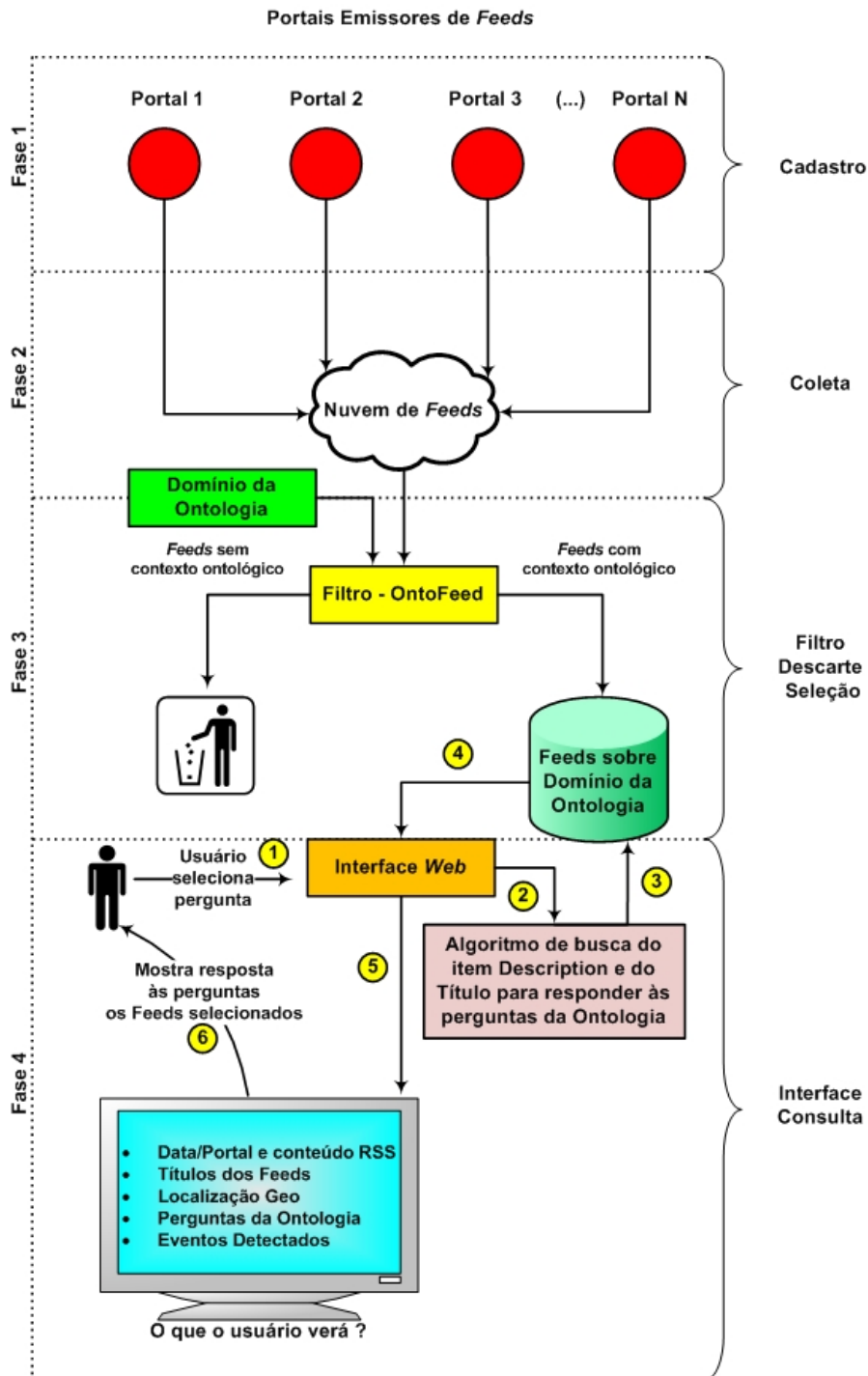


Figura 12 – Diagrama das Fases do OntoFeed

As consultas ou perguntas da ontologia serão construídas pelo administrador do sistema e ficarão disponíveis na fase 4, onde ocorre a interação com o usuário. Caberá ao usuário *Web* selecionar uma entre as perguntas apresentadas e consultar o banco de dados, com o objetivo que o sistema execute inferências entre a pergunta e os Feeds armazenados como relevantes para o tema. A figura 12 mostra a concepção do sistema.

As tabelas que compõem o banco de dados com os seus campos, chaves primárias e estrangeiras estão representadas no modelo de entidades e relacionamentos (figura 13).

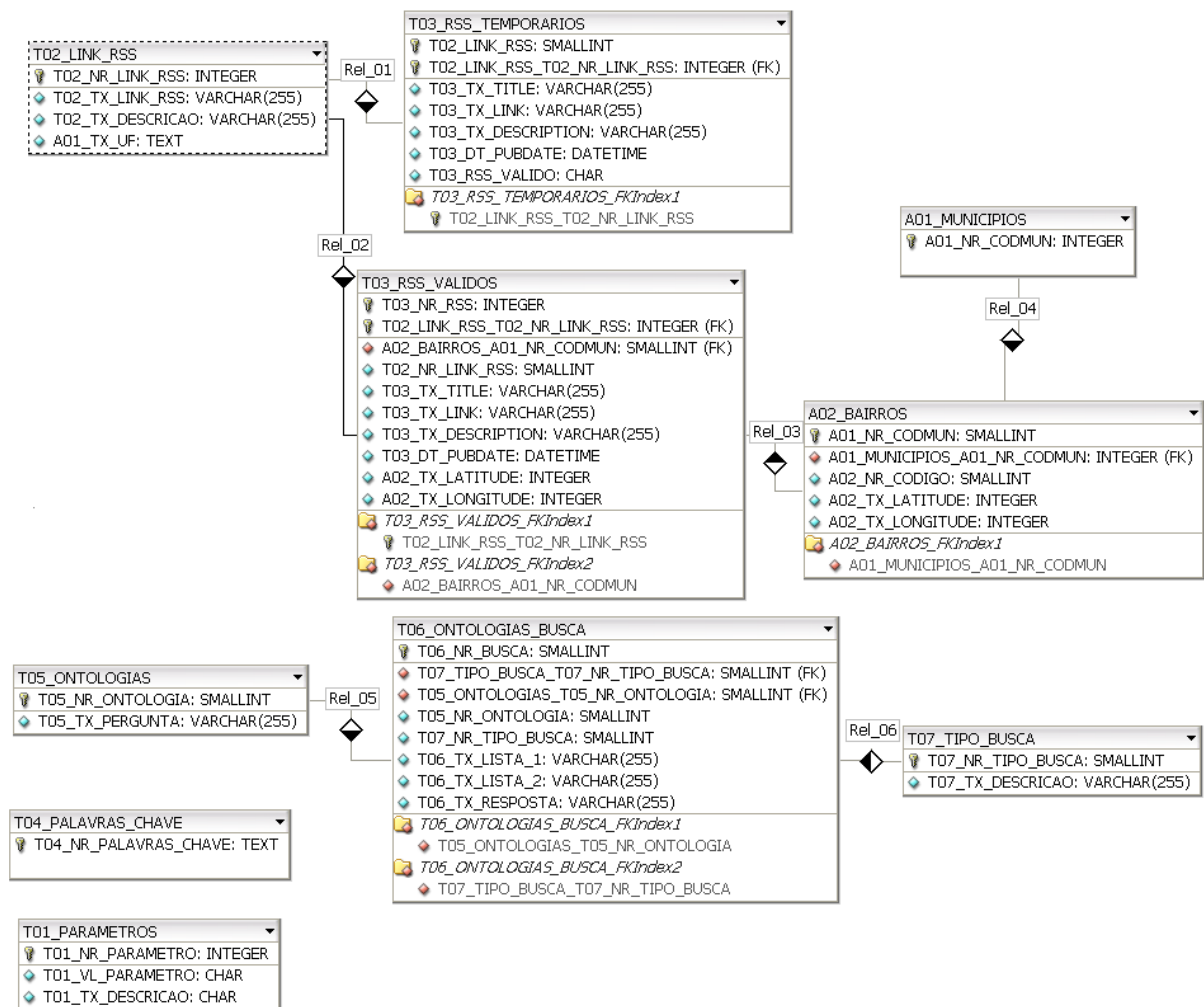


Figura 13 – Modelo Entidades e Relacionamento do OntoFeed

## 2.2 Seleção das Fontes Emissoras de RSS e Pré-Filtro

Nesta etapa são selecionados alguns portais emissores de notícias, no formato RSS, tais como G1, UOL, OGLOBO e R7 (figura 14). Esta seleção pode ser ampliada ou mesmo modificada pelo administrador do sistema a qualquer momento. Estes Feeds são o alvo do programa que analisa o seu conteúdo (description) e o seu título.

**Fontes RSS - Inclusão/Alteração de campos**

Número:

Valor:

Descrição:

Número	Valor	Descrição
1	http://g1.globo.com/dynamo/rio-de-janeiro/rss2.xml	G1 - Rio de Janeiro [alterar] [excluir]
2	http://feeds.folha.uol.com.br/folha/emcimahora/rss091.xml	Folha - Em cima da hora [alterar] [excluir]
3	http://g1.globo.com/dynamo/sao-paulo/rss2.xml	G1 - São Paulo [alterar] [excluir]
4	http://www.r7.com/data/rss/previsaoTempo.xml	R7 - Previsão do Tempo [alterar] [excluir]
5	http://www.r7.com/data/rss/saoPaulo.xml	R7 - São Paulo [alterar] [excluir]
6	http://www.r7.com/data/rss/RiodeJaneiro.xml	R7 - Rio de Janeiro [alterar] [excluir]
7	http://feeds.folha.uol.com.br/folha/cotidiano/rss091.xml	Folha - Cotidiano [alterar] [excluir]
8	http://oglobo.globo.com/rss/plantaopais.xml	O Globo - País [alterar] [excluir]

Figura 14 – Tela de Inclusão de Fontes Emissoras de RSS

Após o cadastramento das fontes emissoras, OntoFeed torna-se capaz de receber os *Feeds* emitidos. Nesta fase de recepção de *Feeds*, o sistema executa uma pré-filtragem das notícias, realizando uma busca por palavras chave (que sejam relevantes ao contexto da ontologia criada) tanto no description como em seu título.

**Parâmetros - Inclusão/Alteração de campos**

Número:

Valor:

Descrição:

Número	Valor	Descrição
1	ontofeed	nome do banco de dados [alterar] [excluir]
2	root	nome do usuário padrão [alterar] [excluir]
3		senha do usuário padrão [alterar] [excluir]
4	2011-06-14 13:00:34	última atualização [alterar] [excluir]
5	Chuv,chuv,Chov,chov,Tempora,tempora	pré-filtro de RSS [alterar] [excluir]

Figura 15 – Tela de Inclusão de Palavras ou Expressões do Pré-Filtro

No caso específico de exemplo desenvolvido (Chuvas e suas consequências), para este pré-filtro foram escolhidas as palavras “chuva” e “temporal” com suas variantes maiúsculas e minúsculas (Chuv\*,chuv\*,Chov\*,chov\*,Tempora\*,tempora\*), como se vê na figura 15. O número de palavras e suas variantes pode ser ampliado ou modificado a qualquer momento pelo administrador do sistema.

**ONTOFeed**

**Últimas notícias**

- 14/06 12:06h | Folha - Em cima da hora Santos desembarca em Montevideú e final será na quarta
- 14/06 12:06h | Folha - Em cima da hora Lily Allen adota sobrenome do marido após o casamento
- 14/06 12:06h | Folha - Em cima da hora Médico que declara laço com indústria também é antiético
- 14/06 12:06h | Folha - Em cima da hora Presos morrem em confronto em prisão na Venezuela
- 14/06 12:06h | G1 - São Paulo Aprenda a fazer bolo de pinhão para festas juninas
- 14/06 12:06h | R7 - São Paulo PM persegue suspeitos de roubo a carro dos Correios na zona norte de SP
- 14/06 12:06h | Folha - Em cima da hora Acordo sobre patentes com Apple é vitória rara para Nokia
- 14/06 12:06h | G1 - Rio de Janeiro Polícia descobre desmanche de carros em sítio na Zona Oeste do Rio
- 14/06 12:06h | Folha - Em cima da hora Líderes neonazistas são condenados a prisão perpétua na Rússia
- 14/06 12:06h | G1 - Rio de Janeiro Aneel autoriza alta na tarifa de energia de Nova Friburgo (RJ)
- 14/06 12:06h | O Globo - País Sobre documentos secretos, Sarney diz que não se pode fazer "WikiLeaks" da história do Brasil
- 14/06 12:06h | G1 - São Paulo Fundação faz campanha para incentivar doação de sangue em SP
- 14/06 12:06h | G1 - São Paulo Com voos suspensos, santistas planejam ir de carro para o Uruguai
- 14/06 12:06h | R7 - Rio de Janeiro Polícia encontra desmanche de veículos em Sepetiba

**Perguntas da Ontologia**

- Quais recursos hídricos estão presentes nesta notícia ?
- Quais índices pluviométricos esta notícia detectou ?
- As chuvas provocaram falta de energia elétrica ?
- Pessoas ou bens ficaram soterrados ?

BuscaRespostas

Selecione uma pergunta da Ontologia

Figura 16 – Ambiente Web do OntoFeed – Feeds à esquerda e extensão ontológica à direita

O pré-filtro é responsável pela criação de dois tipos de tópicos para as notícias recebidas: as relevantes para o tema da ontologia e as irrelevantes. As irrelevantes serão as que não contiverem em suas descrições e/ou títulos ao menos uma das palavras chave, descritas anteriormente. O usuário lerá estas notícias, como normalmente ocorre em um leitor

comum de Feeds. Elas ficarão dispostas com título, data, hora e link e serão mostradas do lado esquerdo da tela do OntoFeed. Já as que contiverem as palavras-chave serão cadastradas no Banco de Dados do Sistema para posteriores consultas e respostas às perguntas da ontologia. Este filtro é executado automaticamente sem a interferência do usuário.

Como apresentado na figura 16, os *Feeds* considerados não relevantes, recebidos dos portais cadastrados, são apresentados ao usuário abaixo do tópico “Últimas notícias”.

### 2.3 Construção da Extensão Ontológica

Nesta etapa, são selecionadas perguntas para serem respondidas pela ontologia. No exemplo desenvolvido, tomou-se por base o que está descrito na doutrina da Defesa Civil sobre chuvas e suas consequências. Para tal, o sistema oferece duas telas onde o administrador pode cadastrar as perguntas e as palavras chave que deverão ser procuradas nos textos dos RSS, armazenados no banco de dados do sistema (figura 17).

**Ontologias - Inclusão/Alteração de campos**

Número:

Pergunta:

Número	Pergunta	
1	Quais recursos hídricos estão presentes nesta notícia ?	[alterar] [excluir]
2	Quais índices pluviométricos esta notícia detectou ?	[alterar] [excluir]
3	As chuvas provocaram falta de energia elétrica ?	[alterar] [excluir]
4	Pessoas ou bens ficaram soterrados ?	[alterar] [excluir]
5	Pessoas ou locais ficaram ilhados ?	[alterar] [excluir]
6	O funcionamento de Aeroportos foi prejudicado ?	[alterar] [excluir]
7	Pessoas foram atingidas por Raio ?	[alterar] [excluir]
8	Existiram queda de barreiras em rodovias ?	[alterar] [excluir]
9	Ocorreram ações de resgate na incidência de soterramento ?	[alterar] [excluir]
10	Os eventos adversos geraram desabrigados e desalojados ?	[alterar] [excluir]

Figura 17 – Tela de Inclusão das Perguntas da Ontologia

Cada pergunta possui uma tela de cadastro de palavras-chave relativas à pergunta, com critérios e restrições. Cabe ao administrador defini-los, podendo ampliá-los ou adequá-los conforme o caso. Como exemplo, na pergunta 1 da ontologia: “Quais os recursos hídricos estão presentes nesta notícia?”, foram selecionadas as palavras: rio, ribeiro, ribeirão, córrego e valão, todos estes recursos hídricos.



Geralmente o transbordamento de recursos hídricos e suas correntezas são os responsáveis pelos danos e prejuízos ocorridos durante e após o evento chuvas. Uma das restrições desta pergunta é o uso da palavra “rio” sempre com letras minúsculas, para não confundir com “Rio” (Rio de Janeiro).

**Palavras-Chave das Ontologias - Inclusão/Alteração de campos**

Nr.Busca:

Ontologia:

Lista1:

Lista2:

Resposta:

Nr.Busca	Ontologia	Lista1	Lista2	Resposta		
1	1	rio, ribeiro, ribeirão, córrego, valão		Recurso Natural Hídrico	[alterar]	[excluir]
2	1	subi	nível, nivel	Recurso Natural Hídrico	[alterar]	[excluir]
3	2	milímetro, milimetro, mm		Índice de chuva	[alterar]	[excluir]
4	3	queda de energia, falta de energia, falta de luz, sem luz, sem energia, abastecimento de energia		Prejuízo Sócio-Econômico >> Falta de Energia Elétrica	[alterar]	[excluir]

Figura 18 – Tela de Inclusão de Palavras-Chave para as Perguntas da Ontologia

Após este cadastro de perguntas (figura 18), o sistema está disponível para que o usuário possa acessá-lo via *Web* e escolher qual das perguntas, pré-definidas pelo administrador, deseja que o aplicativo responda mediante consulta ao banco de dados do programa.

Para melhor explicar o funcionamento de Ontofeed, foi desenvolvido um exemplo no domínio de chuvas e suas consequências. A seção 2.4 apresenta a ontologia construída e a seção 2.5 oferece um exemplo de funcionamento do sistema.

## 2.4 A Construção da Ontologia Chuvas e suas Consequências

A ontologia de domínio “Chuvas e suas consequências” visa fornecer a representação conceitual dos principais desdobramentos ocasionados por chuvas nas cidades brasileiras. Com a finalidade de se organizar este conhecimento de maneira formal, utilizou-se o glossário e os manuais de desastre da Secretaria Nacional de Defesa Civil, como base para se estruturar os termos envolvidos no contexto ontológico.

A estruturação da ontologia foi realizada com a metodologia proposta por Noy e McGuinness (2001) e com o *software* Protégé (versões 3.4.4 e 4.1 beta, que apresenta uma visualização melhor das classes).

**Passo 1:** Determinar o domínio e o escopo da ontologia: O campo de pesquisa tem como alvo, os fatos que ocorrem durante e logo após as chuvas, e conseguem provocar danos e prejuízos, tanto do ponto de vista humano como os que se dão nas áreas material e econômica das regiões atingidas. Segundo o Glossário de Defesa Civil, Estudos de Riscos e Medicina de Desastres (CASTRO, 2004, p.77) um evento adverso é definido como: “Ocorrência desfavorável, prejudicial, imprópria. Acontecimento que traz prejuízo, infortúnio. Fenômeno causador de um desastre.” Nesta abordagem, considera-se que a chuva seja o fator que contribuirá de forma decisiva para a ocorrência de eventos adversos. O objetivo para a criação desta ontologia é prover de um conhecimento formal, sobre eventos adversos e desastres, uma aplicação computacional capaz de analisar o item descrição dos *Feeds* enviados pelos portais de notícias e extrair deles inferências, ampliando as informações apresentadas ao usuário do serviço. Assim, o cliente *Web* não somente lerá a notícia relevante dentro do tema chuva e suas consequências, mas também obterá o acréscimo de ter itens específicos da descrição classificados segundo as normas da Defesa Civil. Desta forma a ontologia deverá fornecer respostas, por meio da classificação dos itens do contexto ontológico às seguintes perguntas:

- 1- Quais recursos hídricos estão presentes nesta notícia?
- 2- Quais índices pluviométricos esta notícia detectou?
- 3- As chuvas provocaram falta de energia elétrica?
- 4- Pessoas ou bens ficaram soterrados?
- 5- Pessoas ou locais ficaram ilhados?
- 6- O funcionamento de aeroportos ficou prejudicado?
- 7- Pessoas foram atingidas por raio?
- 8- Existiram quedas de barreira em rodovias?
- 9- Ocorreram ações de resgate na incidência de soterramento?
- 10- Os eventos adversos geraram desabrigados ou desalojados?

Com isto espera-se que os usuários do serviço (clientes que assinam os *Feeds*) recebam esta funcionalidade de análise do item descrição, permitindo-lhes inclusive operar pesquisas por localização e ocorrência via *Web* de *Feeds* passados.

**Passo 2:** Verificar a possibilidade de reuso de outras ontologias: Efetuaram-se diversas pesquisas e buscas em documentos e na Internet, com a finalidade de encontrar uma ontologia deste domínio. A ontologia *FreezingRain*

(<http://sweet.jpl.nasa.gov/ontology/phenomena.owl#FreezingRain>), elaborada pela NASA, engloba tipos de desastres como terremotos e furacões, que não se aplicam ao Brasil, escopo deste trabalho. Outra ontologia do domínio chuvas também desenvolvida pela NASA, *Rain* (<http://sweet.jpl.nasa.gov/2.0/atmoPrecip.owl#Rain>) dá atenção, entre outras coisas, aos tipos de nuvem geradoras das precipitações pluviométricas, voltando-se para estudos meteorológicos. Por esta razão, não foi possível efetuar o reuso.

**Passo 3:** Enumerar os termos relevantes da ontologia: Na pesquisa pelos termos relevantes considerou-se como fonte balizadora e especialista do conhecimento o Glossário de Defesa Civil, Estudos de Riscos e Medicinas de Desastre como também os Manuais de Desastre da Defesa Civil e o CODAR (Codificação de Desastres, Ameaças e Riscos). Alguns termos selecionadas são apresentados a seguir na tabela 1:

Inundação	Enchente	Enxurrada	Alagamento	Aguaceiro	Desabamento
Deslizamento	Soterrados	Afogados	Desabrigados	Deslocados	Desalojados
Ilhados	Granizo	Vendaval	Tempestade	Mau Tempo	Calamidade
Emergência	Barreiras	Encostas	Vítimas	Defesa Civil	Barrancos
Congestionamento		Interdição de Edificação		Queda de Árvore	
Falta de Luz	Falta d'Água	Queda de Poste		Vôos Cancelados/Atrasados	
Destelhamento	Soterramento	Desmoronamento Edificações		Rio	Córrego
Valão	Transbordamento		Desastre	Busca	Salvamento
Mortos	Feridos	Bombeiros	Bote	Pontes	Túneis
Rede Elétrica	Edificações Alagadas		Correnteza	Danos na Telefonia Fixa	
Raios	Cabeça d'Água		Leptospirose	Desaparecidos	
Acidentes	Pista Escorregadia		Danos nos Serviços de Transporte		
Resgate	Índice Pluviométrico		Bens Privados Destruídos/Danificados		
Cheia	Campanhas de Doação		Bens Públicos Destruídos/Danificados		
Helicóptero	Tromba d'Água		Escorregamento		Avalanche
Afetado	Área Vulnerável		Queda de Barreira		Isolados

Tabela 1 – Termos Candidatos à Classe da Ontologia

**Passo 4:** Definir classes e suas hierarquias: (Do passo 4 até ao passo 7, as ações de construção da ontologia se passarão no software Protégé.) A maneira usada para realizar esta fase foi a *top-down* (de cima para baixo). Os termos que aparecem como objetos tendo sua existência independentemente dos demais poderão figurar como classes ou mesmo superclasses da ontologia. Procedeu-se então à investigação no material doutrinário da Defesa Civil, comparando-o com

as expressões coletadas anteriormente. Nesta etapa foram descobertas redundâncias de termos (expressões diferentes que significam a mesma idéia). Esta análise possibilitou que alguns itens fossem descartados e outros aglutinados sob a mesma semântica. Dois exemplos de termos diferentes, mas de mesmo significado, são: 1) enchente/inundação e 2) deslizamento/escorregamento, que ocorreram como elementos separados no passo 3, mas que agora tornam-se apenas uma coisa (enchente igual a inundação e deslizamento igual a escorregamento).

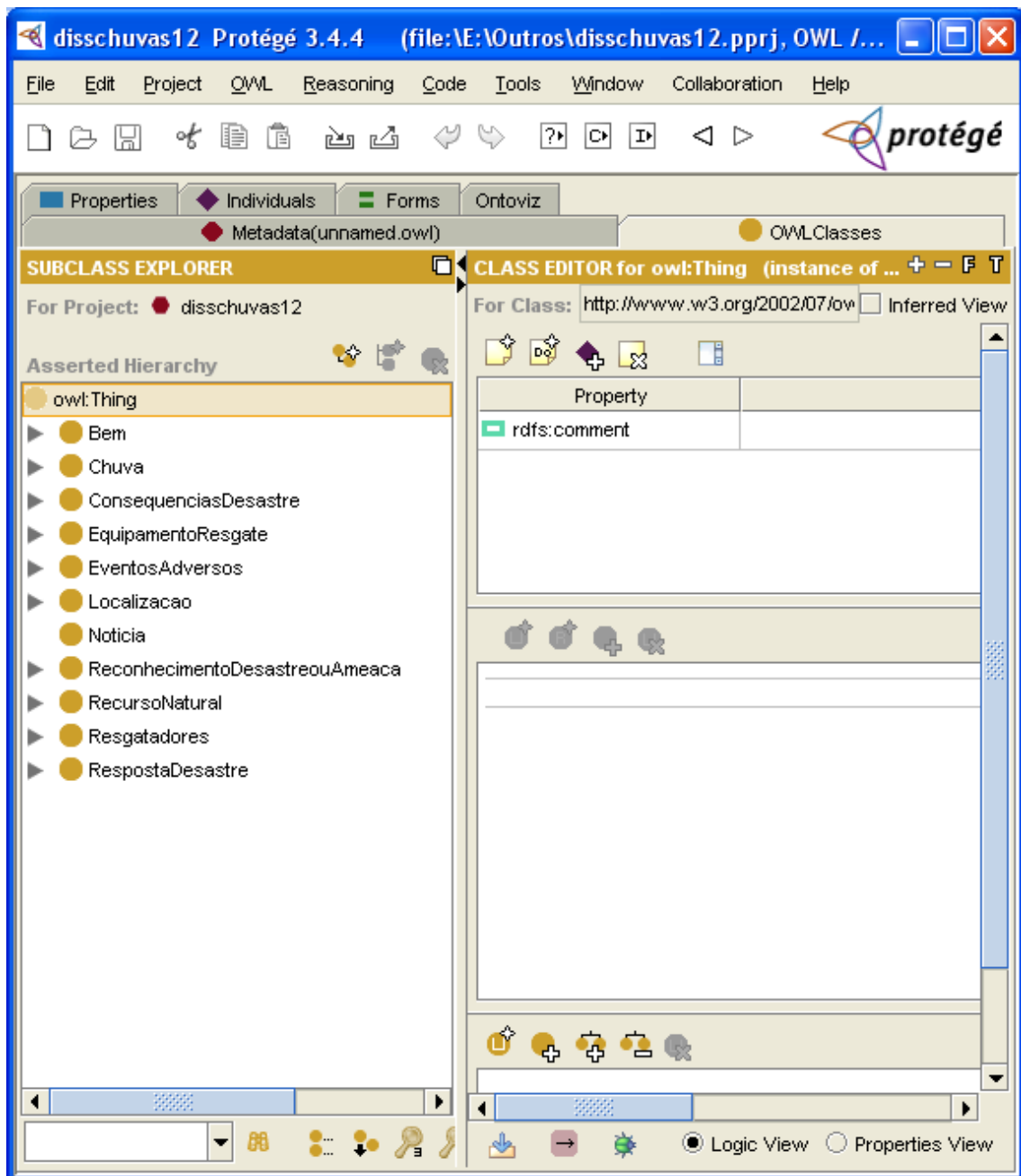


Figura 19 – Classes e subclasses no Protégé [2]

As classes foram obtidas recortando-se as várias facetas que envolvem a análise das chuvas e suas consequências, sob a ótica da doutrina da Defesa Civil, podendo ser visualizadas nas figuras 19, 20 e 21, no software Protégé. Assim, foram criadas as classes:

- Bem, Chuva, ConsequênciasDesastre, EquipamentosResgate, EventosAdversos, Localização, Notícia, ReconhecimentoDesastreouAmeaça, Recurso Natural,
- Somente a classe Notícia não possui subclasses. Toda a taxonomia da ontologia é mostrada por fragmentos das telas do software Protégé em duas partes:

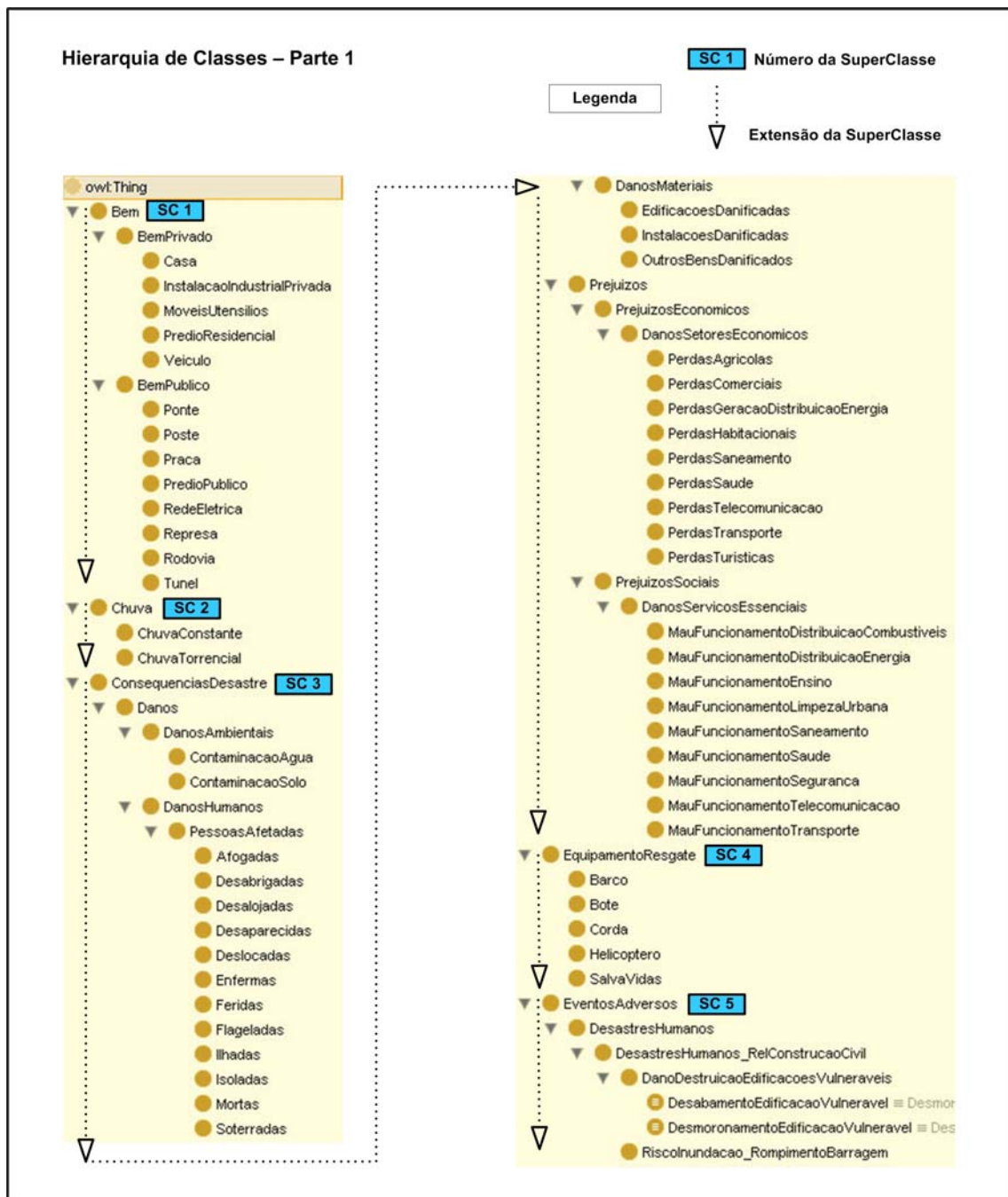


Figura 20 – Classes e subclasses no Protégé [1]

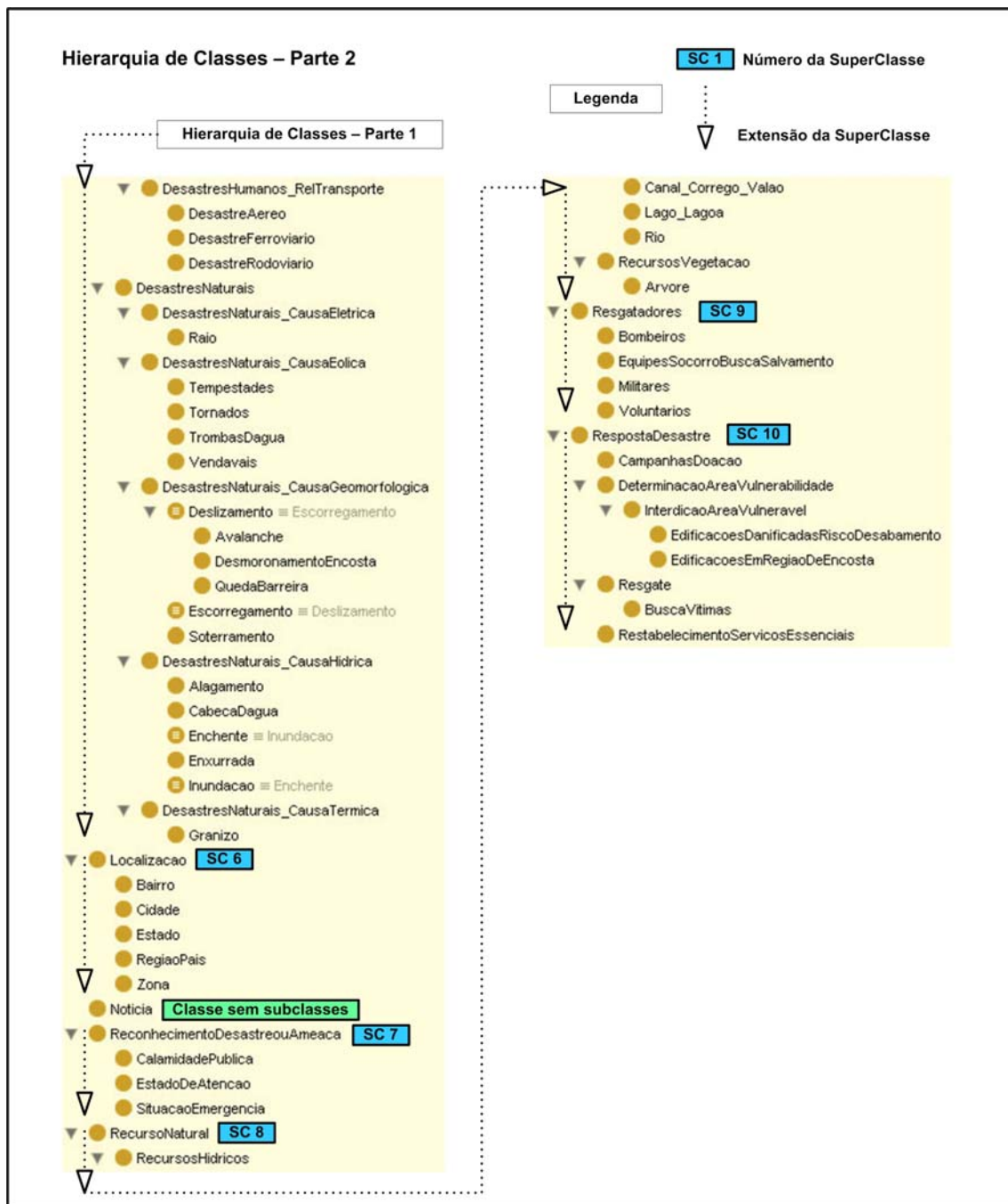


Figura 21 – Classes e subclasses no Protégé [2]

**Passo 5** : Definir as propriedades das classes : Foram definidas onze propriedades de objeto e quatro propriedades de tipos de dados. Nesta modelagem a classe Notícia é apresentada como a principal, pois é dela que se extrairá o conhecimento e onde se operarão as inferências no contexto ontológico montado. Também será ela a responsável por interligar toda a ontologia através de suas onze propriedades de objeto. A própria classe Notícia, por sua vez contará com quatro propriedades de tipos de dados, com a finalidade de tornar única cada uma das

notícias recebidas dos portais informativos. As propriedades de objeto são: temBem, temChuva, temDano, temPrejuízo, temDesastre, temEquipamento, temLocalização, temReconhecimento, temRecursoNatural, temResgatador, temResposta. Já as quatro propriedades de tipos de dados da classe Notícia são: Título, Data, *Link* e Portal. Na propriedade Título está o título da notícia, na Data, a data em que ocorreu o evento, na Link, o link da notícia e finalmente a propriedade Portal identificará o local emissor do RSS. A propriedade de anotação (figura 22) será utilizada para colocar o conteúdo da notícia.

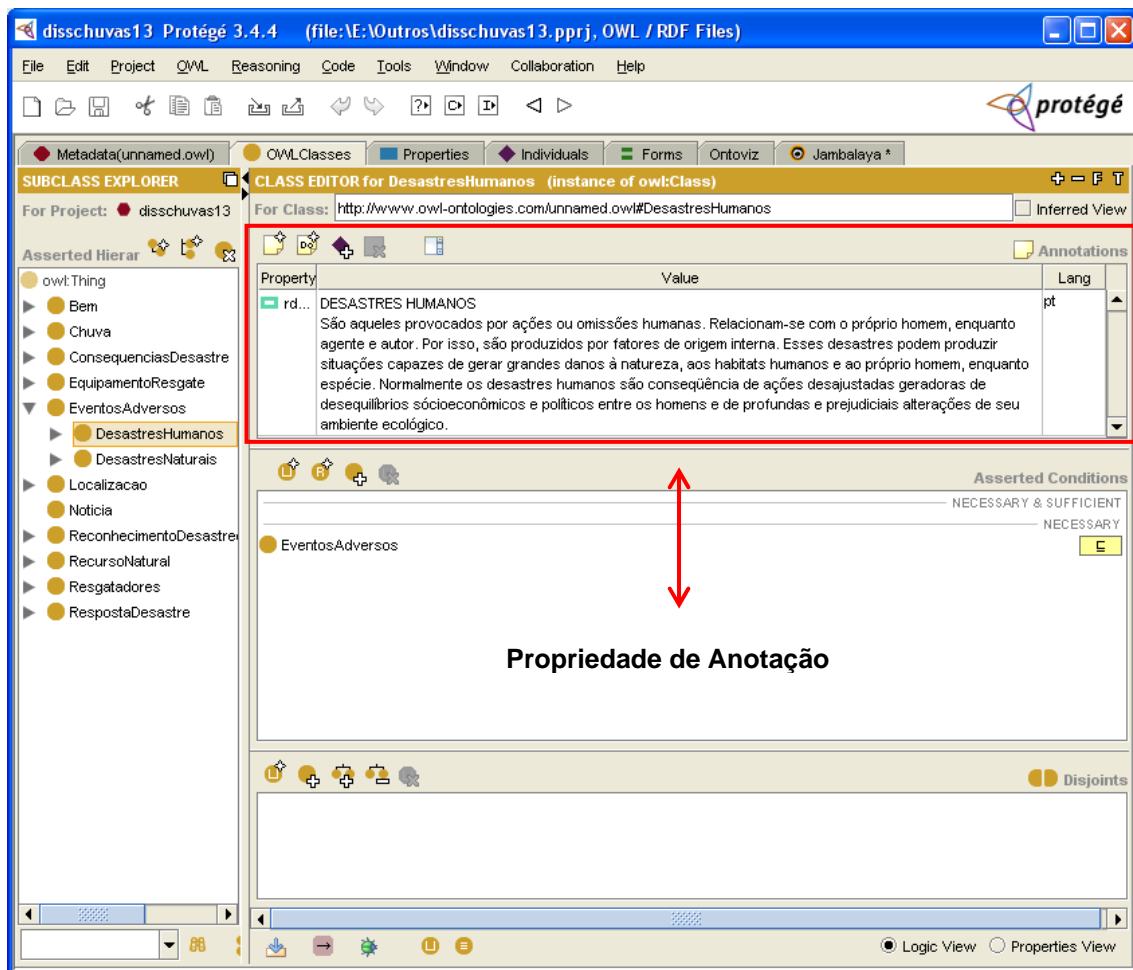


Figura 22 – Uso da Propriedade Anotação no Protégé

**Passo 6** : Definir as características de cada propriedade : As onze propriedades de objeto foram assinaladas como transitivas (figura 23), em virtude de em uma mesma notícia existir a possibilidade de ocorrerem vários eventos pertencentes a classes distintas como por exemplo: deslizamento – (Desastre Natural Geomorfológico), alagamento – (Desastre Natural Hídrico) e pessoas desabrigadas (Dano Humano). As quatro propriedades de tipos de dados embutidas na classe Notícia apresentam as seguintes características: Título (título da notícia), Link e Portal recebem *string*; já Data recebe *date*. Também aqui, foram usadas as propriedades de



anotação disponíveis no Protégé. Seu alvo foi mostrar as definições formais das classes e subclasses componentes da ontologia. As definições, aqui apresentadas são em sua grande maioria extrações do Glossário da Defesa Civil.

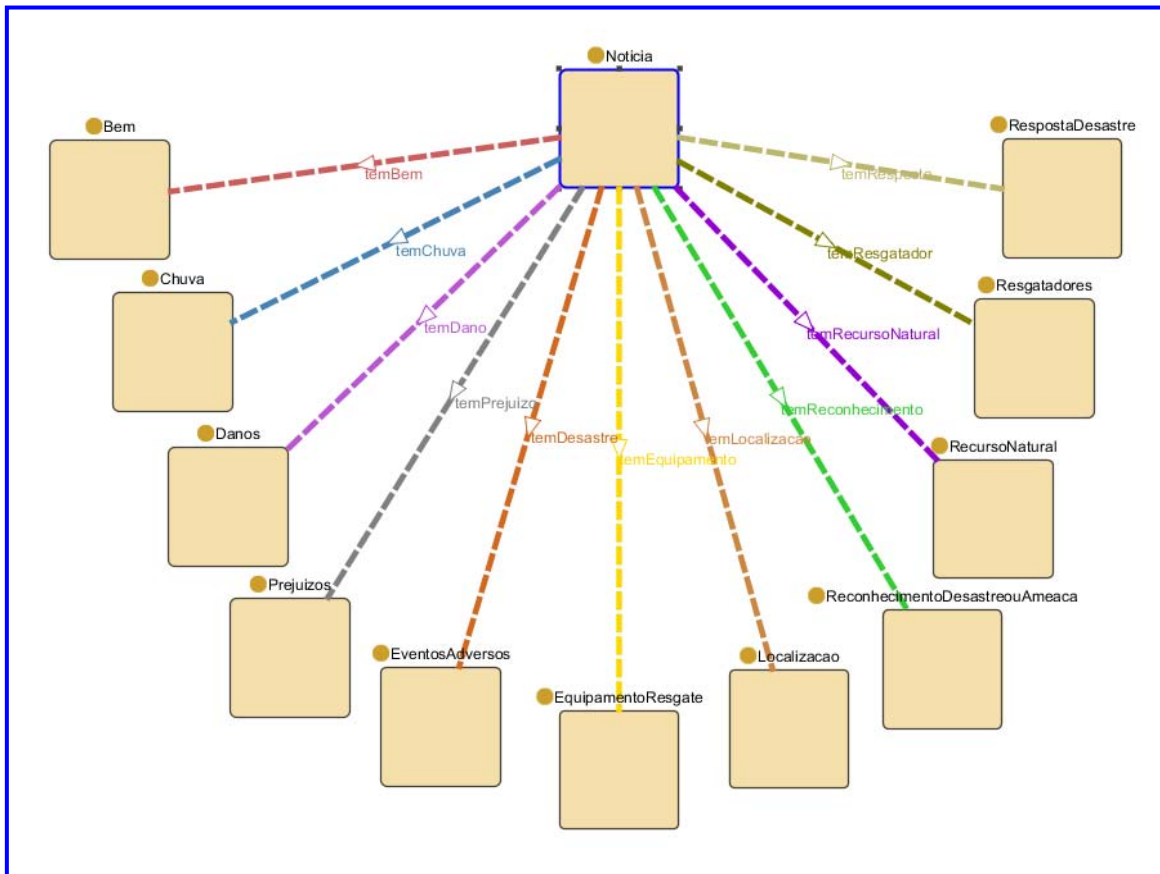


Figura 23 – Classe Notícia e suas Propriedades

**Passo 7** : Criação de instâncias : Para povoar a ontologia foram reunidas diversas notícias oriundas dos *Feeds* recebidos dos portais informativos. Como exemplo, usou-se uma notícia recebida do portal UOL-Cotidiano, com o intuito de detalhar como ocorre a criação de indivíduos. Para que um indivíduo pertença ao conjunto de membros da ontologia faz-se necessário cumprir três condições:

- Ser instância de alguma classe ou subclasse;
- Ser distinto de todos os outros indivíduos e
- Ter suas propriedades devidamente preenchidas.

Quando ocorrer o caso em que uma mesma notícia verse sobre locais diferentes, ela será desdobrada no Protégé (em duas, três ou tantas conforme os locais nela contidos) recebendo



uma numeração única no *software*, para cumprir a condição [ii]. O povoamento da ontologia é um processo simples, conduzido com o suporte de Protégé. Por exemplo, para povoar a classe Notícia, seleciona-se a aba “*individuals*”, escolhe-se adicionar uma nova instância – opção “*create instance*” e surge a tela de inclusão de indivíduos. O exemplo abaixo ilustra o processo (figura 24).

Notícia do UOL – Cotidiano

Link: <http://www1.folha.uol.com.br/cotidiano/854751-ponte-e-levada-pela-correnteza-e-rodovia-go-070-e-interditada.shtml>

Data: 04/01/2011 Título: Ponte é levada pela correnteza e Rodovia GO-070 é interditada

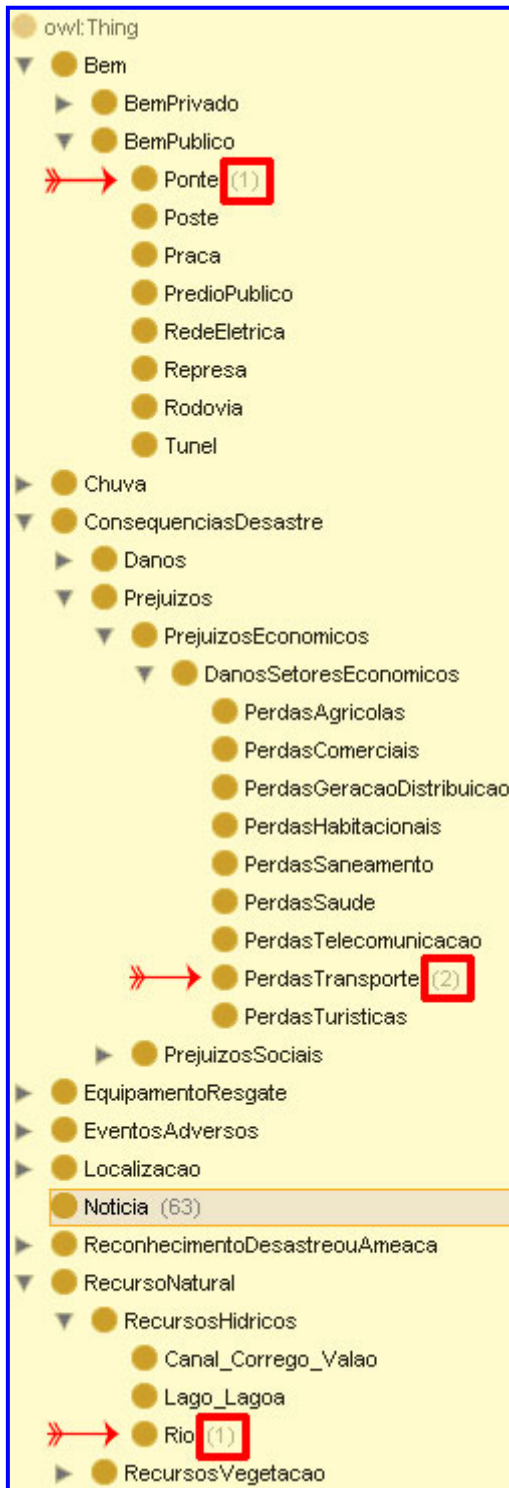
The screenshot shows the Protégé Individual Editor for a Noticia\_1 instance. The main content area displays the 'Corpus ou Conteúdo da Notícia' with the following text: "As chuvas fortes que atingem a região Centro-Oeste há mais de duas semanas continuam provocando estragos. Uma ponte da GO-070, a menos de 90 km de Goiânia, no trecho que liga Itaberaí à cidade histórica de Goiás (antiga Goiás Velho), não resistiu à forte correnteza do rio das Pedras e desabou na madrugada desta terça-feira."

The interface includes several property editors, each with a table for values and a list of instances:

- Titulo:** Value: Ponte é levada pela correnteza...pt
- Data:** Value: 2011-01-04, Type: date
- Link:** Value: http://www1.folha.uol.com.br/...
- Portal:** Value: UOL
- temDano:** Instance: Ponte\_Destruida
- temPrejuizo:** Instances: Queda\_de\_Ponte, Rodovia\_Interditada
- temReconhecimento:** (Empty)
- temDesastre:** (Empty)
- temRecursoNatural:** Instance: Rio\_das\_Pedras
- temEquipamento:** (Empty)
- temResgatador:** (Empty)
- temBem:** Instance: Ponte\_da\_Rodovia\_GO-070
- temLocalizacao:** Instance: Goiás
- temResposta:** (Empty)

Figura 24 – Instanciamento de uma Notícia no Protégé

Após efetuar o lançamento do primeiro indivíduo (Notícia\_1), os itens presentes nela já farão parte do conteúdo das classes, aparecendo na taxonomia com as suas devidas quantidades como mostra a figura 25.



A queda da ponte entra como um Bem Público que foi atingido pelas consequências das chuvas.

Bem >> Bem Publico >> Ponte

Instância: Queda de Ponte da Rodovia GO-070

O dano provocado na Ponte também é classificado como Edificação Danificada dentro da classe Danos.

[...] >> DanosMateriais >> EdificacoesDanificadas

Instância: Ponte Destruída

A queda da ponte e a interdição da rodovia são dois eventos ligados ao transporte que trarão prejuízos ou perdas para este setor.

[...] >> Prejuizos >> [...] >> PerdasTransporte

Instância1: Queda de Ponte

Instância2: Rodovia Interditada

O evento é localizado no estado de Goiás.

Localizacao >> Estado

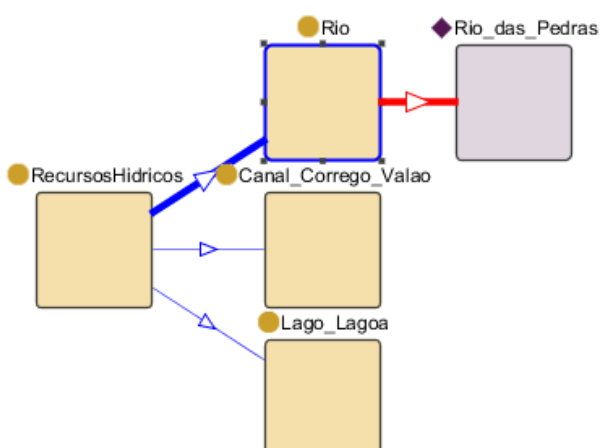
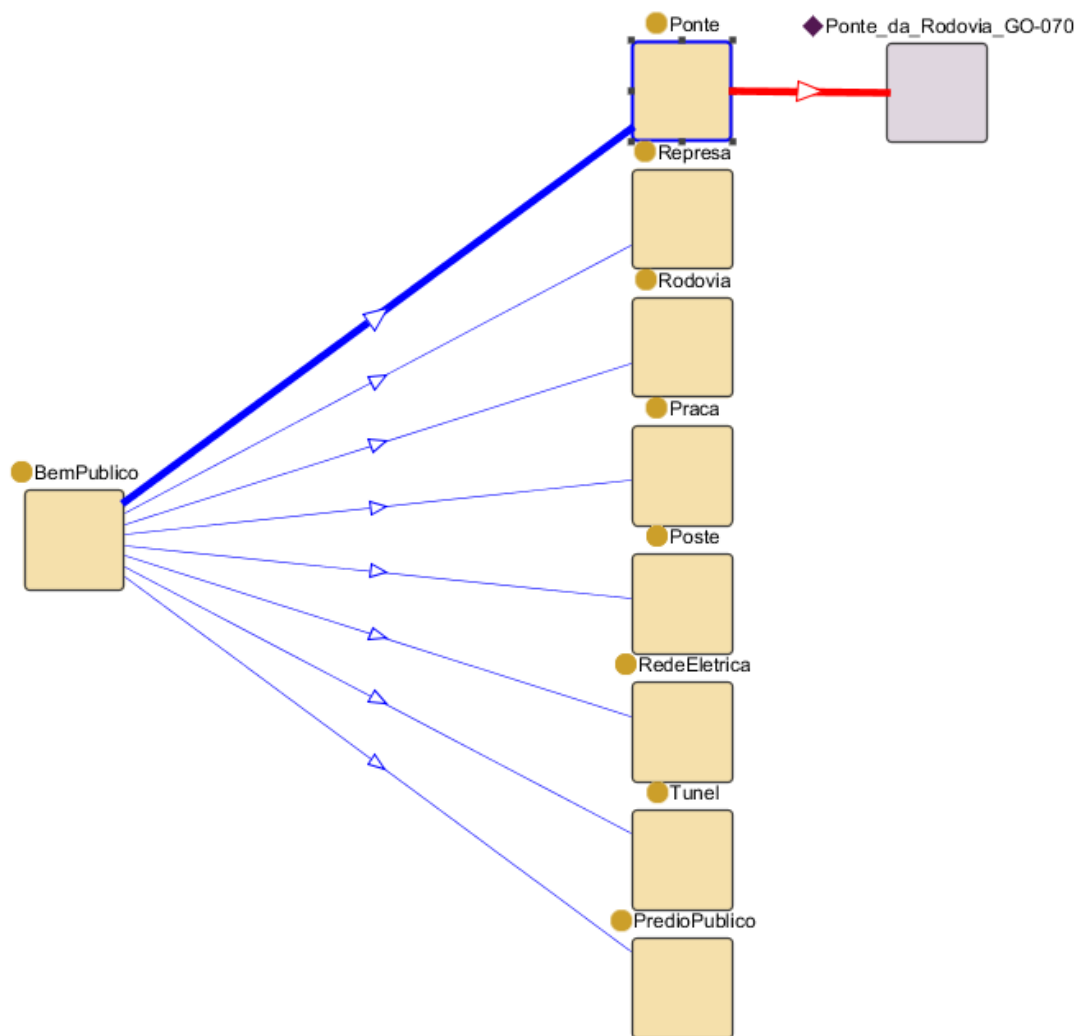
Instância: Goiás

O rio das Pedras é o Recurso Natural envolvido como agente da queda da Ponte. Aqui a característica dele não foi o transbordamento e sim sua correnteza.

RecursoNatural >> RecursosHidricos >> Rio

Instância: Rio das Pedras

Figura 25 – Identificação de indivíduos no Protégé



Através do *plugin* Jambalaya no Protégé pode-se ver superclasses, classes e suas instâncias. Esta ferramenta foi utilizada para realizar a visualização das instâncias geradas pela análise da notícia exemplo.

Figura 26 – Visualização dos indivíduos das classes Recursos Hídricos e Bem Público

Sob a ótica da ontologia construída, a notícia sobre o desabamento da ponte adquire elementos adicionais semânticos, não presentes textualmente no RSS original, os quais possibilitam uma melhor análise do conteúdo informativo (figuras 26 e 27).

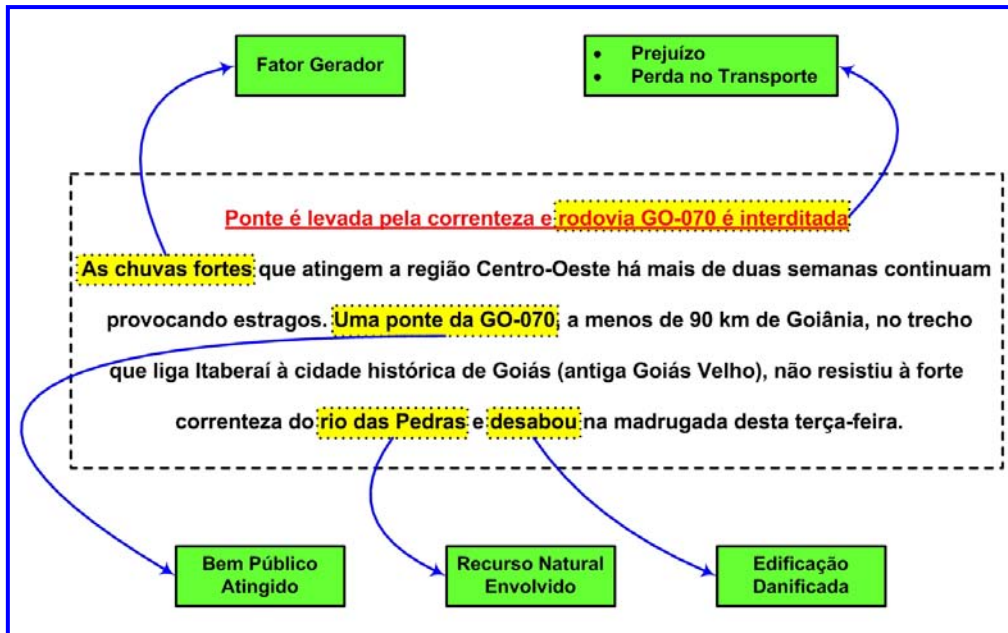


Figura 27 – Classificação dos termos presentes no corpúsculo da notícia

## 2.5 Funcionamento do OntoFeed

Na página do projeto, é vista a dupla operação do OntoFeed:

- Ao lado esquerdo, sob o título “últimas notícias”, seu funcionamento é o mesmo de um leitor comum de *Feeds*, identificando as notícias dos mais variados temas com Data, Hora, Portal e o Título do *Feed* (que ao ser selecionado remete para a página onde se encontra a notícia completa). O usuário pode atualizar o recebimento dos *Feeds* a qualquer momento, bastando para isto clicar sobre o canto direito superior da página do OntoFeed e novas notícias oriundas dos portais cadastrados serão apresentadas.
- Ao lado direito, sob o título perguntas da ontologia, acham-se as perguntas pré-determinadas pelo administrador, onde o usuário poderá escolher uma de cada vez, para que o sistema responda. Para responder à pergunta selecionada pelo usuário, o sistema efetuará uma busca nas notícias armazenadas no banco de dados, procurando as palavras-chave cadastradas pelo administrador e que estão ligadas àquela pergunta, selecionada pelo usuário. O usuário clicará sobre o botão “Busca Respostas” e o que houver no banco de dados como relevante à pergunta será mostrado como se pode ver na figura 29, sob o título “Respostas da Ontologia”.

**ONTOFeed**

**Últimas notícias**

14/06 12:06h | Folha - Em cima da hora Santos desembarca em Montevideu e final sera na quarta

14/06 12:06h | Folha - Em cima da hora Lily Allen adota sobrenome do marido após o casamento

14/06 12:06h | Folha - Em cima da hora Médico que declara laço com indústria também é antiético

14/06 12:06h | Folha - Em cima da hora Presos morrem em confronto em prisão na Venezuela

14/06 12:06h | G1 - São Paulo Aprenda a fazer bolo de pinhão para festas juninas

14/06 12:06h | R7 - São Paulo PM persegue suspeitos de roubo a carro dos Correios na zona norte de SP

14/06 12:06h | Folha - Em cima da hora Acordo sobre patentes com Apple é vitória rara para Nokia

14/06 12:06h | G1 - Rio de Janeiro Polícia descobre desmanche de carros em sitio na Zona Oeste do Rio

14/06 12:06h | Folha - Em cima da hora Líderes neonazistas são condenados a prisão perpétua na Rússia

14/06 12:06h | G1 - Rio de Janeiro Aneel autoriza alta na tarifa de energia de Nova Friburgo (RJ)

14/06 12:06h | O Globo - País Sobre documentos secretos, Sarney diz que não se pode fazer "WikiLeaks" da história do Brasil

14/06 12:06h | G1 - São Paulo Fundação faz campanha para incentivar doação de sangue em SP

14/06 12:06h | G1 - São Paulo Com voos suspensos, santistas planejam ir de carro para o Uruguai

14/06 12:06h | R7 - Rio de Janeiro Polícia encontra desmanche de veículos em Sepetiba

**Perguntas da Ontologia**

Quais recursos hídricos estão presentes nesta notícia ?

Quais índices pluviométricos esta notícia detectou ?

As chuvas provocaram falta de energia elétrica ?

Pessoas ou bens ficaram soterrados ?

BuscaRespostas

Selecione uma pergunta da Ontologia

Figura 28 – OntoFeed: à esquerda é Leitor Comum de Feeds e à direita com Extensão Ontológica

A figura 28 mostra este funcionamento do sistema com a pergunta 1 da ontologia (“Quais recursos hídricos estão presentes nesta notícia?”) sendo assinalada. Após busca no banco de dados o OntoFeed apresenta vários resultados hachurando em amarelo as palavras-chave do contexto da ontologia que respondem à pergunta selecionada (figura 29). As respostas encontradas pelo sistema são agrupadas na página e cada uma delas possui os campos: Data, Hora, Portal, e os *links* Título e Ver Notícia. No *link* Ver Notícia, a seleção do usuário, o remete à página completa da notícia, ao passo que, selecionando Título, o sistema abre a notícia - *Feed* e apresenta para o usuário o que consta no item “*description*” do RSS.



**Perguntas da Ontologia**

Quais recursos hídricos estão presentes nesta notícia ?

Quais índices pluviométricos esta notícia detectou ?

As chuvas provocaram falta de energia elétrica ?

Pessoas ou bens ficaram soterrados ?

BuscaRespostas

**Respostas da Ontologia**

24/01 00:01h | R7 - São Paulo | [ver notícia](#)  
Marginal Tietê e Penha entram em alerta por causa da chuva forte  
Subprefeitura da Penha notificou transbordamento do **rio**, Tietê na região  
Resposta: **Recurso Natural Hídrico**

19/01 00:01h | G1 - São Paulo | [ver notícia](#)  
Após chuva, carro é encontrado dentro de **córrego** no ABC

19/01 00:01h | G1 - São Paulo | [ver notícia](#)  
Nível de **rios** sobe e CGE coloca três subprefeituras de SP em atenção

15/01 00:01h | R7 - São Paulo | [ver notícia](#)  
Mesmo sem chuva, nível do **rio** Atibaia sobe 10 cm em um dia

12/01 00:01h | R7 - Rio de Janeiro | [ver notícia](#)  
População de Itaipava diz que sempre foi prejudicada por chuvas

12/01 00:01h | R7 - Rio de Janeiro | [ver notícia](#)  
Três Rios entra em estado de alerta

08/01 00:01h | R7 - São Paulo | [ver notícia](#)  
Após chuvas, avenida Aricanduva ainda tem dois pontos de alagamento

08/01 00:01h | R7 - São Paulo | [ver notícia](#)  
Cidades do interior de São Paulo ainda consertam estragos causados pelas chuvas do ano passado

04/01 00:01h | G1 - São Paulo | [ver notícia](#)  
Bombeiros buscam homem que sumiu após chuva no interior de SP

04/01 00:01h | R7 - São Paulo | [ver notícia](#)  
Leito de **rio** em São José dos Campos está acima do nível normal

14/12 00:12h | G1 - São Paulo | [ver notícia](#)  
Interior de SP ainda sofre com os estragos das chuvas

Figura 29 – Respostas da Ontologia à Pergunta 1

No *Feed* de 00:01 horas de 24/01, do portal R7, com o título: “Marginal Tietê e Penha entram em alerta por causa da chuva forte”, vê-se hachurado em amarelo o termo “rio”, que embora não conste no título do *Feed* está presente em sua descrição. Abaixo desta descrição, OntoFeed acrescenta a informação “Recurso Natural Hídrico”, que aponta para a classificação ontológica executada com base na doutrina da Defesa Civil. Já no segundo *Feed* trazido pelo sistema como resposta à pergunta 1, o termo hachurado é “córrego” e ele consta no título da notícia.

Com o objetivo de ligar as notícias informadas via *Feed* ao local onde estas ocorreram, utilizou-se além da API do Google Maps o cadastramento de duas tabelas com os bairros componentes dos municípios de São Paulo e do Rio de Janeiro. Isto fez com que o sistema

adquirisse algumas funcionalidades dos sistemas de informação geográfica. Assim as notícias que ocorrerem nas cidades de São Paulo ou Rio de Janeiro e contiverem em seu título ou em sua descrição, os bairros onde aconteceram, receberão o acréscimo do item “Ver Mapa”, (figura 30) na linha onde estão as outras informações do RSS, como: hora, data, portal e *link*.

**Perguntas da Ontologia**

- Quais recursos hídricos estão presentes nesta notícia ?
- Quais índices pluviométricos esta notícia detectou ?
- As chuvas provocaram falta de energia elétrica ?
- Pessoas ou bens ficaram soterrados ?

BuscaRespostas

---

**Respostas da Ontologia**

24/01 00:01h | R7 - São Paulo | ver notícia | **ver mapa**  
 Marginal Tietê e Penha entram em alerta por causa da chuva forte  
 Subprefeitura da Penha notificou transbordamento do rio Tietê na região  
 Resposta: **Recurso Natural Hídrico**

19/01 00:01h | G1 - São Paulo | ver notícia  
 Após chuva, carro é encontrado dentro de córrego no ABC

19/01 00:01h | G1 - São Paulo | ver notícia | ver mapa  
 Nível de rios sobe e CGE coloca três subprefeituras de SP em atenção

15/01 00:01h | R7 - São Paulo | ver notícia  
 Mesmo sem chuva, nível do rio Atibaia sobe 10 cm em um dia

12/01 00:01h | R7 - Rio de Janeiro | ver notícia  
 População de Itaipava diz que sempre foi prejudicada por chuvas

12/01 00:01h | R7 - Rio de Janeiro | ver notícia  
 Três Rios entra em estado de alerta

08/01 00:01h | R7 - São Paulo | ver notícia | ver mapa  
 Após chuvas, avenida Aricanduva ainda tem dois pontos de alagamento

08/01 00:01h | R7 - São Paulo | ver notícia  
 Cidades do interior de São Paulo ainda consertam estragos causados pelas chuvas do ano passado

04/01 00:01h | G1 - São Paulo | ver notícia  
 Bombeiros buscam homem que sumiu após chuva no interior de SP

04/01 00:01h | R7 - São Paulo | ver notícia  
 Leito de rio em São José dos Campos está acima do nível normal

14/12 00:12h | G1 - São Paulo | ver notícia  
 Interior de SP ainda sofre com os estragos das chuvas

Figura 30 – Item “Ver Mapa”

Caso o usuário clique sobre o item “Ver Mapa”, o sistema localizará a notícia no Google Maps e a exibirá com um ponto identificador (na forma de balãozinho) vermelho. Desta forma após a busca das respostas da ontologia, existindo algum RSS (que seja de algum bairro pré-cadastrado), a API do Google Maps fornecerá a latitude e a longitude do bairro e marcará os pontos no objeto GMap, fornecido pela própria API. Os pontos que forem marcados serão do tipo *LabeledMarker*, que funcionam como uma extensão do objeto *Gmarker*, que permite a adição de *Labels* (texto identificador). Para facilitar a visualização

do usuário em cada ponto marcado é adicionado o título da respectiva notícia (figura 31).



Figura 31 – Localização da Região de Referência da Notícia

## 2.6 Considerações Finais

Este capítulo apresentou a aplicação OntoFeed, que amplia os recursos de um leitor de *Feeds* comum, proporcionando buscas semânticas em banco de dados via *Web*, a partir de perguntas cadastradas. Este serviço agrega valor aos *Feeds*, investigando tanto em seu título como na sua descrição, palavras-chave do contexto da ontologia do domínio em uso. Com este recurso, o usuário pode interagir com o sistema selecionando dentre as perguntas disponíveis, uma de cada vez e verificar o funcionamento do OntoFeed, quando este associa à pergunta os *Feeds* relevantes para respondê-la apresentando-os em ordem de data como visto anteriormente. No próximo capítulo apresentam-se as conclusões desta dissertação, suas limitações e trabalhos futuros.