

**Uma Arquitetura para Desenvolvimento da Web
Semântica Baseada em Comunidades Virtuais de
Prática:
Projeto *DWeb - Dream Web*.**

José Mauro da Silva Cedric Luiz de Carvalho

Technical Report - RT-INF_001-06 - Relatório Técnico
February - 2006 - Fevereiro

The contents of this document are the sole responsibility of the authors.
O conteúdo do presente documento é de única responsabilidade dos autores.

Uma Arquitetura para Desenvolvimento da Web Semântica Baseada em Comunidades Virtuais de Prática: Projeto *DWeb - Dream Web*.

José Mauro da Silva *
josemauro@inf.ufg.br

Cedric Luiz de Carvalho †
cedric@inf.ufg.br

Abstract. *The World Wide Web, or simply Web, has been designed to support information resource sharing at global level. Nowadays, this sharing is limited, mainly respect to imperfect information recovery and to the “Hidden Web”. The Semantic Web, comes up as a possible solution for such limitations. This text proposes a conceptual architecture, entitled DWeb, that aims to provide an environment to eases generic application development, allowing to gather intelligence, interoperation and integration, according to the principles of Semantic Web, in the context of virtual communities of practice.*

Keywords: Semantic Web, Community Web, Conceptual Models, Semantic Portal, Ontology, Metadata, XML, RDF, Middleware Services, CORBA Architecture, Cluster, Plug-ins and Ubiquitous Computation.

Resumo. *A World Wide Web, ou simplesmente Web, foi projetada para suportar compartilhamento de recursos de informação em nível global. Atualmente, este compartilhamento é limitado, principalmente, no que tange à recuperação de informações imperfeitas e à Web Oculta. A Web Semântica, surge como uma possível solução para tais limitações. Este texto propõe uma arquitetura conceitual, intitulada DWeb, que visa criar um ambiente de desenvolvimento de aplicações genéricas, dentro dos princípios da Web Semântica, que permita agregar inteligência, interoperação e integração ao ambiente Web existente, a partir de um modelo conceitual de comunidades virtuais de prática.*

Palavras-Chave: Web Semântica, Comunidades Web, Modelos Conceituais, Portal Semântico, Ontologias, Metadados, XML, RDF, Serviços de Middleware, Arquitetura CORBA, Cluster, Plug-ins e Computação Ubíqua.

*Mestrando em Ciência da Computação - GEApIS-INF/UFG.

†Orientador - GEApIS-INF/UFG.

1 Introdução

As limitações tecnológicas da *Web* são notórias e amplamente discutidas pela comunidade científica. Dentre estas limitações pode-se destacar os vários problemas relacionados à recuperação de informações imperfeitas¹[1] e a *Web Oculta*²[2].

O W3C[22] é um consórcio que desenvolve esforços na definição de padrões para a *Web* (especificações, diretrizes, softwares e ferramentas) e serve como um fórum aberto para discussão sobre a *Web*. Um dos seus grupos de trabalho, o grupo *Semantic Web Activity Statement*[19] tem trabalhado no sentido da propositura de uma nova *Web*, resultado da evolução da *Web* atual, à qual se tem dado o nome de **Web Semântica**. Ela pode ser definida da seguinte forma:

“A Web Semântica é uma visão: é a idéia de se ter dados sobre a Web bem definidos e ligados de uma maneira tal que possam ser usados por máquinas não só com o objetivo de apresentação, mas para automação, integração e reutilização dos dados entre aplicações”[18].

Além dessa definição, foram desenvolvidas diversas tecnologias interoperáveis, que atualmente estão estruturadas em uma arquitetura conceitual[7] sobre a qual a *Web Semântica* está sendo desenvolvida. Esta arquitetura (Figura 1), é uma estrutura que está sendo construída pela comunidade científica em diversos centros de pesquisas, universidades e organizações em nível mundial. Ela visa a definição de padrões que permitam posteriormente a comunicação plena entre os pares que transacionam recursos de informação sobre a *Web Semântica* [18] (informações adicionais podem ser encontradas em [21], [29], [17], [8], [16]).

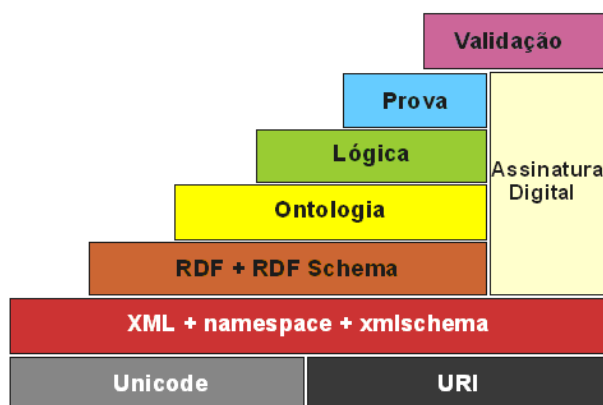


Figura 1: Arquitetura Padrão do W3C para a Web Semântica.[7]

Nesta estrutura, até o momento, já se encontram definidas as cinco camadas inferiores, ou seja: *Unicode*, *URI*, *XML + NameSpace + XMLSchema*, *RDF + RDFSchema* e a camada *Ontologia*.

A camada *Unicode* trata da formatação dos caracteres. Nesta camada se encontra o formato *UTF-8*, que representa os caracteres em 1 *byte* e *UTF-16*, que representa os caracteres em 2 *bytes* e que permite a representação de caracteres em todos os idiomas conhecidos atualmente. Esta camada permite a representação de dados observando-se os aspectos da internacionalização das aplicações *Web*[13].

¹Recuperação de informações imperfeitas se refere às limitações apresentadas pelos mecanismos de buscas que trabalham somente com buscas baseadas em palavras-chaves. Estas buscas são limitadas a avaliações sintáticas, o que leva à uma baixa relevância das informações retornadas com relação aos interesses dos usuários.[1]

²A *Web Oculta*, refere-se ao conjunto de informações da *Web*, apresentadas aos usuários através de páginas dinâmicas que acessam bancos de dados e que não são visualizadas pelos mecanismos de recuperação de informações atuais.

A camada *URI (Uniform Resource Identifier)*, permite a vinculação semântica de recursos identificáveis na Web Semântica. Um *URI* possibilita a referência tanto de recursos de informação internos quanto externos à Web Semântica. Um recurso de informação pode ser qualquer coisa que tenha uma identidade, podendo ser abstrato ou físico. Como exemplos de recursos de informação abstratos, pode-se citar: texto eletrônico, imagem, serviço ou uma coleção de recursos. Dentre os recursos de informação físicos, pode-se citar: seres humanos, empresas ou livros. Um *URI* possibilita a minimização de problemas relacionados à ambigüidades, principalmente na montagem de triplas *RDF* [6], [17].

A camada *XML (eXtensible Markup Language)*, *XMLSchema* e *NameSpace*, define a sintaxe dos documentos transacionados e o espaço de nomes onde os mesmos são definidos, evitando-se, dessa forma, redefinições desnecessárias e conflitos de definições de termos ou estrutura de dados. Esta camada garante a interoperabilidade sintática no ambiente da Web Semântica[21], [24].

A camada *RDF (Resource Description Framework)* e *RDFSchema*, define a estrutura dos dados transacionados, ou seja, os metadados[24]. Os metadados, definidos como dados sobre dados[16], garantem a interoperabilidade estrutural e, desta forma, proporciona o compartilhamento de recursos de informação “compreendidos” por máquina[17].

A camada Ontologia, facilita o entendimento, via vocabulário adicional, através do qual uma semântica formal é definida. Um ontologia, segundo *Tom Gruber*, é uma especificação explícita e formal de uma conceitualização compartilhada[29]. Essa semântica formal, pode ser materializada no ambiente da Web Semântica através da linguagem padrão para construção de ontologias *OWL (Web Ontology Language)* [8], [20].

As camadas superiores: Lógica, Prova, Verdade e Assinatura Digital (Segurança), ainda estão em processo de definição e desenvolvimento pela comunidade científica.

O Projeto *DWeb (Dream Web)*, neste contexto, visa contribuir com a consolidação e o desenvolvimento da Web Semântica oferecendo um ambiente para o estabelecimento de Comunidades Virtuais de Prática (CoPs) na Internet. Este ambiente deverá facilitar a comunicação entre os membros das comunidades e entre as comunidades com a utilização de componentes semânticos associados aos recursos disponibilizados. Melhorando-se o processo de comunicação espera-se tornar mais ágil e eficiente o desenvolvimento de atividades colaborativas dentro e fora de comunidades específicas.

No desenvolvimento da arquitetura *DWeb* deve-se observar os princípios de projeto especificados por Tim Berners-Lee para a *Web* existente, que são: simplicidade, modularidade, descentralização e tolerância, citados em [4] e [3], bem como, os padrões tecnológicos adotados para a arquitetura oficial da Web Semântica proposta pelo W3C em [18], [5], [8].

O *DWeb*, atua sobre os recursos de informação que representam as unidades de informação do ambiente. Estas unidades de informação são referenciadas através de metadados que descrevem os recursos. O uso dos metadados possibilita a atuação de um mecanismo de inferência (componente da arquitetura *DWeb*) tornando processo de recuperação de informações mais preciso por meio de avaliações semânticas sobre estes recursos de informação.

Os recursos de informação, na arquitetura *DWeb*, representam informações estruturadas, semi-estruturadas e não estruturadas. No caso das informações estruturadas, os recursos referenciam conteúdos informacionais armazenados em banco de dados. Através de mecanismos de controles de segurança em profundidade (camadas de segurança)[37], onde as unidades de informação são controladas, cria-se, a possibilidade de minimizar os problemas relacionados à *Web Oculta*, pois os recursos de informação são protegidos de forma individual sendo, desta forma, visualizados de maneira seletiva e controlada. A definição dos aspectos relacionados à segurança sobre os recursos de informação são mapeados pelo próprio criador do recurso.

O *DWeb* está sendo proposto como uma arquitetura para desenvolvimento de aplicações genéricas. As aplicações desenvolvidas se utilizam de serviços desenvolvidos para o modelo de serviço do *DWeb*, o que minimiza o trabalho dos desenvolvedores de aplicações computacionais sobre esta infra-estrutura e, por outro lado, possibilita a minimização dos problemas relacionados à recuperação de informações e à *Web Oculta*.

O *DWeb* trabalha de forma orientada a comunidades virtuais de prática. As comunidades virtuais surgiram em meados dos anos 80 com a criação da USENET entre centros de computação de universidades que utilizavam o sistema operacional UNIX. Uma de suas funções era distribuir notícias sobre vários tópicos referentes à rede.

No início foram desenvolvidos grupos de discussão em ambientes não comerciais. A CompuServ, nos anos 80, explorou comercialmente um grande número de fóruns que permitiam compartilhamento de interesses pessoais e profissionais. No início dos anos 90, a *América on-line (AOL)* se apresentava como um serviço de acesso fácil para um grande contingente de usuários. A publicação do livro *The Virtual Community* [41] por Howard Rheingold em 1994, introduziu o conceito para a comunidade em geral. Ele foi um dos primeiros a declarar que as redes *on-line* estavam emergindo como uma força social muito importante que poderia disponibilizar as experiências de comunidades ricas e autênticas.

As comunidades virtuais se classificam nos seguintes tipos: comunidades virtuais de relacionamentos (que são construídas sobre relacionamentos especiais entre pessoas, como por exemplo relações de família, etc.), comunidades virtuais de lugar (que são baseadas em indivíduos que compartilham o mesmo *habitat* ou local), comunidades virtuais de conhecimentos (que ajudam, de forma séria, a encontrar pessoas com os mesmos objetivos, valores e concepções sobre determinado assunto), comunidades virtuais de memória (que são baseadas em um passado compartilhado ou algo histórico) e, finalmente comunidades virtuais de necessidades (que são baseadas em algum fato ou acontecimento semelhante e que tenham algum fator que afete emocionalmente os participantes).

Em [33], comunidades de prática estão definidas como coleções de indivíduos agrupados por relações informais que compartilham tarefas de trabalho semelhante e um contexto comum.

No Projeto *DWeb* são adotadas as tecnologias da Web Semântica dentro do contexto de comunidades virtuais de prática. Estas fornecem um modelo bastante adequado para se criar um contexto bem definido de forma a permitir a recuperação de informações (recursos de informação) com alto nível de relevância para as necessidades dos usuários.

O restante deste texto é organizado em 5 seções de forma a fornecer uma visão conceitual sobre os problemas da *Web* hoje, os avanços esperados com o desenvolvimento da Web Semântica e, finalmente, propor uma arquitetura conceitual que contemple estes avanços. A Seção 1 apresenta os fundamentos da *Web* existente e da Web Semântica; a Seção 2 apresenta o Projeto *DWeb* e destaca as tecnologias que dão suporte aos macro-requisitos da Web Semântica, que são: inteligência, interoperabilidade e integração; na Seção 3, destaca-se o arcabouço sobre o qual o Projeto *DWeb* implementa a sua arquitetura de funcionalidades genéricas e as vantagens de sua arquitetura ser relacionada a comunidades virtuais de prática; a Seção 4 detalha a arquitetura em três camadas do projeto *DWeb*, destacando seus principais componentes e, finalmente, na Seção 5, são feitas as considerações finais e uma conclusão sobre a arquitetura proposta.

2 O Projeto *DWeb*

O Projeto *DWeb* propõe o uso conjunto das tecnologias da Web Semântica e conceitos de comunidades virtuais de prática como forma de minimizar os problemas da *Web* existente. Ele visa contribuir com o desenvolvimento da Web Semântica em conformidade com os padrões da

comunidade científica, em especial, os padrões divulgados pelo W3C. A seguir é fornecida uma visão geral sobre o Projeto DWeb e demonstrada a forma como ele se insere no contexto da Web Semântica.

2.1 Visão Geral do Projeto DWeb

Os requisitos para o Projeto DWeb foram estabelecidos em conformidade com os requisitos para um projeto do tipo SWMS (*Semantic Web Management System*), o qual é um tipo particular de ambiente de software orientado a ontologia, que tem como principal característica, ser especialmente projetado para ajudar no desenvolvimento de aplicações da Web Semântica[38].

2.1.1 Cenários de Motivação

Projetos³ são frequentemente implementados através de alguma metodologia, onde se identifica o problema e parte-se para a busca de uma solução[45]. Pode-se introduzir o Projeto DWeb, a partir de um cenário imaginário, onde se configure um problema passível de identificação e, em seguida, propor uma solução conceitual para o problema identificado.

Pode-se imaginar alguns cenários onde profissionais de diferentes domínios de conhecimento estejam trabalhando. Por exemplo, um médico está trabalhando em pesquisas relacionadas à cirurgia oftálmica, outro em cirurgia cardíaca e outro em cirurgia do aparelho digestivo. Todos fazem parte da comunidade de médicos. Esta comunidade de médicos tem em seu domínio de conhecimento (domínio medicina) um grupo de sub-domínios, que podem ser classificados em: sub-domínio cirurgia oftálmica, sub-domínio cirurgia cardíaca e sub-domínio cirurgia do aparelho digestivo. Em uma análise mais aprofundada pode-se observar que estes sub-domínios têm conhecimentos específicos de cada sub-domínio e conhecimentos comuns a todos os demais e que poderiam ser categorizados como sub-domínios de cirurgia geral. Cada um desses médicos pode estar trabalhando em uma localidade distinta e nem sequer saber da existência dos outros.

Um dos médicos chegou a alguns resultados relacionados à sua pesquisa e gostaria de compartilhá-los com outros profissionais, ou verificar se algum outro médico já obteve resultado semelhante. O problema é que, para aceitação das suas experiências em nível internacional, ele precisaria publicar o trabalho em organizações de renome mundial, o que demandaria um certo tempo, o qual ele não possui, pois as verbas para as suas pesquisas têm um prazo para serem utilizadas e qualquer tempo consumido é importante para a sua plena utilização.

Em um cenário no domínio de conhecimento do direito, ocorre um problema com uma pessoa para a qual um advogado está montando um processo jurídico sobre erro médico. Casualmente o advogado é um deficiente com baixa acuidade visual e necessita de ferramentas alternativas de comunicação. Para montar o processo jurídico, o advogado necessita saber de algumas informações relativas a um determinado diagnóstico ligado à cirurgia oftálmica e buscar informações do domínio do direito processual, especificamente, onde se possa encontrar jurisprudência sobre o assunto. O problema é que estas informações devem estar contidas no domínio do direito internacional, pois o médico envolvido na questão é de outro país. Um outro lado do problema é que este processo jurídico tem prazo muito restrito para a sua montagem e corre o risco de prescrever, caso não seja dada a sua entrada em tempo hábil. Outro ponto

³Projeto é um empreendimento temporário com o objetivo de criar um produto ou serviço único. Temporário significa que cada projeto tem prazo definido. Único significa que o produto ou serviço é, de alguma forma, diferente de todos os outros produtos ou serviços semelhantes.[45]

importante do problema são as necessidades especiais do próprio advogado, que pode ocasionar dificuldades na aquisição das informações necessárias à montagem do processo jurídico.

Em um cenário no domínio de conhecimento de jornalismo, um jornalista está escrevendo uma matéria sobre erros médicos e necessita de informações sobre o assunto. O jornalista precisa escrever a matéria com enfoque na área médica e jurídica, desta forma, necessita de informações verídicas, com registros históricos, por localidades, por sintomas, com amplitude mundial do domínio medicina e o respectivo veredito do domínio do direito. Para tanto, ele necessita fazer algumas análises sobre as informações de forma a conseguir identificar, semanticamente, erros médicos por categorias de sintomas, porém, de antemão, ele não dispõe destas categorias. O problema é como conseguir informações, em tempo hábil, no contexto preciso (domínio erros médicos), com abrangência mundial e local, com informações verídicas e com alto nível de credibilidade para a publicação da matéria.

Os problemas identificados até então, de uma forma geral, estão relacionados à comunicação entre as pessoas. De uma maneira mais específica o problema é, como possibilitar a estes especialistas em medicina o compartilhamento de seus conhecimentos e experiências sobre cirurgia, de forma transparente, formal, coerente, dinâmica, contextualizada, abrangente e organizada, garantindo-se a possibilidade de aprimoramentos individuais e em grupo de forma democrática. Neste mesmo contexto, como permitir que o advogado obtenha acesso a informações e/ou conhecimentos de um outro domínio de conhecimento da mesma forma que no seu domínio, mesmo sendo uma pessoa com necessidades especiais devido à sua baixa acuidade visual. E ao jornalista, além do compartilhamento de domínio de conhecimento distintos, como obter acesso à informações sobre o seu domínio de conhecimento e de outros domínios com possibilidade de avaliação semântica destes, com respostas rápidas e com alto nível de confiança.

Em todos os casos a solução tradicional seria a busca de informações documentárias (refere-se de uma forma geral a livros, artigos técnicos/científicos, documentação da área, etc.) específicas de cada domínio do conhecimento, mas, em todos os cenários, os profissionais não dispõem de tempo para averiguações exaustivas, informações de domínios de conhecimento distintos e que tenham uma abrangência em nível mundial.

Em uma análise menos atenta poderia ser proposta como solução a implementação de um sistema computacional tradicional, provavelmente baseado na *Web*. Em uma análise mais elaborada, poder-se-ia notar que a *Web* atual possui diversas limitações tecnológicas, principalmente no que se refere à recuperação de informações baseadas em palavras chaves[1] e à *Web* Oculta[2], o que poderia, em última análise, acarretar problemas a estes profissionais, principalmente, em relação ao tempo de busca, à qualidade e o tratamento semântico das informações recuperadas, notadamente, no que tange à relevância sobre cada tema tratado.

Desta forma, pode-se identificar que estes profissionais necessitam de um mecanismo de comunicação com as seguintes características:

- Facilidades quanto ao registro, intercâmbio, organização e gestão de conhecimentos e que tal mecanismo considere alguns requisitos inerentes às dificuldades de comunicação⁴ de cada pessoa;
- Acesso à informações de uma forma totalmente transparente, democrática, intuitiva e integrada em domínios do conhecimentos distintos;
- Abrangência global através da infra-estrutura da Internet existente;

⁴Refere-se às pessoas idosas, usuários de navegadores alternativos, usuários de tecnologias assistivas, etc.

- Delimitação de contexto (escopo) dos conhecimentos tratados em domínio específicos;
- Utilização de tecnologias da *Web* existente acrescida dos aprimoramentos da *Web Semântica*.

Assim, a maioria dos problemas identificados pelos cenários apresentados poderiam ser minimizados ou extintos. É neste contexto que o Projeto *DWeb* se propõe colaborar na resolução dos problemas observados na *Web* atual. Porém, para atingir estes objetivos deve-se levar em consideração alguns requisitos de projeto como se pode ver a seguir.

2.1.2 Requisitos Para o Projeto DWeb

Basicamente, pode-se estabelecer quatro grupos de requisitos. Os clientes da aplicação devem se conectar remotamente ao *SWMS* e serem apropriadamente autorizados. Conseqüentemente, requisitos de conectividade, acessibilidade e segurança são necessários.

Por outro lado, um sistema *SWMS* deve responder aos aspectos estáticos da *Web Semântica* (camadas definidas pelo *W3C*[7]). Em particular, deve oferecer suporte a todas as linguagens que já estão padronizadas pelo *W3C*. Uma propriedade desejável é a tradução entre linguagens diferentes, o que aumenta a interoperabilidade entre módulos de software existentes, principalmente, os que estão focados em uma linguagem específica.

Os aspectos dinâmicos da *Web Semântica*⁵, resultam em um outro grupo de exigências provenientes dos acessos, armazenamentos e integração de dados, bem como a consistência, concorrência, durabilidade e raciocínio.

Outro aspecto importante se refere à necessidade de delimitação de contexto para implementação de sistemas de raciocínio lógico.

Finalmente, é esperado que o sistema facilite a criação de uma infra-estrutura de *plug-ins*, que seja extensível. No último grupo de exigências estão as transações com manipulação flexível de módulos.

Lista Classificada de Requisitos de Projeto:

1. Contexto

- Contextualização: os sistemas baseados em conhecimento requerem uma delimitação contextual precisa para que sejam capazes de fornecer resultados coerentes e corretos. Desta forma, o Projeto *DWeb* deve fornecer um mecanismo que facilite esta delimitação contextual em diversos domínios de conhecimento simultaneamente. Com esta contextualização é possível minimizar os problemas originados da ambigüidade ocorridos durante o tratamento semântico dos recursos de informação.

2. Conectividade, Acessibilidade e Segurança

- Conectividade: o Projeto *DWeb* deve habilitar o acoplamento fraco entre aplicações, onde se permite acesso através de protocolos que atendam aos padrões da *Web*, bem como, o acoplamento forte embutidos dentro de outras aplicações. Isto pode ser feito oferecendo-se um padrão sofisticado de *API*. Em outras palavras, um cliente deve poder obter acesso local ao sistema e/ou conectar-se remotamente através de serviços *Web*, por exemplo.

⁵Refere-se aos aspectos de validação, inferência, monitoramento, versionamento, evolução, transação, acesso, etc.[38]

- **Acessibilidade:** é a capacidade de um produto ser flexível o suficiente para atender às necessidades e preferências do maior número possível de pessoas, além de ser compatível com tecnologias assistivas usadas por pessoas com necessidades especiais[14]. Neste contexto, este requisito visa a inclusão digital de um maior número possível de pessoas através da habilitação da acessibilidade no uso do ambiente do Projeto *DWeb*.
- **Segurança:** garantir segurança de informação significa proteger a informação contra revelação sem autorização, transferência, modificação, ou destruição, seja acidental ou intencional. Para realizar isto, qualquer operação deve somente ser acessível para clientes devidamente autorizados. Mecanismos de identificação devem ser estabelecidos através do emprego de técnicas de autenticação. Dados confidenciais devem ser criptografados para comunicação e armazenamento persistente. Finalmente, devem existir meios para monitorar (*logging*) operações confidenciais[37].

3. Linguagens da Web Semântica

- **Suporte a Linguagens:** uma exigência trivial é a de que o Projeto *DWeb* deve dar suporte a todos os padrões da Web Semântica, tais como ontologias e metadados. Ele deve prover suporte aos padrões da Web Semântica existente como *RDF*, *RDFS* e *OWL*, enquanto, deve também ser flexível o suficiente para suportar linguagens futuras resultantes das especificações das camadas de lógica, prova e verdade.
- **Interoperabilidade Semântica:** usa-se o termo interoperabilidade semântica no sentido de indicar uma tradução entre linguagens de diferentes ontologias e diferentes semânticas. Neste contexto, além das linguagens proprietárias, existem também as linguagens construídas especificamente para a Web Semântica, que são: *RDFS*, *OWL Lite*, *OWL DL* e *OWL Full*. Usualmente editores de ontologia são focados para trabalhar com uma linguagem em particular e não possuem habilidades de trabalhar com outras linguagens. Conseqüentemente, o *DWeb* deve permitir tradução entre linguagens e semânticas diferentes. Um problema a ser considerado é que freqüentemente uma tradução não pode ser alcançada sem perda de informação.
- **Mapeamento de Ontologias:** Em contraste com Interoperabilidade Semântica, o mapeamento visa a tradução entre diferentes ontologias de uma mesma linguagem. Certas comunidades, usualmente, têm suas próprias ontologias e poderiam usar o Mapeamento de Ontologias para facilitar o intercâmbio de dados.

4. Gerenciamento de Dados

- **Armazenamento de Ontologias:** tipicamente, aplicações para Web Semântica devem ter acesso e serem capazes de armazenar dados ontológicos. Adicionalmente, uma ontologia para um domínio pode ser construída sobre outras ontologias, com outros níveis de abstrações e outros conceitos. O Projeto *DWeb* deve possibilitar o armazenamento de ontologias a serem oferecidas para as aplicações do ambiente.
- **Integração:** assume-se que o ambiente da Web Semântica seja distribuído e composto por fontes heterogêneas de recursos de informação. Portanto, o sistema *DWeb* deve ter a capacidade de acessar e agregar informações distribuídas e heterogêneas. Meios para detectar e trabalhar com conjuntos de dados duplicados são também requeridos. Por exemplo, os mesmos dados podem ser acessados através de serviços de raciocínios diferentes, os quais podem fornecer serviços únicos e distintos para

dados. Também, meios para consultar e compor dados distribuídos devem ser fornecidos. Assim, devem ser disponibilizados recursos que contemplem a homogeneização das fontes heterogêneas de informação. Para suprir este requisito, o Projeto *DWeb* deve contemplar o mecanismo de integração de fontes heterogêneas de dados referenciado como Mediador/Tradutor[48].

- **Localização:** considerando-se que assumiu-se que os dados na Web Semântica são distribuídos, e que a *Web* é um grande espaço de informações não supervisionado, meios para localização inteligente de dados baseados em *RDF* e focados em ontologias são requeridos. Baseado em uma descrição semântica do objetivo de procura, o sistema deve poder descobrir informação pertinente sobre a *Web*.
- **Internacionalização:** O Projeto *DWeb* deve permitir a todos os usuários criar ontologias e suas instâncias em diferentes linguagens, suportando conjuntos de caracteres latinos e não latinos.
- **Consistência:** consistência de informações é um requisito fundamental em qualquer aplicação. Para o *DWeb*, cada alteração de uma ontologia deve resultar em uma ontologia também consistente. Para atingir este requisito, deve-se definir regras de evolução de ontologias. Os módulos de alteração de ontologias devem ser implementados aderentes a estas regras. Também devem ser feitas todas as atualizações de ontologias dentro do conceito de transações que assegurem as propriedades comuns de atomicidade, consistência, isolamento e durabilidade (*ACID*).
- **Semântica Formal:** uma semântica formal especificada para uma ontologia deve ser clara e não ambígua.
- **Durabilidade:** semelhante à consistência, a durabilidade é uma das exigências que devem estar contidas em qualquer aplicação com uso intenso de dados. Ela pode ser conseguida reutilizando-se a tecnologia desenvolvida para bancos de dados.
- **Concorrência:** deve ser possível o acesso concorrente e a modificação de dados. Isto pode ser conseguido usando-se processamento transacional, onde objetos podem ser modificados no decorrer de uma transação.
- **Raciocínio:** o motor (*engine*) de raciocínio é o componente central de aplicações Web Semântica e pode ser utilizado para várias tarefas, tais como, validação semântica e dedução de informação implícita. O Projeto *DWeb* deve prover acesso a tais motores que podem realizar o serviço de raciocínio requerido.

5. Flexibilidade

- **Estensibilidade:** a estensibilidade é um requisito que se aplica à maioria dos sistemas de software. Princípios de engenharia de software criam funções adicionais que tentam evitar mudanças de sistema necessárias no futuro. Conseqüentemente, a estensibilidade também é desejável para o Projeto *DWeb*. Além disso, este projeto deve conter múltiplas camadas e dados para modelar a Web Semântica. Assim, são necessários múltiplos módulos de software, tais como, analisadores *XML*, armazenamento *RDF*, ferramentas de mapeamento de bancos de dados relacionais e ontologias, armazenamento de ontologias e módulo de raciocínio *OWL*. Desta forma, a estensibilidade considera novas *APIs* de dados e módulos de software correspondentes que são uma exigência importante para tal sistema. Em primeiro lugar, dados de várias *APIs* orientarão, no futuro, suporte a cada camada e novos módulos de

software para serem integrados. Segundo, uma aplicação particular pode ter comportamento proprietário. Isto requer que sejam usadas novas soluções para poder estender as características.

- Supervisão de Módulos de Software: para um cliente, deve haver uma possibilidade de declarar o estado de seu trabalho precisamente. Conseqüentemente, são requeridos meios para supervisionar inteligentemente os módulos de softwares. Baseado em uma descrição semântica do objetivo de procura, o sistema deve poder descobrir em qual cliente está o seu foco em um determinado momento.
- Dependência: o sistema deve permitir expressar dependências entre módulos de software diferentes. Um exemplo é a administração de dependências que indique que: o módulo **A** é dependente do módulo **B**.

2.1.3 Principais características do Projeto DWeb

O Projeto *DWeb*⁶, (acrônimo *Dream Web*, ou seja a “*Web* dos sonhos”), tem como meta criar um meio universal para a troca de dados, informações e conhecimentos. Para tanto, o usuário deve ser contemplado com uma visão transparente, acessível[14] e intuitiva com relação aos recursos de informação providos e manipulados pelo seu ambiente.

O *DWeb* visa contribuir com a implantação, desenvolvimento e consolidação da Web Semântica. Neste contexto propõe um ambiente de manipulação semântica de recursos de informação baseado em comunidades virtuais de prática e tecnologias da Web Semântica.

A segmentação informacional em comunidades permite a estruturação dos recursos de informação em domínios de conhecimentos específicos[25]. Desta forma, torna-se possível a manipulação semântica dos recursos de informação em um contexto preciso, onde a definição de ontologias de domínio se torna coerente.

O *DWeb* propõe um ponto único de acesso aos sistemas computacionais sobre sua infraestrutura. A restrição em um ponto único de acesso é uma forma de minimizar os problemas relacionados à falta de integração do sistemas computacionais nas organizações. Pode-se citar como exemplo uma organização fictícia que tenha 300 sistemas computacionais em atividades com 1000 funcionários ativos. Se esta organização não tiver preocupação com o aspecto da integração dos seus sistemas computacionais, uma anomalia que provavelmente ocorrerá será, a replicação (parcial ou total) de lista de funcionários (300 x 1000), bem como das funções de controle de acesso (*login* de acesso) relativos aos sistemas. Outro ponto importante que pode ser observado é a gestão dos funcionários com relação aos 300 sistemas e os 1000 funcionários. Além disso, outro ponto relevante diz respeito à manutenção de uma estrutura de *menus*, os quais, freqüentemente estão atrelados a controles de acesso. Desta forma, a integração dos sistemas em um ponto único de acesso se mostra de extrema importância em relação a estes aspectos[24]. O *DWeb*, propõe uma arquitetura, onde existe um ponto único de acesso e um ambiente totalmente integrado pela sua infra-estrutura. Nesta arquitetura, a integração se dá em nível de *middleware* e de sistemas.

A integração em nível de *middleware* se dá pela fato do *DWeb* ser construído em uma arquitetura baseada em mensagens síncronas e assíncronas. Neste modelo, os serviços são fracamente acoplados o que permite a inserção e retirada dos mesmos de forma facilitada e permite melhoramentos no tratamento de ameaças à segurança do ambiente.

A integração em nível de sistemas se dá, principalmente, por componente de *Workflow*, que tem a função de integrar os processos de negócio e automatizar a criação de menus com

⁶Referenciado a partir deste momento, apenas por *DWeb*.

opções, que promove a característica de personalização. A base do controle de acesso é o *RBAC* (*Role Base Access Control*) o qual é baseado em papéis, e também, orientado a comunidades virtuais de prática.

Outro ponto importante da arquitetura do *DWeb* é o fato dela ser orientada a serviços. Estes serviços estruturam a arquitetura permitindo alto nível de escalabilidade, pois os serviços (teoricamente) podem crescer indefinidamente e sofrer alterações de forma bastante transparente ao desenvolvedor de aplicações.

Sobre esta infra-estrutura cria-se a possibilidade de contemplar os principais requisitos da Web Semântica: **inteligência**, **interoperabilidade** e **integração**, detalhados a seguir.

2.2 A Inteligência

A inteligência é um dos requisitos mais complexos do *DWeb*, e que exige um maior empenho de pesquisa científica, principalmente, no sentido da representação e manipulação de conhecimentos. A finalidade principal é atender solicitações de serviços e retornar, aos solicitantes, os resultados baseados em tratamento inteligente de recursos de informação. De acordo com a enciclopédia britânica, citada por Sabbatine[44], a inteligência pode ser definida como: “habilidade de se adaptar efetivamente ao ambiente, seja fazendo uma mudança em nós mesmos ou mudando o ambiente ou achando um novo ambiente”.

Segundo Sabbatine, “...esta é uma definição inteligente, porque ela incorpora aprendizado (uma mudança em nós mesmos), manufatura e abrigo (mudança do ambiente) e migração (encontrando um novo ambiente). A inteligência é uma entidade multifatorial, envolvendo coisas tais como a linguagem, pensamento, memória, raciocínio, consciência (a percepção de si mesmo), capacidade para aprendizagem e integração de várias modalidades sensoriais. De modo a nos adaptar efetivamente, o cérebro deve usar todas estas funções...”. Continua Sabbatine, “...inteligência não é um processo mental único, mas sim uma combinação de muitos processos mentais dirigidos à adaptação efetiva do ambiente...”.

Reconhecer quais são os componentes da inteligência é muito importante em termos de montar uma “teoria da inteligência” e também para desenvolver uma aplicação computacional que se utilize dela. Ainda segundo Sabbatine, uma das teorias mais sólidas e interessantes foi proposta por *Sternberg*, e se relaciona diretamente ao que se sabe sobre a evolução. Ele propõe que a inteligência, conforme pode ser visto na Figura 2, é constituída de três aspectos integrados e interdependentes: mundo interno, as relações com o mundo externo, as experiências que relacionam o mundo externo e o interno.

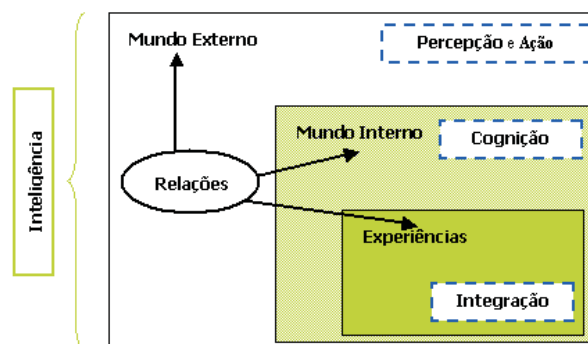


Figura 2: Componentes da Inteligência, segundo *Sternberg*[44].

O *DWeb*, mais modesto em suas pretensões, trabalha a Inteligência Computacional como

um conjunto de processos computacionais formado por raciocínio computacional automatizado (mecanismo de inferência), conhecimento e a memória. Estes elementos estão detalhados a seguir.

2.2.1 Raciocínio Computacional Automatizado

O conhecimento, é manipulado através de mecanismos de inferência[43], ou seja raciocínio computacional automatizado, que atuam sobre os recursos de informação através de tecnologias de metadados e ontologias. A arquitetura *DWeb* é gerida pela tecnologia de agentes racionais, ou seja agentes inteligentes, que são organizados de acordo com as suas funções e/ou objetivos a serem atingidos.

Um agente inteligente é um software que pode ter as seguintes características: autonomia, comunicabilidade, comportamento adaptativo, confiabilidade, cooperatividade, comunicação, flexibilidade, inteligência, mobilidade, persistência, personalização, orientação a objetivos, representatividade, responsabilidade e sociabilidade[49]. De forma geral, o agente capta o ambiente e a ele reage de acordo com suas próprias regras internas de operações. Diferentemente de aplicações convencionais, o próprio agente determina, quando e como executará as suas funções.

Estes agentes inteligentes, irão interoperar em um ambiente de multi-agentes com diversas características, tais como: autonomia, mobilidade, comunicação, etc. Eles deverão ter a capacidade de “aprender” significados de termos a partir de especificações formais, e utilizar-se deles para tomadas de decisões inteligentes, aprimorando a interação com outros agentes que entrarem em comunicação com o ambiente e/ou com os clientes de aplicação *DWeb*.

De *Russell e Norvig*[43], pode-se simular a inteligência humana através da tecnologia de agentes racionais, como se pode ver na Figura 3:

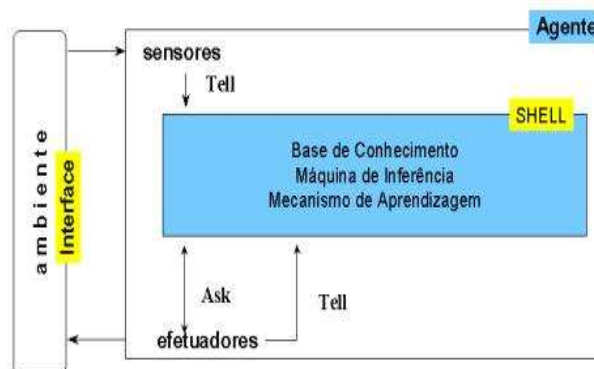


Figura 3: Componentes de Agentes Racionais, segundo *Russell e Norvig*[43].

Os agentes inteligentes (racionais), implementam um mecanismo que contempla (teoricamente) os requisitos fundamentais para a inteligência computacional. A sua utilização, integrada com metadados e ontologias, possibilita ações inteligentes relacionadas a recursos de informação no *DWeb* e utiliza estes mecanismos computacionais para promover a inteligência computacional no ambiente de comunidades *DWeb*.

A inteligência necessita de um componente adicional para aplicar o raciocínio computacional, o Conhecimento.

2.2.2 O Conhecimento

O conhecimento é representado por metadados, ontologias e recursos de informação. Os metadados são mecanismos de descrição de recursos de informação. As ontologias são representações de um domínio, utilizando conceitos e relacionamentos entre eles. Seu uso facilita o entendimento, via vocabulário adicional, através do qual uma semântica formal é definida para representar os conhecimentos do *DWeb*[43]. Os recursos de informação são informações e/ou conhecimentos, em meio eletrônico ou não, disponibilizados pelas comunidades virtuais de prática e referenciados por metadados.

Tom Gruber[29], refere-se à ontologia como uma especificação explícita e formal de uma conceitualização compartilhada, onde, o termo especificação explícita, simboliza as regras do mundo conceitual (conceitos, propriedades, funções, axiomas definidos explicitamente). O termo formal, refere-se ao processo de inferência aplicável sobre esta semântica formal, ou seja, manipulável por máquina. O termo conceitualização, refere-se aos fatos do mundo real, representados via um modelo abstrato de algum fenômeno do mundo e, finalmente, o termo compartilhada, refere-se ao conhecimento consensual entre grupos de pessoas. Essa semântica formal é implementada e materializada no ambiente do *DWeb* através da linguagem padrão *W3C* para construção de ontologias *OWL*[20].

A inteligência necessita de um local para guardar estes conhecimentos: a memória.

2.2.3 Memória

Como mecanismo de memória, o *DWeb* utiliza bases de dados distribuídas e heterogêneas. Estas bases de dados são compostas de uma arquitetura semelhante a um armazém de dados acessível via linguagem *XML*. Estes acessos são intermediados por um mecanismo de homogeneização de mediação/tradução de bases heterogêneas[48].

A mediação é um dos componentes do *DWeb*, que têm a finalidade de dar a transparência estrutural aos recursos de informação heterogêneas e distribuídas. Esta transparência se efetiva quando da disponibilização aos Clientes, Serviços e Módulos do *DWeb* de uma estrutura de dados virtual [48].

A tradução, é um serviço *DWeb* que se integra ao serviço de mediação com vistas a traduzir as consultas recebidas do mediador, em consultas às bases de dados reais. Este é um componente distribuído por toda a arquitetura do *DWeb*.

Além da inteligência, a Web Semântica tem outro macro requisito: a interoperabilidade, a qual é tratada a seguir.

2.3 A Interoperabilidade

A interoperabilidade, permite o compartilhamento de recursos de informação no contexto da Web Semântica. O *DWeb*, disponibiliza um mecanismo de interoperabilidade que trabalha especificamente no nível de aplicação. São contemplados os seguintes tipos de interoperabilidade: sintática, estrutural e semântica. A interoperabilidade sintática, é promovida por mecanismos *XML*, quando do transporte e representação de dados e metadados no ambiente *DWeb*.

A interoperabilidade estrutural, representa o modelo de dados distintos especificando como os recursos estão organizados, os tipos, valores possíveis por tipo e estão organizados em *RDF* e *RDFS* no *DWeb*.

A interoperabilidade semântica, proporciona a compreensão plena de conteúdo via mecanismos de inferência. A interoperabilidade semântica, ocorre na manipulação sobre fatos e regras registrados respectivamente em metadados, ontologias e recursos de informação. Estes

mecanismos são incorporados no *DWeb* e detalhados na arquitetura conceitual do mesmo, onde pode-se observar os componentes que permitem a sua implementação.

O último macro requisito da Web Semântica é a integração das fontes heterogêneas de informações, tema tratado a seguir.

2.4 A Integração

A integração, permite a troca de informações em ambiente heterogêneo e distribuído de forma transparente para os seus clientes. Esta troca de informações é implementada no *DWeb* através de mecanismo de mediação/tradução. Este mecanismo acrescenta uma camada adicional entre as aplicações e as fontes heterogêneas de informação [48].

As fontes heterogêneas de informação devem sofrer um processo de homogeneização, a ser proporcionado por este mecanismo, de forma a permitir integração transparente das mesmas e as aplicações solicitantes realizando traduções de consultas às bases de dados reais.

A seguir, detalha-se a unidade de informação do *DWeb*, que são os recursos de informação.

2.5 Recursos de Informação

O *DWeb*, é um ambiente computacional voltado para a manipulação de recursos de informação disponibilizados por comunidades virtuais de prática[25]. Estes recursos de informação englobam documentações armazenadas em meio computacional ou não. Todos os documentos (eletrônicos ou não) são manipulados através de mecanismos de metadados, que lhes dão atributos tais como: autor, sumário, data de publicação, etc.

Os recursos de informação manipulados pelo *DWeb*, são classificados, segundo a estruturação dos dados, em: dados estruturados, dados semi-estruturados e dados não estruturados. Os dados estruturados são dados armazenados em sistemas gerenciadores de banco de dados tradicionais. Os dados semi-estruturados são os dados armazenados em arquivos do tipo *HTML* (*Hypertext Markup Language*), *e-mail* e *news*. Neste tipo de informação é possível alguma extração de características ou atributos através de processamento automático. Os dados não estruturados são provenientes de arquivos em formatos proprietário, como por exemplo: *PDF* (*Portable Document Format*), *PostScript* e *MS Word*[12].

O *DWeb* representa recursos de informação através de recursos de metadados, o que torna possível a implementação do modelo *mediador/tradutor* para recuperação de informações.

Os recursos de informação são manipulados em uma arquitetura flexível e genérica promovida por serviços e módulos *DWeb*, conforme apresentados a seguir.

2.6 Serviços Básicos e Módulos DWeb

O *DWeb* trabalha com o conceito de arquitetura evolutiva, ou seja, uma arquitetura flexível e que permite a inserção, alteração e remoção de funcionalidades de forma transparente aos desenvolvedores *DWeb*. Para implementar este conceito, o *DWeb* trabalha como uma arquitetura de serviços e módulos.

Os Serviços Básicos são componentes de *software* que executam diversas funcionalidades a serem compartilhadas pelos Módulos *DWeb*.

Os Módulos *DWeb* são os diversos sistemas computacionais, que dão um caráter generalista à arquitetura *DWeb*, que compartilham os Serviços Básicos de forma a aumentar o nível de reutilização do ambiente.

A seguir, é apresentado o conjunto de princípios básicos que norteia o desenvolvimento do *DWeb* de forma a atingir seu objetivo geral.

2.7 Princípios Básicos do Projeto

O processo de construção do *DWeb* parte de um conjunto de princípios básicos. Estes princípios básicos norteiam o processo de desenvolvimento do *DWeb* e mantém a coerência com o objetivo geral de Projeto. A seguir, relaciona-se estes princípios básicos.

- Explorar a representação formal do domínio do discurso⁷ para descrever recursos de informação nos termos de um modelo conceitual de comunidade virtual de prática;
- Desenvolver técnicas de representação de informações e/ou conhecimentos desenvolvidos por comunidades e que possibilite a reutilização destes;
- Desenvolver técnicas que permitam obter transparência de acesso e manipulação em relação aos dados, informações e conhecimentos a todas as comunidades;
- Criar um ambiente que permita realizar desenvolvimento evolutivo e flexível de aplicações para Web Semântica;
- Propiciar um ambiente orientado por padrões tecnológicos abertos;
- Propiciar um ambiente de pesquisa científica e tecnológica interdisciplinar, aderente aos padrões da comunidade científica nacional e internacional, com foco nos avanços tecnológicos e bem estar do ser humano;
- Propiciar um ambiente seguro, com vista a atingir as metas de integridade, disponibilidade e privacidade sobre os recursos de informação de comunidade[37];
- Propiciar a gestão de conhecimento orientada a comunidades virtuais de prática[25];
- Criar um ambiente generalista, que possibilite a coexistência de sistemas computacionais distintos.

De forma geral, o objetivo geral do Projeto *DWeb* é *Criar um ambiente de desenvolvimento de aplicações genéricas, dentro dos princípios da Web Semântica, que permita agregar inteligência, interoperação e integração ao ambiente da Web existente, a partir de um modelo conceitual de comunidades virtuais de prática.*

A seguir, são detalhados os conceitos adotados pelo *DWeb* para Comunidades Virtuais de Prática e as vantagens decorrentes de sua utilização como fundamentação do modelo arquitetônico proposto pelo projeto.

⁷Deve-se entender o domínio do discurso, como a formalização de várias categorias conceituais (ontologias e/ou *thesaurus*) que estão relacionadas às atividades e/ou entidades, as quais estão desenvolvendo processos diários de trabalho (interesses ou práticas comuns). Estas categorias são organizadas em hierarquias de conceitos e relações. Os conceitos e relações têm nomes e, estes nomes, constituem terminologias que são o instrumental para o intercâmbio de informações e conhecimentos e representam o processo cooperativo de comunidades[12], [42].

3 Comunidades DWeb

Segundo Teixeira[25], o advento da Internet como meio de comunicação ágil, flexível e de baixo custo, foi o fato propulsor da adoção, em larga escala, das comunidades virtuais. Ele destaca ainda, que estes grupos foram organizados utilizando-se das tecnologias de *e-mail*, *chats* e *websites* para se comunicarem, onde profissionais de áreas específicas passaram a trocar informações relevantes sobre o seu dia-a-dia, ou seja, as suas **melhores práticas** (tais como, experiências, histórias, ferramentas) e as formas como estruturam seus processos, além de compartilhar soluções para seus problemas mais comuns[35].

As comunidades segmentam os recursos de informação e conhecimentos em termos de interesses comuns que elas possuem, o que, desta forma, permite a segmentação destes em domínios de conhecimentos específicos. Ao restringir as comunidades em domínios específicos, cria-se a possibilidade de utilização de tecnologias da Web Semântica, dentre elas as ontologias de domínios [9], [32].

Esta segmentação conceitual proporciona um modelo bastante adequado para o desenvolvimento da Web Semântica, pelo fato de se aproximar muito, em termos funcionais, do modelo de “pseudo-organização” da *Web* existente e permitir aprimoramento desta, no que tange aos seus requisitos principais de inteligência, interoperabilidade e integração.

O *DWeb*, atua no contexto destas comunidades, possibilitando a manipulação e compartilhamento de conhecimento de forma transparente. Estes conhecimentos são oriundos de uma semântica formal representada em ontologias, metadados e recursos de informação. Os recursos de informação são informações ou conhecimentos, em meio eletrônico ou não, disponibilizados pelas comunidades virtuais de prática, referenciados por metadados e representam unidades de informações. As ontologias organizam conceitos de forma hierarquizada compondo um conjunto de regras para avaliação semântica de fatos. Os metadados referenciam recursos de informação (ex: documentos, dados, etc.) construídos pelas respectivas comunidades e representados em linguagem formal destacando diversos atributos sobre os mesmos.

Uma comunidade, para o *DWeb*, é definida em termos de conceitos (ex: pessoas, artefatos, eventos, etc.) e propriedades (ex: estilo, material, etc.) com interesses comuns (ex: comércio, educação, cultura, medicina, computação, etc.). Estas comunidades têm em comum alguma identificação e conhecimentos formalizados que são utilizados para criar recursos de informação (ex: documentos, dados, etc.) [12].

Os usuários de comunidades *DWeb* têm funcionalidades diferenciadas pelos seguintes papéis: gestor de ambiente, gestor de comunidades e usuários participantes de comunidades. O gestor de ambiente têm a incumbência de gerir todas as comunidades catalogadas no *DWeb*, dentre as suas funções pode-se citar: criação, alteração e exclusão de comunidades e atribuição do papéis de gestor de comunidades. O gestor de comunidades têm a incumbência de gerir comunidades específicas e alocar perfis de acesso aos usuários participantes de comunidades. Os usuários participantes de comunidades são usuários catalogados na comunidade aos quais são alocados papéis pelo gestor da comunidade. Eles participam ativamente no acesso e/ou manipulação de dados, informações e conhecimentos.

O modelo conceitual ou esquema conceitual descreve, integra e recupera recursos de informação de comunidades e captura sua estrutura semântica. Desta forma é possível construir aplicações para a Web Semântica que tenham foco na arquitetura de comunidades *DWeb*, ou seja, estruturar um ambiente da Web Semântica com múltiplas visões sobre uma mesma infraestrutura, de forma a contemplar uma arquitetura com múltiplos sistemas computacionais e com possibilidade de reutilização de funcionalidades, recursos de informação, bem como, de conhecimentos desenvolvidos pelas comunidades e entre comunidades [9].

4 Arquitetura Conceitual do Projeto DWeb

O *DWeb* é constituído de uma arquitetura conceitual centrada no triângulo de relacionamentos de recursos de informação desenvolvidos e compartilhados de forma transparente em comunidades e entre comunidades, onde são destacados os elementos: **membro de comunidade, gestão e comunidade**.

A Figura 4, representa uma visão dessa arquitetura conceitual em alto nível que dá origem ao modelo organizacional do *DWeb*. Nela, pode-se identificar as suas principais características, que são: **integração, interoperabilidade e inteligência**, as quais são aplicadas sobre os recursos de informação criados por **membro de comunidade** e **geridos** pelo *DWeb*. O *DWeb* oferece a estes **membros de comunidade** e às suas **comunidades** diversos tipos de facilidades de **transparência** (conforme detalhado na Seção 4.2) no que tange aos acessos e manipulação dos recursos de informação e conhecimentos oriundos de **comunidades**.

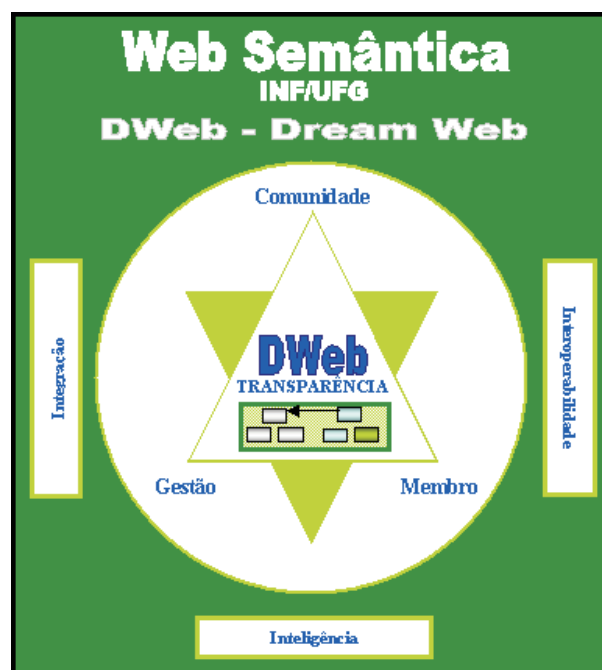


Figura 4: Visão da Arquitetura Conceitual do Projeto *DWeb* em Nível Macro.

Nesta relação informacional, é possível o compartilhamento de recursos de informação dentro de um contexto voltado para atender as necessidades de informações e conhecimentos de comunidades[25].

O elemento **membro**, representa a visão de cada participante de comunidade que é usuário de uma aplicação da Web Semântica que manipula recursos de informação e conhecimentos de uma comunidade *DWeb*.

O elemento **comunidade**, representa todos os grupos de comunidades com interesses comuns que estão compartilhando os recursos de informação e conhecimentos. Este compartilhamento, se dá entre membros da comunidade, representado por pessoas e entre comunidades distintas.

O elemento **gestão**, representa todos os mecanismos de infra-estrutura do *DWeb* que possibilitam atingir os principais objetivos da Web Semântica. Pode-se destacar os objetivos de agregação de conhecimentos, obtidos via mecanismo de inferência, acesso interoperável nos níveis sintático, estrutural e semântico e, finalmente, a integração de fontes de dados heterogêneas e distribuídas.

Neste contexto, o caráter generalista do *DWeb* é demonstrado em todos os níveis de sua arquitetura. Na arquitetura conceitual, o fato do *DWeb* ser voltado para construção de aplicações para Web Semântica e direcionado à comunidades pode ser evidenciado. As comunidades, contemplam tanto os aspectos informais como os aspectos formais de estruturas organizacionais que as contenham. Desta forma, os recursos de informação e conhecimentos são genéricos, ou seja, não são direcionados a nenhum tipo de aplicação específica.

A seguir, distribui-se os elementos: **membro**, **gestão** e **comunidade** uma arquitetura de três camadas conceituais.

4.1 Visão Arquitetônica Geral do Projeto *DWeb*

A arquitetura geral do *DWeb* suporta os requisitos principais da Web Semântica, que são interoperabilidade, integração e inteligência. Para que estes requisitos possam ser melhor implementados, desenvolvidos e consolidados particiona-se a arquitetura em três camadas conceituais. Os três elementos (membro, gestão e comunidades) podem ser mapeados sobre estas camadas, conforme se pode ver na Figura 5.

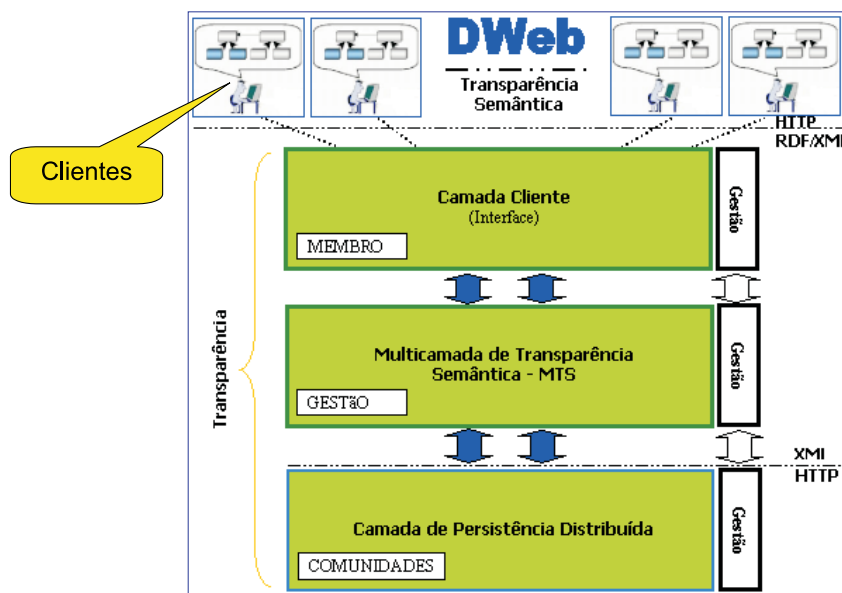


Figura 5: Arquitetura Conceitual do Projeto *DWeb* com Alocação de seus Elementos.

Desta forma, concentra-se em cada camada conceitual funcionalidades que lhes são próprias de forma a possibilitar um funcionamento integrado com as demais camadas. Além disso, isto possibilita a implementação de uma arquitetura de componentes que interagem na forma de troca de mensagens com baixo acoplamento entre os mesmos.

Com estas características o *DWeb* contempla diversos tipos de transparências nos acessos e manipulação de recursos de informação, conforme pode ser visto a seguir.

4.2 Transparência do *DWeb*

O ponto central da arquitetura conceitual do ambiente *DWeb* é o seu alto nível de transparência. A transparência modela os diversos tipos de facilidades disponibilizadas às comunidades. Para fornecer esses tipos diferentes de transparência o *DWeb* é organizado em uma arquitetura de serviços, que são componentes computacionais aderentes à arquitetura (maiores

detalhes podem ser vistos na Subseção 4.8.5). Estes serviços aumentam a flexibilidade do ambiente, graças à sua arquitetura de *plug-ins*. A arquitetura de *plug-ins*, refere-se a um mecanismo que permite incluir, alterar, retirar módulos e serviços de forma transparente ao desenvolvedor de aplicações *DWeb* e, desta forma, possibilita a característica evolutiva do *DWeb*. A seguir são mostradas as relações entre serviços e as respectivas capacidades em termos de transparências que estas proporcionam:

- o serviço de *middleware*: proporciona transparência de comunicação interaplicação e serviços de mensagens;
- o serviço de interface com membro de comunidade: proporciona transparência aos diversos dispositivos de acesso (computação ubíqua);
- o serviço de integração de fontes de dados: proporciona a transparência estrutural de fontes heterogêneas de dados através do modelo mediador-tradutor;
- o serviço de raciocínio automatizado: proporciona a transparência de processos de inferência aplicados aos recursos de informação compartilhados por comunidades;
- o serviço de gerenciamento de ambiente: proporciona a transparência de processos gerenciais do *DWeb* e transparência de segurança computacional[37];
- o serviço de interconexão de multicamada de transparência semântica: proporciona a transparência semântica aplicada aos recursos de informação e a sua escalabilidade;
- o serviço de personalização e rastreamento de membro de comunidade: proporciona a transparência de personalização e fluxo de trabalho, onde autorizações de acesso a recursos de informação, serviços e módulos são automaticamente mapeados para as pessoas;
- o serviço de comunidades: proporciona a transparência de recursos de informação, são serviços voltados especificamente para as comunidades;
- o serviço de módulos (sistemas computacionais): proporciona a transparência de integração aos diversos sistemas computacionais do *DWeb*;
- o serviço de persistência de recursos de informação: proporciona a transparência de localização de fontes heterogêneas de informações distribuídas;
- o serviço de *plug-ins*: proporciona a transparência de funcionalidades de aplicações e evolução no contexto do *DWeb*;
- o serviço de inteligência: proporciona a transparência de manipulação de ontologias aos diversos sistemas computacionais do ambiente;
- o serviço de busca de recursos de informação: proporciona a transparência na busca semântica de recursos de informação distribuídos pelo ambiente.

Estas facilidades estão disponíveis em um ponto único de acesso.

4.3 Ponto Único de Acesso

Nesta arquitetura, o ponto único de acesso aos recursos de informação é uma das características fundamentais. Esta concepção permite a construção de um modelo no qual monta-se uma arquitetura de *datawarehouse* virtual em XML, onde se disponibiliza acesso transparente aos recursos de informação em um ambiente heterogêneo e distribuído[12].

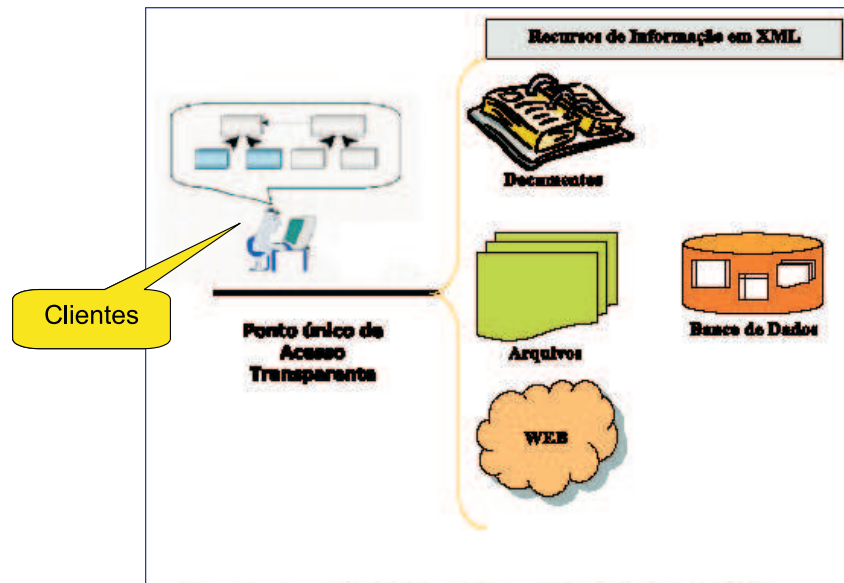


Figura 6: Visão do ponto único de acesso a recursos de informação.

Apesar deste modelo aparentemente restringir o acesso, a idéia é exatamente o oposto. O que se pretende é tornar disponível os acessos aos dados, informações e conhecimentos em um único ponto. Desta forma, o usuário não necessita procurar exaustivamente por funcionalidades (serviços e/ou módulos) e recursos informação, pois estes são mapeadas automaticamente pelo componente de *Workflow*, ou seja, os acessos são personalizados [50], [30], [24].

Sobre este ponto único de acesso ocorre a manipulação de informação e conhecimentos.

4.4 Contexto do Conhecimento

Na era do conhecimento, os níveis de capacitação dos indivíduos e das organizações são fatores fundamentais para a sua sobrevivência. A tecnologia permite que informação e conhecimento sejam permutados com facilidade, aproximando pessoas, comunidades e nações. A complexidade do processo se deve ao fato de que o valor não está mais no domínio da informação, mas sim em como trabalhar com o conhecimento relacionado a esta informação.

Sistemas Inteligentes são exemplos de sistemas que utilizam a tecnologia da informação para manipular conhecimentos especializados e que apresentam resultados qualitativos e quantitativos. Esses sistemas permitem que um maior número de pessoas obtenha acesso ao conhecimento a partir de sua aquisição, sistematização, representação e processamento em ambientes computacionais.

Pode-se usar a aquisição de conhecimento como mecanismo de gestão do conhecimento organizacional (explícito e tácito). O conhecimento explícito é o conhecimento de fácil aquisição e pode ser encontrado em livros, manuais, etc. O conhecimento tácito está na mente das pessoas e é de difícil aquisição. O DWeb cria estas facilidades para explorar a aquisição do conhecimento tácito através do motor de *Workflow*[30], [24] (Subseção 4.7.6).

Os Sistemas Inteligentes podem manipular símbolos que representam entidades do mundo real, eles são capazes de trabalhar eficazmente com o conhecimento, através de mecanismos de inferência. Para se consolidar os conhecimentos acerca destes mecanismos é necessário distinguir dado, informação e conhecimento.

4.5 Dado, Informação e Conhecimento

A distinção entre dado, informação e conhecimento é necessária para o entendimento da natureza e dos tipos de conhecimento humano e, desta forma, estabelecer uma ligação entre conhecimento e as tecnologias que os manipulam [40].

Pode-se considerar que o dado é um elemento puro sobre um determinado evento. Normalmente são utilizados no ambiente operacional, podendo ser armazenados e recuperados de Banco de Dados e/ou através de alguma forma documental. O dado necessita de complementos para oferecer embasamento ao entendimento da situação.

A informação é o dado analisado e contextualizado. Ela envolve a interpretação de um conjunto de dados. Desta forma, faz-se necessária a inserção de parâmetros de comparação. Uma análise somente é possível após o estabelecimento do contexto de referência, gerando-se a informação.

O conhecimento se refere à habilidade de se criar um modelo mental que descreva o objeto e também as ações que se deve implementar para tomadas de decisões inteligentes. Em poucas palavras, pode-se afirmar que o processo de gerar conhecimento é um processo no qual uma informação é comparada à outra e combinada em muitas ligações úteis e com significado semântico. Isto implica que o conhecimento é diretamente dependente da experiência das pessoas que as utilizam. O conhecimento pode ser abstraído em diversos níveis como, por exemplo:

- Fatos: são as relações entre objetos;
- Conceitos: são as abstrações de idéias de natureza hierárquica;
- Regras: são as operações que orientam ações a serem tomadas;
- Metarregras: são responsáveis pela criação de novas regras.

No processo da tomada de uma decisão acontece o uso explícito de um conhecimento, ou seja, o conhecimento é usado como base de avaliação para o referido processo. O conhecimento pode ser representado na forma de uma combinação de estruturas de dados e procedimentos interpretativos que levam a um comportamento já conhecido, fornecendo informações a um sistema computacional, que terá a função então de planejar e tomar as decisões necessárias na execução de ações. Desta forma pode-se classificar o conhecimento da seguinte forma:

- Declarativo: **o que é** - descreve os fatos e eventos;
- Procedural: **como funciona** - prescritivo e difícil de explicar;
- Senso Comum: **juízo certo e errado** - declarativo e procedural;
- Item Heurístico: **único para cada indivíduo** - envolve avaliação sistemática.

O conhecimento necessário à solução de um problema pode ser oriundo de várias fontes de informação. Desta forma, uma decisão pode ser tomada por meio da análise lógica, que foi feita baseada em dados, entrevistas, relatórios ou outros meios e, eventualmente, pode estar apoiada em dados heurísticos ou intuitivos.

A seguir, detalha-se a arquitetura conceitual do *DWeb* em suas três camadas funcionais básicas.

4.6 Detalhamento da Visão de Alto Nível do Projeto DWeb.

O *DWeb* pode ser melhor compreendido pelo detalhamento de sua arquitetura em uma visão de alto nível, conforme pode ser visto na Figura 7. Por facilidades didáticas e organizacionais a arquitetura conceitual do *DWeb* divide-se em três camadas, quais sejam: **cliente**, **multicamada de transparência semântica**⁸ e a camada de **persistência distribuída**.

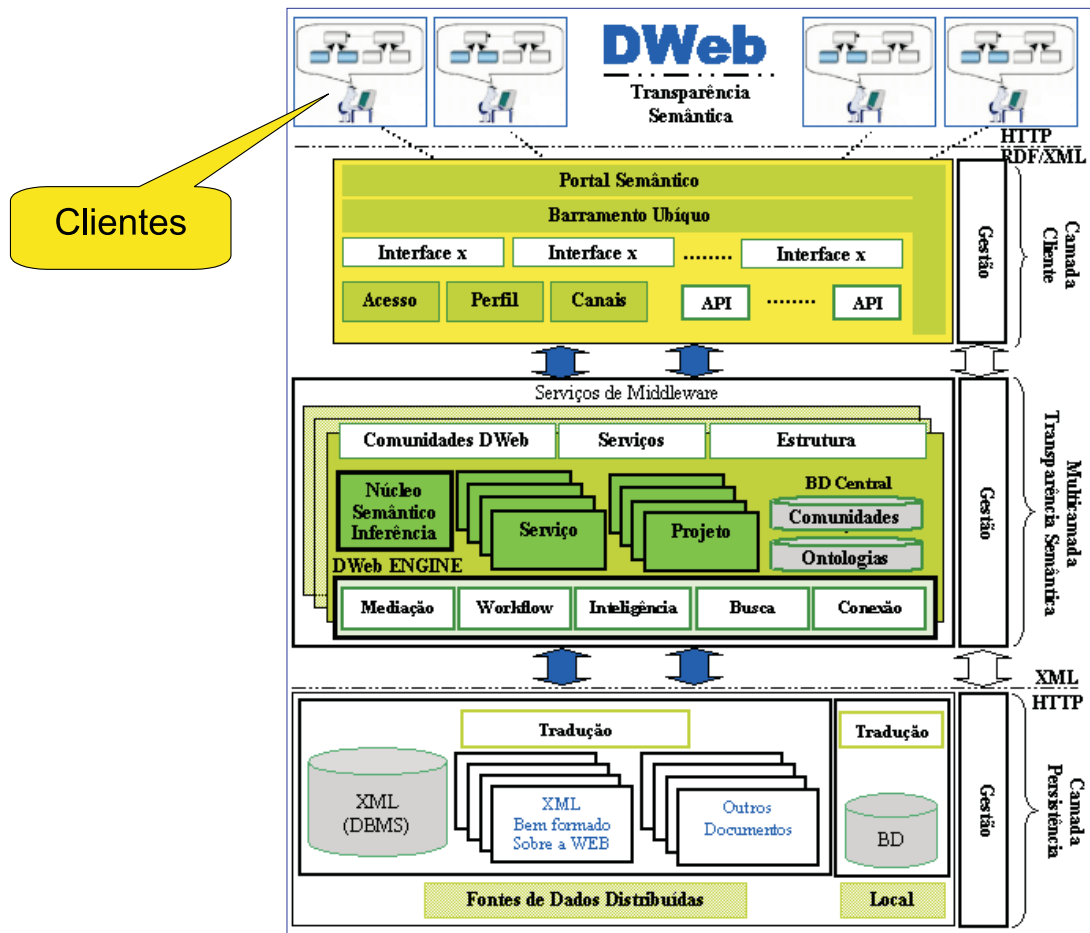


Figura 7: Arquitetura Conceitual do Projeto *DWeb* para Web Semântica.

A seguir, detalha-se cada uma destas três camadas, descrevendo-se os seus principais elementos com suas respectivas funcionalidades no contexto da arquitetura *DWeb*.

4.7 Camada Cliente

Esta camada (Figura 8) tem a responsabilidade de gerir todos os mecanismos que tratam da comunicação entre **Clientes** e o *DWeb*. Estes mecanismos criam uma interface mais amigável com o **Cliente**, sempre se preocupando com os aspectos de acessibilidade na Internet ou acessibilidade na *Web*, o que significa permitir o acesso à *Web* por todos, independente de tipo de usuário, situação ou ferramenta. A acessibilidade na *Web* beneficia também pessoas idosas, usuários de navegadores alternativos, acesso móvel e usuários de tecnologias assistivas (tecnologias que permitam acessos aos recursos de informação por pessoas com algumas dificuldades, tais como, dificuldades motoras, visuais, etc.[14]).

⁸A partir deste momento denominada pelo acrônimo *MTS*.

A camada **Cliente** deve dar suporte ao desenvolvimento de conteúdo das páginas e de ferramentas preparados para a acessibilidade. Os autores de conteúdo de páginas *Web* e ferramentas devem ter em mente a diversidade do público. Muitos usuários podem ter dificuldades (total ou parcial) de visão, audição, locomoção ou para processar facilmente algum tipo de informação. Outros podem utilizar-se apenas do teclado para navegar. Muitos podem, ainda, utilizar-se de navegador textual. A criação de páginas *Web* com *design* acessível beneficia pessoas com deficiências, mas também, pessoas sem deficiência em situações e características diversas.

Além destes aspectos esta camada deve se preocupar com detalhes relacionados à ubiquidade[14], utilidade, segurança da aplicação[37], etc..

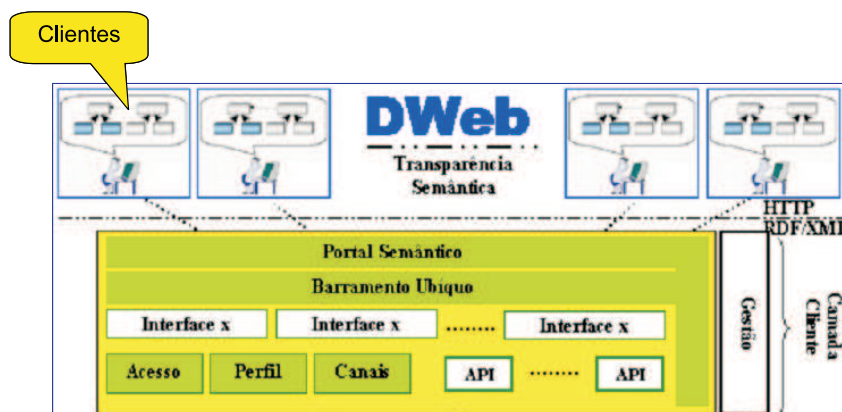


Figura 8: Camada Cliente (Interface) da Arquitetura do Projeto *DWeb*.

4.7.1 O Cliente

No topo da arquitetura, estão simbolizados os **Clientes** da aplicação *DWeb*. Estes **Clientes** devem se vincular a uma ou mais comunidades, para poder obter acesso aos serviços, recursos de informação e conhecimentos desta(s) comunidade(s).

Para obter acesso à comunidade, o **Cliente**, tem que ter a chave de ingresso na comunidade. Esta chave de ingresso na comunidade, é fornecida pelo gestor de comunidade.

Estes **Clientes** são diferenciados pelo mecanismo de gerenciamento baseado em papéis *RBAC*[50] e *Workflow Engine*[30]. O *RBAC* (*Role Base Access Control*) funciona como um serviço básico do *DWeb* que realiza gestão de acesso através de papéis atribuídos ao **Cliente** em cada comunidade à qual o mesmo está vinculado. O *Workflow Engine* é um serviço básico do *DWeb* com a responsabilidade de gerenciar o fluxo de trabalho orientado a comunidades [24].

Desta forma, para obter acesso diferenciado, os **Clientes** devem ser inscritos em um ou mais papéis nas comunidades às quais obteve ingresso. Estes papéis, são gerenciados pelo serviço de gerenciamento de fluxo de trabalho, que personaliza as opções de serviços *DWeb* disponíveis ao **Cliente**.

4.7.2 Portal Semântico Orientado a Comunidades *DWeb*

Os acessos serão disponibilizados por meio de navegadores *Web* (*browsers Web*), em portais semânticos. Estes navegadores devem estar configurados para operarem com o protocolo *IIOP* (*Internet Inter-ORB Protocol*). Este protocolo pode funcionar em túneis no protocolo *TCP/IP* (*Transmission Control Protocol/Internet Protocol*) sobre o *HTTP*[15] (*Hypertext Transfer Protocol*) [28], [24].

Através dele, torna-se possível mecanismos seguros no transporte de recursos de informação e a sua escolha é devida ao fato do mesmo ser um padrão no serviço de *middleware CORBA* [28] (*Common Object Request Brucker Architecture*) e se adequar à *MTS*, onde se localizam os principais serviços fornecidos pelo *DWeb*.

Um portal semântico *DWeb*, permite categorizar recursos de informação ao longo de um esquema multidimensional de comunidades *DWeb*, que pode ser enriquecido a todo momento pelos usuários *DWeb*. Também fornece a possibilidade para ingressar em serviços relativos a todas as comunidades de que fazem parte. Os recursos podem ser categorizados por técnicas de aprendizado de máquina, isto é: descrição de recursos de informação por seus produtores usando um sistema descritivo com base em metadados[24], [27].

No momento do primeiro acesso, o **Cliente** visualizará as informações básicas no Portal Padrão do *DWeb*, onde se encontram disponíveis uma série de serviços básicos *DWeb* aos **Clientes**. Dentre estes serviços, o **Cliente**, de posse da chave de ingresso de cada comunidade, pode se inscrever e, a partir de então, usufruir de toda a arquitetura do *DWeb*. Caso o **Cliente** não tenha as chaves, existe uma opção para que o mesmo faça solicitação ao gestor de comunidade que deseja fazer parte. Para isto, o **Cliente** deverá fornecer uma série de informações que serão analisadas para emissão da chave de ingresso específico do **Cliente** para cada comunidade.

4.7.3 Barramento Ubíquo

A arquitetura é provida por uma camada de adaptação automática ao dispositivo que o **Cliente** está utilizando (*PDA*, Telefones Celulares, *Desktop*, etc.). Esta camada tem o nome de “Barramento Ubíquo” e faz o rastreamento do **Cliente** montando uma interface “mais adequada” ao seu dispositivo de acesso. Ela é construída a partir de informações registradas na Base de Dados Central do *DWeb*, onde se encontram informações relativas aos dispositivos virtuais vinculados aos **Clientes** e, desta forma, a camada de “Barramento Ubíquo” faz o tratamento de adaptação e montagem da interface a cada operação de conexão com o ambiente[24], [14], [27].

4.7.4 Interfaces Comunitárias

São disponibilizadas, quando da definição e construção das comunidades, uma interface com a qual os **Clientes** terão acesso. Esta interface é específica para cada comunidade e é montada através de parâmetros de configuração de interface. O registro da interface fica vinculado à comunidade e às informações de definição da mesma no momento de sua criação pelo gestor de comunidade[27].

O Gestor de Comunidade, pode criar e gerenciar uma comunidade mediante chave de autorização de comunidade. Esta autorização é concedida e/ou revogada ao gestor de comunidade pelo Gestor do Ambiente do *DWeb*.

4.7.5 API's de Comunicação

A camada cliente disponibiliza serviços autoconfiguráveis por comunidades. Estes serviços utilizam-se de *API's* (*Application Programming Interface*) para acessar as diversas funcionalidades (Módulos) e serviços do ambiente *DWeb*.

Cada serviço *DWeb*, é acessível através de uma *API*, que deve ser documentada e divulgada para que as equipes de análise, projeto e codificação de módulos ou serviços as utilizem. Desta forma, os critérios de utilização, parametrização, funcionalidades, etc., são divulgados de forma pública e que facilite a reutilização dos serviços por todas as comunidades *DWeb*.

4.7.6 Controle de Acesso

Conforme especificado pelo W3C[23], o controle de acesso deve prover: gerenciamento de identificação, associação de grupos (identidade e propriedades) e gerenciamento de privilégios[24].

O serviço de controle de acesso geral do *DWeb* é disponibilizado pelo *Workflow Engine*[24] e é direcionado ao conceito de comunidades virtuais de prática. Todo o gerenciamento de acesso é centrado no domínio informacional de cada comunidade.

Da mesma forma, este serviço é integrado ao serviço de perfil de usuário, o qual tem a função de adequação a tecnologias assistivas e preferências de usuários. Este conceito é implementado através de agente de usuário, o qual é definido pelo W3C como um *software* para ter acesso a conteúdo *Web*, incluindo navegadores (*browsers*) gráficos, navegadores textuais, navegadores de voz, telefones móveis, jogos *multimedia* e alguns *softwares* que usam tecnologias assistivas junto com navegadores, tais como, leitores de tela, ampliadores de tela e software de reconhecimento de voz[14].

Um outro ponto da personalização do usuário são os canais pelos quais os recursos de informação trafegam, o que garante a característica ubíqua do ambiente e funciona de forma integrada ao perfil e controle de acesso[14].

4.8 Multicamada de Transparência Semântica

A camada *MTS* (como se pode ver na Figura 9) funciona dentro de um domínio de sistema de negócio *BSD* (*Business System Domain*) em *middleware CORBA*. Cada *BSD* contém uma ou mais camadas da transparência semântica. Estas camadas devem conter todos os componentes que permitem a execução de serviços e módulos. Dentre estes componentes destacam-se os Serviços de *Middleware*, as Comunidades *DWeb*, os Serviços de Controle, o Núcleo Semântico de Inferência, os Serviços Básicos, os Projetos, a Base de Dados Central e os *Engines DWeb*.

O *DWeb* propõe um mecanismo que permite a interligação entre camadas (componente de conexão) de transparência semântica e, desta maneira, formando um *cluster*, o qual foi denominado *MTS*. A *MTS* replica e interliga as camadas de transparência semântica (*CTS*) com a finalidade de se tornar um ambiente mais escalonável na realização de trocas de mensagens interaplicação na forma de mensagens síncronas, através do protocolo *IIOP* e, assíncrona, através da *API Java/JMS* (*Java Message Service*)[28], [24].

Este mecanismo de comunicação, através de mensagens, possibilita um melhor gerenciamento do aspecto segurança, pois aumenta a independência das aplicações[37], e torna o ambiente mais flexível, o que é fundamental para a arquitetura de *plug-ins*[24].

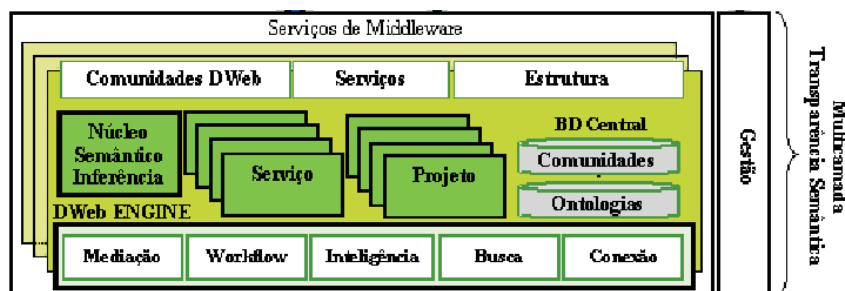


Figura 9: Multicamada de Transparência Semântica do Projeto *DWeb*.

A *MTS* é a camada de intermediação que provê serviços às demais camadas (Camada

Cliente e Camada de Persistência). Para atingir este objetivo, ela é construída na forma de uma arquitetura baseada em serviços.

O termo multicamada, advém do mecanismo que estabelece um protocolo de comunicação entre camadas de transparência semântica, permitindo a sua replicação e o aumento da escalabilidade dos serviços prestados às demais camadas. Este mecanismo é estabelecido através da troca de mensagem entre camadas de transparência semântica.

Com este mecanismo forma-se um agrupamento de computadores (*hardware* e *software*) com balanceadores de carga, cuja finalidade é aumentar a escalabilidade e utilização distribuída de recursos de informação, com todas as facilidades de ambiente único e processamento mais ágil.

4.8.1 Serviços de *Middleware*

O *DWeb* utiliza serviços de *middleware* para obter interoperabilidade em nível de *middleware*. A opção por serviços de *middleware* é a arquitetura *CORBA*. Esta arquitetura foi selecionada por ser um padrão aberto e com alto nível de interoperabilidade [28], [24].

Na arquitetura orientada a serviços de *middleware* do *DWeb*, os elementos das comunidades e/ou organizações se vinculam através de um modelo de rede. Este modelo de rede e alguns de seus principais elementos podem ser vistos na Figura 10.

Neste modelo, pode-se identificar três domínios de alto nível: domínio de sistemas internos, domínio *Web* pública e domínio Internet pública.

No domínio relativo aos sistemas internos existe um conjunto de elementos que estão relacionados a seguir:

- Internet privada (recursos de *intranet*): pode-se citar como exemplo: diretórios de rede, *links* de comunicação, roteadores, redes locais, etc.;
- Serviços de diretórios: definem a estrutura organizacional da empresa/comunidade virtual de prática e as pessoas que dela fazem parte;
- Serviços de segurança: definem um contexto de segurança distribuída baseada em *BSD - Business System Domain*. A autenticação de usuários e servidores é feita com certificados digitais que permitem aos usuários se conectar uma vez e serem autenticados automaticamente para acesso a cada sistema;
- Serviços de intermediação de mensagens: incluem serviços de troca de mensagens, serviços de transformação de mensagens e serviços de notificação;
- Serviços de arquivos, correios e lista de trabalho: as listas de trabalho se referem a um componente de iteração com o componente de *Workflow*;
- Serviços de dispositivos: fornecem serviços de acesso a dispositivos;
- Serviços de estrutura organizacional: fornecem informações sobre a estrutura organizacional;
- Serviços de domínio de negócio (*BSD - Business System Domain*): servidores de aplicação, servidores de banco de dados, servidor de segurança, etc.;
 - Filas de mensagens: gerencia filas de entrada e saída de mensagens;
 - Aplicações: os servidores de aplicação privados fornecem as funcionalidades específicas e necessárias aos usuários internos;

- Servidores *Web*: os usuários interagem via servidores *Web BSDs*;
- Serviços de *firewall* e acesso de usuários remotos (segurança): autenticação, autorização, etc.

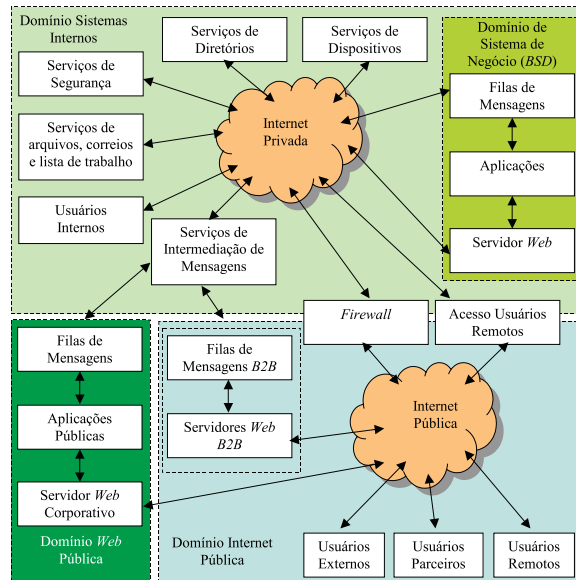


Figura 10: Arquitetura Orientada à Serviços de Middleware do *DWeb* baseada em [24].

O domínio *Web* pública[24], está relacionado aos serviços que as comunidades e/ou organizações oferecem aos usuários externos, que são:

- Filas de mensagens: têm a função de enviar e receber mensagens aos *BSDs* internos e associados;
- Aplicação *Web* pública: os servidores de aplicações públicas fornecem as funcionalidades específicas e necessárias aos usuários externos;
- Servidores *Web* corporativos: os servidores *Web* públicos funcionam de forma semelhante aos servidores *Web BSDs* internos, exceto pelos controles de segurança que são diferentes.

O domínio da Internet pública[24] especifica como os recursos de informação são acessíveis através dos seguintes elementos:

- Filas de mensagens *B2B*: as trocas de mensagens *HTTP* permitem as comunicações *B2B* (*Business-to-Business*) entre processos de negócio de comunidades e/ou organizações e parceiros de negócio (comunidades e/ou organizações);
- Servidores *Web B2B*: os servidores de aplicação *Web B2B* fornecem as funcionalidades específicas e necessárias à troca de mensagens *B2B* entre processos de negócios de comunidades e/ou organizações e parceiros de negócio (comunidades e/ou organizações);
- *Firewall*: fornece a conexão para que os usuários internos acessem os recursos de informação através da *Web* na Internet pública, restringindo o acesso à rede privada por usuários e sistemas computacionais externos;
- Servidor de acesso remoto: permite que usuários obtenham acesso aos sistemas computacionais utilizados por comunidades e/ou organizações através da rede pública de forma remota.

4.8.2 Comunidades DWeb

As comunidades *DWeb* (Seção 3) modelam os aspectos informais nos relacionamentos de troca de informações e conhecimentos. Eventualmente, podem modelar aspectos formais quando de sua vinculação à Unidade Organizacional (Subseção 4.8.3).

4.8.3 Estrutura Organizacional DWeb

O ambiente *DWeb* contempla a construção de estruturas organizacionais. Estas estruturas são formadas de Unidades Organizacionais⁹, que se agrupam para formar os organogramas organizacionais correspondentes ao ambiente. Os organogramas representam subdivisões funcionais das organizações catalogadas no *DWeb*.

A *UO* representa as divisões formais da organização que está sendo modelada para funcionar sob a arquitetura *DWeb*. Ela pode representar: a própria organização, diretorias, departamentos, divisões, etc. Neste contexto, uma Unidade Organizacional tem uma relação de um-para-um com Comunidades Virtuais de Prática. Desta forma, antes de se criar uma *UO* deve-se primeiro criar uma Comunidade Virtual à qual ela estará vinculada durante todo o seu ciclo de vida, que é criação, ativação, manutenção, desativação e possível exclusão.

4.8.4 Núcleo Semântico de Inferência

O Núcleo Semântico de Inferência é o principal serviço do ambiente *DWeb*. Ele é uma aplicação computacional que realiza inferências (raciocínios) sobre uma Base de Conhecimento. Esta Base de Conhecimento possui sentenças¹⁰ lógicas, que representam explicitamente o conhecimento sobre o mundo selecionado com a finalidade de resolver problemas. Nesta aplicação, o conhecimento é o ponto central e contribui com a formação da característica de inteligência no *DWeb*. Este conhecimento é delimitado no contexto de comunidades virtuais de prática [40], [43].

O Núcleo Semântico de Inferência deve conter duas vertentes: uma relativa ao conhecimento e outra relativa à sua representação. O conhecimento, pode ser entendido e manipulado pelo ser humano. A representação do conhecimento em meio computacional possibilita o seu processamento pelo computador, como por exemplo, a representação simbólica. Em uma perspectiva do conhecimento, observa-se que a Base de Conhecimento deve descrever **o que** o sistema deve fazer. Da perspectiva simbólica, as sentenças da Base de Conhecimento devem indicar **como** o sistema irá proceder. Desta forma, se separa o problema expresso pela análise e modelagem da atividade de representar este método em um formalismo matemático com enfoque eficiente em termos computacionais.

O entendimento do processo de resolução de problemas é fundamental para o entendimento do funcionamento do Núcleo Semântico de Inferência. Inicialmente considera-se a capacidade de raciocínio, que nada mais é do que a forma como se chega a algumas conclusões sobre determinados problemas analisados que, em outras palavras, definem o mecanismo de geração de novos conhecimentos. Aliado a este processo deve-se ter alguma forma de otimização das conclusões, pois, se a conclusão não vier em tempo hábil poderá não ter valor útil. No processamento cerebral, o uso de paralelismo promovido pelos neurônios acelera o processo de resolução de problemas. No Núcleo Semântico de Inferência, tenta-se otimizar este tempo através de heurísticas que agregam valor decisório à capacidade de raciocínio.

⁹A partir deste momento, uma Unidade Organizacional será referenciada por *UO*.

¹⁰São expressas em linguagens de representação de conhecimento e possuem uma sintaxe especificando-as de maneira bem formada.

Uma heurística para acelerar o processo de resolução de problemas consiste em guiar o processo de raciocínio. Os especialistas possuem a capacidade de determinar uma seqüência de operações de raciocínio que encontre soluções mais rapidamente. Esta técnica permite o aproveitamento de raciocínios anteriores e é de fundamental importância na resolução de problemas. Com a utilização destas heurísticas otimiza-se em muito o tempo final das inferências, simplesmente cortando caminho.

O conhecimento sobre como resolver problemas constitui uma classe de conhecimento. Neste contexto é importante destacar que as heurísticas dependem, em grande parte, do conhecimento sobre a solução de problemas, ou seja, da experiência dos solucionadores dos problemas que o Núcleo Semântico de Inferência se propõe a resolver.

Foram criadas diversas linguagens de representação do conhecimento. Dentre as linguagens lógicas, a lógica proposicional, é a mais simples. Ela representa somente os fatos. A lógica de predicados de primeira ordem¹¹ é uma lógica mais poderosa e que permite representar a maioria dos fatos, objetos e as relações entre eles [26], [31]. Outra forma de lógica mais elaborada e poderosa é a lógica temporal [31] que tem a possibilidade de representar tudo o que a lógica de predicados de primeira ordem representa, mais a variante de temporalidade. Os sistemas de regras de produção (par condição-ação) também podem ser utilizados para a representação do conhecimento.

Existem diversos modos de raciocínio lógico ou processos de inferência. Estes processos se caracterizam pela maneira como fazem a descoberta de conhecimento em uma Base de Conhecimento. Os procedimentos de inferência operam sobre os axiomas e os fatos específicos do problema para derivar os fatos que interessam aos usuários solicitantes.

Algumas regras simples de inferências podem ser aplicadas em sentenças com quantificadores para obter sentenças sem quantificadores. A inferência em lógica de predicados pode ser feita convertendo-se a Base de Conhecimento para lógica proposicional e usando-se esta lógica no processo de inferência. Dentre os diversos modos de raciocínio, pode-se destacar o encadeamento para frente (*Forward Chaining*) e o encadeamento para trás (*Backward Chaining*), os quais devem ser implementados no DWeb.

O encadeamento para frente se caracteriza por partir de sentenças (regras de inferência) da Base de Conhecimento para produzir novas conclusões (afirmações). São também chamados de procedimentos dirigidos a dados (*Data Driven*), pois o processo de inferência não é direcionado para resolver um problema particular.

No encadeamento para trás, parte-se de uma hipótese que se quer provar procurando regras na Base de Conhecimento, retroativamente, para analisar as assertivas que suportam a hipótese em questão. Se a premissa (antecedente) é uma conjunção, seus termos são processados um a um. Este algoritmo dirige a sua busca por objetivos, no sentido de atingir a premissa e é, normalmente, utilizado em sistemas de programação em lógica, tais como: *Prolog*.

O DWeb deve implementar o Núcleo Semântico de Inferência que raciocina logicamente sobre recursos de informação heterogêneos e distribuídos. Este raciocínio deve considerar as representações em metadados e Bases de Conhecimentos representados por ontologias de domínio no contexto de comunidades virtuais de prática.

4.8.5 Arquitetura de Serviços DWeb

A arquitetura de serviços DWeb (Figura 11), tem como característica fundamental o caráter de reutilização de serviços e evolução das aplicações DWeb (Módulos). Esta arquitetura

¹¹Alguns autores referem-se a esta lógica com um nome mais curto: Lógica de Predicados.

especifica um padrão de acesso e utilização dos serviços pelas aplicações *DWeb* e a forma como estes são estruturados internamente.

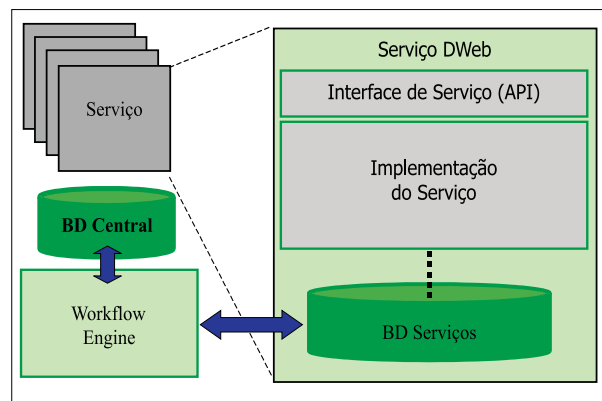


Figura 11: Camada Cliente da Arquitetura do Projeto *DWeb*.

Todos os serviços definidos e construídos no *DWeb*, deverão ter uma *API* (Interface de Programa de Aplicação) que permita aos demais serviços e módulos a sua completa utilização. Esta *API* deve ser documentada e divulgada às equipes de *DWeb*. Nela, a interface com os serviços deverá fornecer a sintaxe de comandos, os parâmetros, resultados, restrições, funcionalidades e tudo mais que for necessário à sua correta utilização.

A implementação dos serviços, deve seguir padrões tecnológicos de engenharia de software, conforme a *IEEE 12207.0 - 1996*, a qual se refere ao padrão para engenharia de software - Ciclo de Vida do Processo de Software do *IEEE/EIA* especificados em [47] e [46]. Nestas implementações serão definidas as classes de serviços *DWeb*, em sua maioria construídas utilizando a tecnologia Java. A utilização da tecnologia Java está sendo escolhida pelo seu amplo uso pela comunidade científica, por ser padrão aberto e o alto poder de computação que ela possui[24].

O *BD* de serviços é uma base de dados de controle que está localizada no *BD Central*, em uma *CTS* do *DWeb*. Ela mantém uma base de informações necessárias à utilização do serviço. São informações específicas de cada serviço para cada comunidade.

O *DWeb* mantém, em sua Base de Dados Central, uma estrutura de dados (*BD Controle de Serviço*) que possibilita um controle sobre os serviços disponibilizados no ambiente *DWeb*. Todos os serviços disponibilizados deverão ser registrados (registro de serviço *DWeb*) nesta base de dados. Desta forma, os **Clientes** podem obter acesso aos serviços disponibilizados pelo ambiente.

Algumas das informações registradas e mantidas para controle dos serviços são: código, descrição resumida, descrição completa, autoria, documentação de atualização, serviços vinculados, módulos vinculados, serviços dependentes, situação de uso (ativo, desativado permanentemente, desativado temporariamente, funcionamento exclusivo, etc.), autorização do gestor do ambiente *DWeb*, *LOG* de monitoramento, etc.

Todo Módulo que usa o serviço deve consultar esta base de informações para obter garantia de que o serviço é um serviço ativo no ambiente e sua utilização é legal. Com esta forma de atuação pretende-se aprimorar a segurança em profundidade de camadas implementada pelo *DWeb*[37].

4.8.6 Projetos (Módulos) DWeb

No *DWeb* operam diversos sistemas computacionais simultaneamente, compartilhando os diversos serviços do ambiente.

Os Módulos ou Projetos *DWeb* são semelhantes aos serviços básicos. Eles modelam os sistemas computacionais que estão compartilhando a infra-estrutura de serviços *DWeb*. Estes Projetos *DWeb* fazem uso dos Serviços *DWeb* através da *API* de cada Serviço, onde os desenvolvedores usam os serviços transparentemente e tratam as funcionalidades dentro de cada sistema computacional.

4.8.7 Engines DWeb

O Projeto *DWeb* possui em sua arquitetura um conjunto de Serviços especiais de extrema importância e complexidade. Estes serviços são os *Engines DWeb* e são constituídos por um grupo de 5 serviços: Mediação, *Workflow*, Inteligência, Busca Semântica e Conexão.

1. Engine Mediação

O *Engine Mediação* é um mecanismo que implementa a homogeneização dos recursos de informação que são heterogêneos e distribuídos. Para executar este serviço, a arquitetura do *DWeb* deve prover um mecanismo de *mediação/tradução* [48].

Nesta arquitetura, os usuários executam solicitação (*query*) de informações e/ou conhecimentos de comunidades ao **mediador** (*mediator*), e este, encaminha estas solicitações aos **tradutores** (*wrappers*), que têm a função de traduzir estas requisições para as fontes de informações reais, que são heterogêneas e distribuídas e retornar os resultados aos **mediadores**, que têm a função de apresentar estes resultados de forma homogênea aos usuários.

2. Engine Workflow

O *Engine Workflow* fornece um serviço básico que automatiza os processos de negócio orientados a comunidades e/ou organizações. Em síntese, um sistema de *Workflow* é montado sobre definições de quem faz o que, de que forma, quando e quais os caminhos que levam e trazem os pacotes de dados e informações (recursos de informação) que dão vida a um processo de negócio[24], [30].

Neste contexto, são definidos os principais elementos do *Engine de Workflow*, que são: os papéis, as regras e as rotas.

Os papéis são definidos como: conjunto de características e habilidades necessárias para executar determinada tarefa ou tarefas pertencentes a uma atividade. Estes papéis são implementados no *DWeb* através de mecanismo *RBAC* (**R**ole **B**ase **A**ccess **C**ontrol).

As regras são definidas como: atributos que definem de que forma os dados que trafegam no fluxo de trabalho devem ser processados, roteados e controlados pelo *Engine de Workflow*.

As rotas são os caminhos lógicos que, definidos sob regras específicas, têm a função de transferir a informação dentro do processo, ligando as atividades ao fluxo de trabalho.

O gerenciamento de fluxo de trabalho é um elemento-chave da arquitetura de integração provida pelo *DWeb*[24].

3. *Engine* de Inteligência

O *Engine de Inteligência* é composto de mecanismos para representação de conhecimentos, baseados principalmente em ontologias e manipulação destes conhecimentos através de mecanismos de inferência. Uma outra vertente da inteligência está nos mecanismos de representação de recursos de informação promovidos pelos metadados[24].

4. *Engine* de Busca Semântica

O mecanismo de busca semântica ou *Engine* de Busca Semântica, se baseia em uma arquitetura multiagente que agrega todo o potencial da tecnologia de agentes. Os agentes de softwares, navegarão no ambiente do *DWeb* em busca de informações[1]. Neste ponto, deve-se notar que a filosofia é de prover uma arquitetura distribuída de informações, em contraste com os mecanismos de busca tradicionais[10], que são centralizados, visando otimizar a busca de informações e distribuindo também o processamento em cada CTS, evitando um crescimento exponencial em cada base de índices local[1].

Para a localização de informações no ambiente local do *DWeb*, um agente de localização de informações locais, pertinentes ao contexto das comunidades virtuais de prática, é disponibilizado localmente. Para localizar informações externas, o mecanismo de busca semântica, utiliza agentes de busca externa, que têm a função de negociar com outros ambientes na Web Semântica, a localização, em seus índices locais, de uma informação relevante. Estas informações poderão ou não estar disponíveis ao mundo exterior com base em critérios locais de segurança[37].

O fornecimento de informações solicitadas é de responsabilidade dos módulos internos. Dentre esses módulos, existe o *NLP (Natural Language Processing)*, que tem a função de analisar o texto ou parâmetros fornecidos para a realização de busca. Nesta análise, são identificados os termos baseados em tecnologia de thesaurus e ontologias e suas estruturas lingüísticas [19],[39].

O texto, após ser analisado pelo *NLP*, é passado ao Extrator de Informações (IE) que tem a função de extrair informações de documentos em formatos heterogêneos (ex.: arquivos *PDF (Portable Document Format)*, mensagens de correio eletrônico, páginas *Web*, banco de dados, etc.) e converter em um único formato homogêneo. Estas informações devem passar por filtros usando a tecnologia mediação-tradução [11], [48].

A extração de informações apoiar-se-á em um mecanismo de extração de padrões usando tecnologia de aprendizado de máquina[11]. Este mecanismo proverá ao IE os padrões necessários para tomadas de decisões na busca de informações e criação de índices.

Os recursos de informação solicitados pelo usuário são fornecidos com um sumário. Este sumário orienta os usuários na obtenção dos recursos de informação evitando tráfego desnecessário de documentos extensos. A composição destes sumários é feita a partir de informações oriundas de metadados registrados na catalogação dos documentos ou após processamento de geração automática de sumários.

5. *Engine* de Conexão

O *Engine* de Conexão implementa a comunicação entre camadas de transparência semântica. Ele fornece uma arquitetura fortemente escalonável, onde esta arquitetura é fortemente integrada e distribuída com a finalidade de executar serviços e módulos *DWeb*, ou seja, implementa um mecanismo de otimização do tempo de execução de solicitação de serviços *DWeb*[34].

Para realizar este requisito, o *DWeb* deve implementar a arquitetura de agrupamento (*Cluster*).

Nesta arquitetura, existe um elemento geral do *DWeb* que implementa funcionalidades múltiplas, ajustáveis através de configuração para o funcionamento desejado pelo gestor do ambiente *DWeb*.

Este elemento é o *Engine* de Conexão *DWeb*. Com esta característica um ambiente pode funcionar como uma entidade de transparência semântica independente, ou como um agrupamento de entidades de transparência semântica[24], [34], [36].

Na modalidade agrupamento de entidades de transparência semântica, o *Engine* de Conexão *DWeb* pode ser configurado para trabalhar como um coordenador de solicitação de serviços. Nesta modalidade, somente uma camada de transparência deverá ser configurada para o serviço de balanceamento. O *Engine* de Conexão *DWeb* deve gerir todas as solicitações de serviços *DWeb* e os seus respectivos retornos de solicitação de serviços *DWeb* para o seu solicitante[24], [36].

O agrupamento *DWeb*, possui facilidades para distribuição de solicitação de serviços *DWeb* em múltiplos níveis. Cada nível deverá ser composto de uma unidade de balanceamento e uma quantidade de nós executores, configurável pelo gestor do ambiente *DWeb*.

Este artifício arquitetônico, permite uma arquitetura escalonável, teoricamente e virtualmente infinita e abre uma janela muito adequada para gerenciamento de comunidades em alto nível, pois pode-se segmentar conjuntos de comunidades em servidores *DWeb* e otimizar ainda mais as solicitações de serviços, através de direcionamento de solicitação de serviços em segmentos organizados por agrupamentos dedicados.

4.9 Camada Persistência Distribuída

Esta camada (Figura 12) implementa o aspecto distribuído do *DWeb*. A filosofia do *DWeb* é manter os recursos de informação na origem e disponibilizá-los via mecanismo de busca semântica na arquitetura mediador/tradutor [48]. Em caso de necessidade, por exemplo para otimização e segurança, pode-se configurar uma comunidade virtual de prática para ter replicação dos recursos de informação na MTS.

Os recursos de informação são mantidos localmente e o serviço de tradução os tornam disponíveis às comunidades *DWeb* e acessíveis via serviço de mediação centralizado na MTS da arquitetura do *DWeb*. O serviço de busca semântica passa, dessa forma, a ter acesso aos recursos de informação com possibilidade de aplicação do serviço de inferência. Desta maneira, pode-se raciocinar sobre os recursos de informação e obter as informações e conhecimentos desejados de forma transparente e precisa.

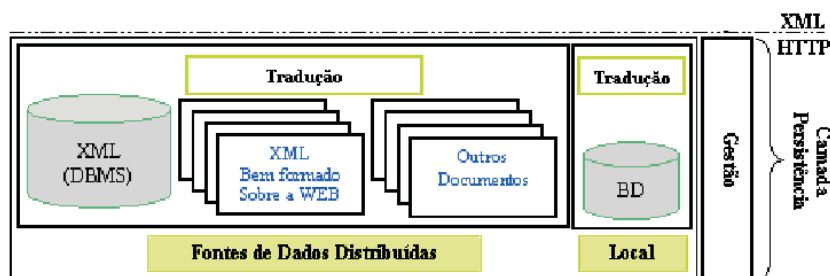


Figura 12: Camada de Persistência Distribuída da Arquitetura do Projeto *DWeb*.

4.9.1 O Mecanismo de Mediação/Tradução

A heterogeneidade semântica tem sido apontada como um dos problemas mais importantes e difíceis de se tratar adequadamente. Ela envolve a interoperabilidade e cooperação entre diversas fontes de recursos de informação, onde se acentua as diferenças sintáticas, semânticas e estruturais entre os diversos sistemas. Na área de Banco de Dados, problemas relacionados aos aspectos de intercâmbio, compartilhamento e integração de dados têm sido uma preocupação de pesquisadores, mas não se encontram totalmente bem resolvidos e documentados na literatura técnica.

No contexto da *Web* não se aplica a abordagem de integração de dados, necessitando evoluir para novos paradigmas. Um destes paradigmas é denominado de “sistemas virtuais”. Em tais sistemas empreende-se uma visão de integração de dados sem as respectivas materializações através de mecanismos de mediação/tradução.

Tais mecanismos devem prover uma interface uniforme para acesso aos dados, de tal forma que os usuários e as aplicações tenham uma visão heterogênea, autônoma e distribuída das fontes de dados, bem como mecanismos para acessar conteúdos e suas descrições implementadas em metadados.

O elemento de mediação, deste mecanismo, é uma aplicação computacional (serviço *DWeb*) que estabelece uma camada ativa entre os usuários e as fontes de dados do lado do cliente, independente destas fontes. Ela pode acessar diversas fontes heterogêneas de informação, bem como outros elementos de mediação através de tradutores.

Segundo [48], alguns dos serviços fornecidos pelo elemento mediador são:

- Invocação de transação de tradução para fontes de dados legada;
- Seleção real de fonte de dados relevante;
- Estratégias de otimização de acesso para prover resposta rápida e de baixo custo;
- Imposição de filtros de segurança para vigiar dados privados;
- Resolução de terminologias de domínio e ontologia diferentes;
- Resolução de escopo misto;
- Interpolação ou extrapolação para emparelhar diferenças em dados temporais;
- Redução de dados históricos para instantâneos limitados;
- Omissão de replicação de informação;
- Poda de dados enfileirados de baixa qualidade ou relevância;
- Informe de divergências de valores esperados ou tendências; etc.

Os tradutores (*wrappers*) têm como principal tarefa a tradução de consultas submetidas pelo elemento de mediação para uma linguagem da fonte hospedeira do destino.

A implementação deste mecanismo representa um desafio muito interessante para efeito da pesquisa científica. O *DWeb* se propõe a investigar, principalmente, nos seguintes tópicos:

- O número de fontes de dados pode ser muito elevado tornando a integração e resolução de conflitos um problema difícil de se resolver;

- O dinamismo a que estas fontes estão sujeitas, tais como, criação e remoção de fontes teriam que ser realizadas de forma a não causar impacto;
- Os tipos de fontes de dados podem variar muito, desde fontes estruturadas (ex.: banco de dados), semi-estruturadas (ex.: arquivos de *e-mails*, páginas *Web*) e não estruturadas de informação, não fornecendo nenhuma informação capaz de auxiliar no processo de integração.

A integração é tratada pelo *DWeb* como um mecanismo de mediação/tradução (*mediator/wrappers*) onde o mediador é implementado na camada *MTS* e os elementos de tradução (*wrappers*) são implementados na camada de persistência distribuída.

A camada de persistência distribuída contempla um elo de ligação com a camada *MTS*, que é, o elemento que realiza o serviço de tradução (*wrappers*). Este elemento faz parte do conceito do mecanismo de mediação/tradução. O serviço de mediação, localizado na *MTS*, intermedia o sistema de informação, ligando recursos de informação e o programa de aplicação. Ele implementa uma interface única sobre a qual os recursos de informação são virtualmente visualizados.

4.9.2 As Fontes Distribuídas/Locais de Informações

As fontes de dados locais e distribuídas são compostas por recursos de informação estruturados, semi-estruturados e não estruturados. Cada recurso de informação é tratado adequadamente pelo seu respectivo serviço de tradução (*wrappers*). O serviços de tradução, da camada de persistência distribuída, se integra ao serviço de mediação, da camada *MTS*, onde a homogeneização do recurso de informação ocorre[48].

As fontes de informação local ou distribuída sofrem o mesmo tratamento de forma a serem agrupados pelo serviço de mediação da camada *MTS* e, assim, serem disponibilizados aos clientes ou módulos (projetos) solicitantes do serviço.

5 Considerações Finais

O *DWeb* pretende contribuir com o desenvolvimento da Web Semântica, principalmente, no que tange ao seu aspecto de interdisciplinaridade. As disciplinas incorporadas inicialmente são: Inteligência Artificial, Engenharia de Software, Banco de Dados, Redes de Computadores e Linguística Computacional.

Para prover uma arquitetura de integração organizacional voltada para comunidades virtuais de prática, o *DWeb*, em uma visão de alto nível, deve ter uma arquitetura com as seguintes características: computação distribuída, aplicações baseadas em componentes, processos orientados a eventos, acoplamento fraco de funções de negócio, informações de suporte a decisão, gerenciamento de fluxo de trabalho, acesso à Internet, personalização de interface e possuir mecanismos que possibilitem a implementação de uma inteligência computacional.

Para contemplar os macro requisitos da Web Semântica, o *DWeb* é implementado em uma arquitetura que se divide em três camadas conceituais, que são: Cliente, *MTS* e camada de Persistência Distribuída. Este modelo arquitetônico possibilita vantagens, tais como: flexibilidade, escalabilidade, reutilização e também, especificamente, os macro requisitos da Web Semântica, que são: integração, interoperabilidade e inteligência.

A camada Cliente, preocupa-se, principalmente, com os aspectos da personalização do ambiente *DWeb* junto a cada usuário de forma a contemplar os aspectos da acessibilidade, usabilidade, utilidade e personalização.

A camada *MTS*, preocupa-se, principalmente, com os aspectos gerenciais e transacionais. Nela, as comunidades virtuais de prática, encontram respaldo tecnológico baseado em uma arquitetura de serviços e módulos, o que torna o ambiente genérico. Além disso, contempla os seguintes serviços: *middleware*, mediação, comunidades, estrutura organizacional, *workflow*, inteligência, busca de informações e conexão.

A camada Persistência Distribuída, preocupa-se, principalmente, com os aspectos ligados à distribuição e tradução de recursos de informação e conhecimentos provenientes de comunidades virtuais de prática. Desta forma, se integra aos Clientes através do mecanismo de mediação/tradução com o(s) *MTS*(s).

O *DWeb* propõe minimizar as limitações da *Web* existente, principalmente, no que tange à pesquisas de informações imperfeitas[1] e a *Web Oculta*[2], para tanto, pretende-se utilizar, principalmente, das tecnologias Web Semântica e dos conceitos de comunidades virtuais de prática.

Atualmente, a recuperação de informação na *Web* é imperfeita por apresentar um baixo índice de relevância em relação às necessidades de informações dos usuários. Este baixo índice de relevância origina-se do fato dos mecanismos de recuperação de informação atuais realizarem buscas baseadas na sintaxe de palavras-chave. A Web Semântica se utiliza de novas tecnologias que permitam a realização de buscas baseadas na semântica das informações (seu significado).

O *DWeb* propõe a junção de tecnologias da Web Semântica com conceitos de comunidades virtuais de prática como forma de contextualizar as informações. Desta forma, as buscas serão restringidas às bases de conhecimentos e recursos de informação associadas a este contexto, permitindo assim, um aumento substancial nos índices de relevância dos resultados obtidos.

Outra vertente dos problemas da *Web* existente, diz respeito ao fato de que a maioria das informações estão em bancos de dados. Desta forma, as informações ficam inacessíveis aos mecanismo de busca de informações atuais, fato este conhecido como *Web Oculta*. O *DWeb* propõe a intermediação do tratamento de recursos de informação estruturados (banco de dados) via mecanismo de mediação/tradução.

O *DWeb* propõe a criação de mecanismos de software de navegação especial, baseados em tecnologias assistivas, com vista a obter um aumento da acessibilidade através de tratamento de recursos de informação baseado nos seus metadados de referência. Desta forma, pode-se apresentar os recursos de informação aos usuários com diferentes níveis de dificuldades. Dentre estes tratamentos pode-se citar: leitura de resumo, aumento do tamanho da página *Web* apresentada, etc.

A transparência oferecida pelo ambiente proposto é fundamental por reunir os usuários com diferentes níveis de conhecimentos em computação para o trabalho cooperativo e, desta forma, integrando-os em comunidades virtuais de prática. Esta aproximação é importante por permitir comunicação global, facilitando acesso a informações pertinente ao contexto, enquanto contribui para a implementação desejada e efetiva da Web Semântica.

6 Agradecimento

A Profa. Dra. Ana Paula Laboissière Ambrósio, pela avaliação do presente texto e pelas sugestões feitas, as quais muito contribuíram para a melhoria do texto original.

Referências

- [1] BAEZA-YATES, R; NETO, B. R. **Modern Information Retrieval**. ACM Press, 1st edition, 1999.
- [2] BERGMAN, M. K. **The Deep Web: Surfacing Hidden Value**. The Journal of Electronic Publishing, 7(1), August 2001.
- [3] BERNERS-LEE, T. **Keynote on Evolvability**. <http://www.w3.org/DesignIssues/Evolution.html>, viewed May 2005, 1998.
- [4] BERNERS-LEE, T. **Principles of Design**. <http://www.w3.org/DesignIssues/Principles.html#KISS>, viewed May 2005, 1998.
- [5] BERNERS-LEE, T. **Semantic Web Road map**. <http://www.w3.org/DesignIssues/Semantic.html>, viewed May 2005, 1998.
- [6] BERNERS-LEE, T. **Uniform Resource Identifiers (URI): Generic Syntax**. <http://www.ietf.org/rfc/rfc2396.txt>, viewed May 2005, 1998.
- [7] BERNERS-LEE, T. **Semantic Web Architecture**. <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html>, viewed April 2005, 2004.
- [8] BERNERS-LEE, T; HENDLER, J; LASSILA, O. **The Semantic Web**. Scientific American, May 2001.
- [9] BOGO, L. H. **Criação de Comunidades Virtuais a Partir de Agentes Inteligentes, Uma Aplicação em eLearning**. Master's thesis, Universidade Federal de Santa Catarina - UFSC, Florianópolis - Santa Catarina - Brasil, 2003. In Portuguese.
- [10] BRIN, S; PAGE, L. **The Anatomy of a Large-Scale Hypertextual Web Search Engine**. In: PROCEEDINGS OF THE 7TH INTERNATIONAL WORLD WIDE WEB CONFERENCE, p. 107–17, 1998.
- [11] C., A. K. **Extraction Knowledge: Artificial Intelligence Tools Offer New Ways to Explore Web Content**. Intelligente Portals Feature, 2005.
- [12] CHRISTOPHIDES, V; ET AL.. **Community Webs (C-Webs): Technological Assessment and System Architecture**. <http://cweb.inria.fr/>, viewed May 2005, 2000.
- [13] CONSORTIUM, W. W. W. **Unicode and ISO/IEC 10646 in parallel define the Universal Character Set (UCS)**. <http://www.w3.org/International/O-unicode.html>, viewed May 2005, 1996.
- [14] CONSORTIUM, W. W. W. **Web Content Accessibility Guidelines 1.0**. <http://www.w3.org/TR/WAI-WEBCONTENT/>, viewed May 2005, 1999.
- [15] CONSORTIUM, W. W. W. **HTTP - Hypertext Transfer Protocol**. <http://www.w3.org/Protocols/>, viewed may 2005, 2000.
- [16] CONSORTIUM, W. W. W. **Metadata Activity Statement**. <http://www.w3.org/Metadata/Activity>, viewed May 2005, 2004.

- [17] CONSORTIUM, W. W. W. **Resource Description Framework (RDF)**. <http://www.w3.org/RDF/>, viewed May 2005, 2004.
- [18] CONSORTIUM, W. W. W. **Semantic Web**. <http://www.w3.org/2001/sw/>, viewed May 2005, 2004.
- [19] CONSORTIUM, W. W. W. **W3C Semantic Web Activity Statement**. <http://www.w3.org/2001/sw/#activity>, viewed May 2005, 2004.
- [20] CONSORTIUM, W. W. W. **Web Ontology Language (OWL)**. <http://www.w3.org/2004/OWL/>, viewed May 2005, 2004.
- [21] CONSORTIUM, W. W. W. **Extensible Markup Language (XML)**, 2005.
- [22] CONSORTIUM, W. W. W. **Leading the Web to Its Full Potential...** <http://www.w3.org/>, viewed May 2005, 2005.
- [23] CONSORTIUM, W. W. W. **W3C ACL System**. <http://www.w3.org/2001/04/20-ACLs.html>, viewed May 2005, 2005.
- [24] CUMMINS, F. A. **Enterprise Integration**. OMG Press/John Wiley and Sons. inc., 1st edition, 2002.
- [25] FILHO, J. T. **Comunidades Virtuais: Como as Comunidades de Práticas na Internet estão Mudando os Negócios**. Senac, 1th edition, 2002. In portuguese.
- [26] FITTING, M. **First-order logic and automated theorem proving**. Springer-Verlag New York, Inc, 2nd edition, 1995.
- [27] GAJOS, K; WELD, D. S; ANDERSON, C; DOMINGOS, P; ETIZIONE, O; LAU, T; WOLFMAM, S. **SUPLE: Automatically Generating User Interfaces**. University of Washington, Mar 2004.
- [28] GROUP, O. O. M. **OMG Specifications: The MDA, UML, CWM, CORBA, the OMA**. <http://www.omg.org/gettingstarted/>, viewed may 2005, 2005.
- [29] GRUBER, T. R. **Toward Principles for the Design of Ontologies Used for Knowledge Sharing**. International Workshop on Formal Ontology, Padova, Italy, August 1993.
- [30] HOLLINGWORTH, D. **Workflow Management Coalition - The Workflow Reference Model**. WfMC - Workflow Management Coalition, January 1995.
- [31] HUTH, M. **Logic in Computer Science - Modelling and Reasoning about Systems**. Cambridge University Press, 2nd edition, 2004.
- [32] LESSER, E. L; STORCK, J. **Communities of Praticce and Organizational Performance**. IBM Systems Journal, 40(4), 2001.
- [33] LESSER, E; PRUSAK, L. **Communities of Practice, Social Capital and Organizational Knowledge**. IBM Institute for Knowledge Management, 1999.
- [34] MENASCE, D. A; ALMEIDA, V. A. F. **Capacity Planning for Web Services**. Prentice Hall PTR, 1th edition, 2002.

- [35] MOOR, A. D; HEUVEL, W; GENESERETH, M. R. **Web Service Selection in Virtual Communities**. In: PROCEEDING OF THE 37TH HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES/IEEE, 2004.
- [36] NGUYEN, J. V. **Designing ISP Architectures**. Prentice Hall PTR, 1th edition, 2002.
- [37] NORTHHCUTT, S; ET AL. **Inside Network Perimeter Security**. New Riders Publishing, 2th edition, 2003.
- [38] OBERLE, D; VOLZ, R; STAAB, S; MOTIK, B. **KAON SERVER - A Semantic Web Management System**. University of Karlsruhe, Institute AIFB, Germany, Jan 2002.
- [39] POPESCU, A.-M; ETIZIONI, O; KAUTZ, H. **Towards a Theory of Natural Language Interfaces to Databases**. IUI'03 USA - ACM 1-58113-586-6/03/0001, March 2003.
- [40] REZENDE, S. O. **Sistemas Inteligentes - Fundamentos e Aplicações (RECOPE-IA - Rede Cooperativa de Pesquisa em Inteligência Artificial in Portuguese)**. Editora Manole - Printed in Brazil, 1th edition, 2003.
- [41] RHEINGHOLD, H. **The Virtual Community**. Secker e Warburg, 1st edition, 1994. Also available at <http://www.rheingold.com/vc/book/>, viewed November 2005.
- [42] ROCHE, C. **Ontology: A Survey**. <http://ontology.univ-savoie.fr/activites/publications/IFAC%202003.pdf>, viewed May 2005, 2003.
- [43] RUSSELL, P; STUART, N. **Artificial Intelligence. A Modern Approach**. Prentice-Hall, 15th edition, 2004.
- [44] SABBATINI, R. M. **The Evolution of Human Intelligence (Brain and Mind, Electronic Magazine on Neuroscience)**. http://www.cerebromente.org.br/n12/mente/evolution/evolution_i.html, viewed August 2004, 2001.
- [45] SOARES, A. J; TIBO, M; TOMAZ, K. P; MAGALHÕES, D. **PMBOOK 2000 - O Universo de Conhecimento em Gerência de Projetos**. PMI - Project Management Institute (traduzido para o português pelo PMI MG), Jan 2002.
- [46] THAYER, R. H; CHRISTENSEN, M. **The Supporting Processes**, volume 2 de **Software Engineering**. IEEE Computer Society., 2th edition, 2002.
- [47] THAYER, R. H; DORFMAM, M. **The Development Process**, volume 1 de **Software Engineering**. IEEE Computer Society., 2th edition, 2002.
- [48] WIDERHOLD, G; GENESERETH, M. R. **The Basis for Mediation**. In: CONFERENCE ON COOPERATIVE INFORMATION SYSTEMS, p. 140–157, 1995.
- [49] WOOLDRIDGE, M; JENNINGS, N. R. **Intelligent Agents: Theory and Practice**. <http://www.doc.mmu.ac.uk/STAFF/mike/ker95/ker95-html.h> (Hypertext version of Knowledge Engineering Review paper), 1995.
- [50] ZURKO, M. E; SIMON, R; SANFILIPPO, T. **A User-Centered, Modular Authorization Service Built on an RBAC Foundation**. In: IEEE SYMPOSIUM ON SECURITY AND PRIVACY, p. 57–71, 1999.