

Web Semântica e sua aplicação na tomada de decisão médica

¹ ANDERSON DUTRA MOURA, ² LUCIANO CEBULA, ³ LUIZ FERNANDO DE ALMEIDA GARRETT, ⁴ MARCOS AUGUSTO HOCHULI SHMEIL

¹PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ-PUCPR
RUA IMACULADA CONCEIÇÃO, 1155, BAIRRO PRADO VELHO
80215-901 CURITIBA (PR) - BRASIL
anderson@s4.com.br

²PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ-PUCPR
RUA IMACULADA CONCEIÇÃO, 1155, BAIRRO PRADO VELHO
80215-901 CURITIBA (PR) - BRASIL
luciano@lumendesign.com.br

³PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ-PUCPR
RUA IMACULADA CONCEIÇÃO, 1155, BAIRRO PRADO VELHO
80215-901 CURITIBA (PR) - BRASIL
l.garrett@pucpr.br

⁴ PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ-PUCPR
RUA IMACULADA CONCEIÇÃO, 1155, BAIRRO PRADO VELHO
80215-901 CURITIBA (PR) - BRASIL
shm@ppgia.pucpr.br

Resumo

Este artigo apresenta o conceito de Web Semântica, projeto que busca tornar compreensível o conteúdo disponibilizado na Internet para usuários humanos e computacionais, criando um ambiente onde o conhecimento pode ser armazenado e explorado com maior facilidade, permitindo tomadas de decisão médica com alta qualidade.

PALAVRAS-CHAVE: Web Semântica, Internet, Inteligência Coletiva, Tomada de Decisão Médica.

Abstract - This paper shows the Semantic Web's concept, project that tries to make comprehensible the contents of the Internet to human and computer users, creating an ambient where the knowledge can be stored and explored easily, that allows medical-decision making with high quality.

Key-words: Semantic Web, Internet, Collective Intelligence, Medical-Decision Making.

1. Introdução

A revolução da informação teve seu início com a invenção da escrita na Mesopotâmia e desde então, tem sofrido constantes evoluções. Peter Drucker afirma que a revolução da informação será uma revolução de conceitos. Ela visa definir o significado

da informação e qual seu propósito, por exemplo: definir sobre os novos conceitos das tarefas que são realizadas com a ajuda da informação e das instituições e como efetuar essas tarefas [1].

Uma possibilidade de evolução se apresenta através da Inteligência Coletiva – um ambiente onde cada um dos membros comenta o que faz, tentando construir um saber comum, ao mesmo tempo em que têm liberdade de propor teorias diferentes.

De posse do conceito de Inteligência Coletiva e, apoiado por pesquisadores como Pierre Lévy, que afirmam que a Web é o ambiente ideal para o desenvolvimento deste conceito, este artigo tem como objetivo: a) Analisar o projeto Web Semântica, desenvolvido pelo W3C (*World Wide Web Consortium*), como uma padronização da linguagem utilizada na Internet, tornando-a compreensível para usuários humanos e computacionais; b) verificar se a Web Semântica pode ser considerada uma possível integradora dos dados contidos na Web; c) identificar possíveis aplicações da Web Semântica como base para o desenvolvimento de aplicações a serem utilizadas no processo de tomada de decisão médica.

2. Metodologia

Através de pesquisa bibliográfica e documentos científicos publicados na *web*, buscaremos uma ligação entre a tecnologia e conceitos que envolvem a *web* Semântica e os conceitos de gestão do conhecimento, inteligência coletiva e apoio à tomada de decisões de autores com reconhecidos trabalhos nesta área.

3. Inteligência Coletiva

É uma inteligência distribuída por toda à parte, incessantemente valorizada, coordenada em tempo real, que resulta em uma mobilização efetiva de competências. Ela se fundamenta no reconhecimento e o enriquecimento mútuo das pessoas e no conceito de sinergia, no qual, grupos podem encontrar mais ou melhores soluções do que a soma das soluções encontradas por indivíduos [2].

Como uma possibilidade de gerar um ambiente que permite um comportamento seletivo, resolvendo os problemas gerados pelo excesso de informação, surge a Inteligência Coletiva.

Para a criação desse ambiente são necessários dois fatores: a competição e a cooperação. A competição nascida da liberdade e a cooperação vinda do vínculo social, da amizade. Assim, a Inteligência Coletiva não surge da guerra de todos contra todos, nem da cooperação regulada ou obrigatória, mas de um ambiente que permita as diferenças de idéias, as lutas, os conflitos, que são naturais e que permitem ao novo se expressar [3].

Em seus estudos, Heylighen identificou que o seu desenvolvimento da Inteligência Coletiva é cerceado por obstáculos, que são [4]:

- a) Limites cognitivos dos indivíduos - Por mais competente que o participante seja, sua inteligência individual é limitada, e isto impõem uma restrição fundamental em sua habilidade de cooperar, sendo o resultado obtido através da cooperação dos membros do grupo;
- b) Jogos de Poder - é a luta pelo reconhecimento individual no grupo como o mais inteligente, significando a desconsideração da idéia do indivíduo considerado inferior, criando uma hierarquização da comunicação e abrindo espaço para o “estrelismo”. O resultado mais comum é que os ouvintes raramente prestam atenção no locutor, mesmo diante de sugestões inteligentes;

c) Complexidade de coordenação - é a dificuldade de se coordenar e assimilar a crescente quantidade de informações geradas e recebidas pelos participantes dessa sociedade.

Segundo Crawford, informações são encontradas em sistemas e ferramentas utilizadas pelo homem, mas o conhecimento, em forma de *expertise*, apenas nos humanos. Assim o conhecimento se auto-reproduz e pode ser disseminado, ele é compartilhável [5].

4. Inteligência Coletiva e a Internet

A exemplo da comunidade científica, onde cada um comenta o que faz, na tentativa da construção de um saber comum, no mesmo tempo em que cada indivíduo pode propor teorias diferentes, a Internet, com seus milhões de documentos à disposição, permite essa troca entre todas as pessoas do mundo, dentro dessa nova civilização. Mas essa nova civilização da inteligência coletiva passa por um esforço individual. A inteligência coletiva não vai pensar no lugar do indivíduo, justamente porque ela é o resultado da multiplicação dos esforços individuais para pensar em conjunto. Assim sendo, a formação dessa cultura, que tem como ferramenta a Internet, e que vai utilizar a Inteligência Coletiva e a cooperação intelectual como base, passa por um esforço individual de colaboração [3].

Dessa forma, Lévy afirma que o coletivo inteligente é a nova figura da cidade democrática. A cidade inteligente tem por finalidade seu próprio crescimento, sua densificação, sua extensão, seu retorno sobre si mesma e sua abertura ao mundo. Sua dinâmica está baseada na escuta de todas as idéias, a sua avaliação e a colocação em prática das idéias selecionadas, seguindo-se da validação dos resultados [2].

5. Web Semântica

A Web Semântica busca a representação dos dados da Internet. Este projeto se baseia em um esforço colaborativo, liderado pelo W3C (<http://www.w3.org/2001/sw/>), com participação de pesquisadores e parceiros corporativos. Ela se baseia no conceito de RDF (*Resource Description Framework*), que integra uma variedade de aplicações utilizando XML (*eXtensible Markup Language*) para sintaxe e URI (*Universal Resource Identifier*) para nomenclatura.

A Web Semântica não é uma *Web* separada, mas uma extensão da Internet atual, onde cada informação possui um significado bem definido, permitindo que computadores e pessoas possam trabalhar em cooperação. A propriedade essencial da Internet é a universalidade. O poder de um link de hipertexto é que “qualquer coisa pode ser vinculada a qualquer coisa”. A tecnologia Web, entretanto, não pode identificar a diferença entre um rascunho e um documento, entre informação comercial e acadêmica ou entre culturas, idiomas e mídias diferentes. Uma informação pode ter várias definições distintas, dependendo do assunto. Esta é uma das diferenças entre a informação produzida para consumo humano e para máquinas. Até o momento, a Web se desenvolveu rapidamente como um repositório de documentos para pessoas, ao invés de um repositório de dados e informações que pode ser processada automaticamente, sendo este o principal objetivo da Web Semântica [6].

Segundo o W3C, para que a Web Semântica funcione, os computadores deverão ter acesso a coleções estruturadas de informação e conjuntos de regras que eles poderão utilizar no processo de análise. A representação do conhecimento, como esta tecnologia é eventualmente chamada, ocupa uma posição semelhante a do hipertexto antes do

advento da Web. Isto contém a semente de aplicações importantes, mas para atingir seu potencial pleno, deverá estar ligado a um sistema global e único.

Sistemas de representação de conhecimento tradicional, normalmente, estão centralizados, obrigando que cada um deles compartilhe exatamente a mesma definição de conceitos comuns, como país ou veículo. Mas o controle centralizado é muito rígido, onde o tamanho e o escopo do sistema são inversamente proporcionais as facilidades de gerenciamento. Entretanto, estes sistemas limitam cuidadosamente as questões que podem ser realizadas, para que o computador possa fornecer respostas confiáveis ou responder a todas as perguntas realizadas. O problema é mencionado pelo Teorema matemático de Gödels:

Qualquer sistema que é complexo o suficiente para ser útil também armazena questões sem resposta, mais parecidas com versões sofisticadas do paradoxo básico “Esta sentença é falsa” [7].

Buscando evitar estes problemas, sistemas de representação de conhecimento tradicionais geralmente possuem seus próprios limites e conjunto idiossincrático de regras para gerar inferências sobre seus dados.

Os pesquisadores de Web Semântica, entretanto, consideram que os paradoxos e questões sem respostas são um preço a ser pago para se obter versatilidade. Seu grande desafio é criar uma linguagem que expresse dados e regras simultaneamente, permitindo a análise dos dados e regras a eles relacionados em qualquer sistema de representação de conhecimento, para sua posterior exportação para a Web. Adicionar lógica a Web, o que significa utilizar regras para realizar inferências, escolher cursos de ações e responder perguntas, é objetivo atual da comunidade que trabalha com Web Semântica. Uma união de matemática e engenharia de decisão está sendo utilizada para atingir este objetivo. Entre as tecnologias utilizadas para o desenvolvimento da Web Semântica é possível destacar:

- *eXtensible Markup Language* (XML), que permite ao usuário criar suas próprias *tags* – etiquetas ocultas como comentários em páginas Web ou seções de texto em uma página. *Scripts* ou programas podem utilizar estas *tags* de maneira sofisticada, mas os criadores dos *scripts* precisam conhecer a função das *tags*. Em resumo, o XML permite que usuários adicionem estruturas aos seus documentos [6];
- Resource Description Framework (RDF). Criado para aumentar a qualidade da informação fornecida pelo XML, que codifica isto em conjuntos de tuplas, onde os elementos correspondem ao sujeito, verbo e objeto de uma expressão elementar. Estas tuplas podem ser escritas utilizando *tags* XML. No RDF, um documento realiza afirmações que coisas em particular (pessoas, páginas Web, etc.) possuem propriedades (como “é irmã de” ou “é o autor de”) com certos valores (outra pessoa ou página Web). Esta estrutura caminha para se tornar o modo de descrever a grande maioria de dados processados por máquinas.

O XML permite a estruturação de dados, permitindo sua utilização tanto na Web, quanto em outros sistemas. Segue abaixo um exemplo da representação em HTML, utilizada na web e não-computacionalmente compreensível:

```
<HTML>
<BODY>
<H3>Identificação do Paciente</H3>
<UL>
<LI>Nome: Maria da Silva</LI>
<LI>Sexo: Feminino</LI>
<LI>Nascimento: 24/07/1990</LI>
```

```
<LI>Endereço: Rua Brasil, 23</LI>
<LI>Médico: Dr. João Santos</LI>
</UL>
</BODY>
</HTML>
```

Segue abaixo a representação em XML da mesma informação fornecida acima:

```
< Identificação>
< Nome>Maria da Silva< /Nome>
< Sexo>Feminino</ Sexo>
< Nascimento>24/07/1990</ Nascimento>
< Endereço>Rua Brasil, 23</ Endereço>
< Médico>Dr. João Santos</ Médico>
< Identificação>
```

Para uma melhor compreensão do papel do RDF no processo de representação do conhecimento, são fornecidos os seguintes exemplos:

a) Considere com exemplo a sentença abaixo:

Dr. João Santos é o examinador da paciente Maria da Silva.

Esta sentença é constituída pelos seguintes itens:

Sujeito (Recurso)	Maria da Silva
Predicado (Propriedade)	Examinador
Objeto (Literal)	Dr. João Santos

Tabela 1 – Constituição da sentença.

Agora, esta sentença será representada graficamente como uma declaração RDF utilizando um diagrama. Neste diagrama, os nodos representam recursos e as setas representam as propriedades. Nodos que representam strings literais serão desenhados como retângulos. A sentença acima seria representada pelo diagrama abaixo:

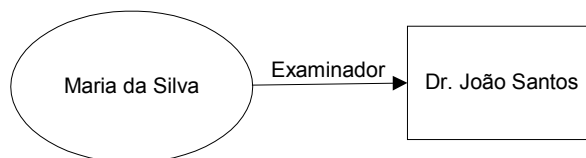


Figura 1 – Representação gráfica da sentença.

A direção da seta é importante. A flecha sempre começa no sujeito e aponta para o objeto da declaração. O diagrama acima poderia ser lido "Maria da Silva possui examinador Dr. João Santos" ou, genericamente, "<subject> POSSUI <predicate> <object>".

Caso queiramos fornecer alguma informação a mais sobre as características do examinador deste recurso (paciente), a sentença seria alterada para do seguinte modo:

O indivíduo cujo nome é Dr. João Santos, e-mail <jsantos@saude.org.br>, é o examinador de Maria da Silva.

O objetivo desta sentença é tornar o valor da propriedade *Examinador* uma entidade estruturada. Em RDF, uma entidade é representada por um outro recurso. A sentença acima não fornece um nome para este recurso, isto é anônimo, então no gráfico abaixo ele será representado por um oval vazio:

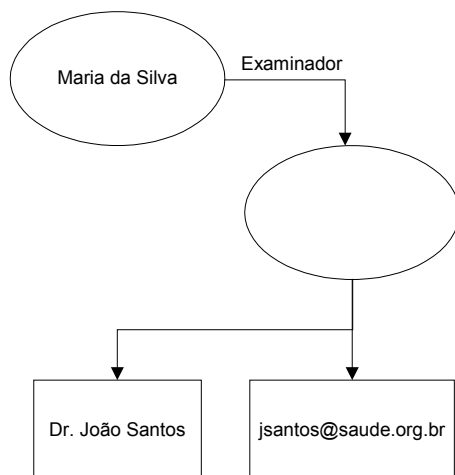


Figura 2 – Propriedade com valor estruturado.

Este diagrama poderia ser interpretado como "Maria da Silva possui examinador alguém e alguém possui nome Dr. João Santos e e-mail jsantos@saude.org.br".

A entidade estruturada do exemplo anterior pode também ser definida com um único identificador. A escolha do identificador é feita pelo criador da aplicação. Evoluindo com o exemplo, caso o código CRM seja usado como identificador único para um recurso "pessoa", como por exemplo "CRM 85740/PR". Reescrevendo as duas sentenças temos: *O indivíduo que possui o CRM 85740/PR é chamado Dr. João Santos e possui o endereço de e-mail jsantos@saude.org.br. A paciente Maria da Silva foi examinada por este indivíduo.*

O modelo RDF para estas sentenças é:

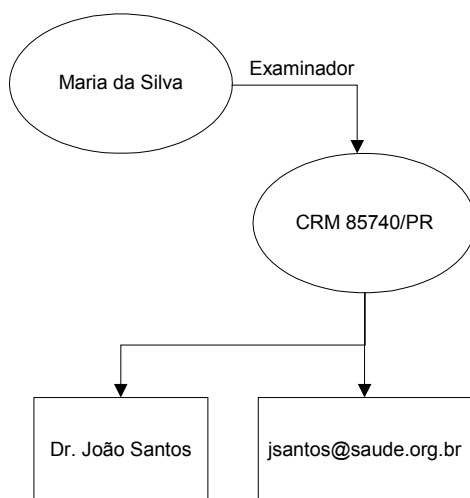


Figura 3 – Valor estruturado com identificador.

É válido observar que este diagrama é idêntico ao anterior com acréscimo do URI para o recurso anônimo anteriormente.

Este processo de representação pode ser aplicado para identificação de pessoas, exames e casos clínicos, entre outros.

6. Evolução do conhecimento

Segundo Berners-Lee, os esforços humanos balançam entre a efetividade de um pequeno grupo atuando independentemente e a necessidade de comunicação com a comunidade global. Este pequeno grupo pode inovar suas atividades, mas isto gera uma subcultura, cujos conceitos não serão compreendidos por outras pessoas. Coordenar ações em um grande grupo pode se tornar um processo lento e exigir um aumento na comunicação interna. O mundo trabalha entre estes dois extremos, cuja tendência é começar pequeno, fonte de uma idéia pessoal, e aumentar gradativamente. Neste ponto, a Inteligência Coletiva pode atuar como ambiente para armazenamento, divulgação e interpretação das atividades desenvolvidas por grupos pequenos ou grandes [6].

Um processo fundamental é a união de subculturas quando uma comunicação comum e mais abrangente é necessária. Eventualmente, dois grupos independentes desenvolvem conceitos muito similares e definir uma relação entre eles traria grandes benefícios.

A Web Semântica, ao lançar todo conceito simples através de uma URI, permite a expressão de novos conceitos com o mínimo esforço. Esta linguagem lógica e unificada habilitará estes conceitos a se unirem, progressivamente, através da Internet. Esta estrutura permitirá que softwares especialistas (agentes) possam realizar análises significativas do conhecimento e trabalhos realizados pela humanidade, provendo uma nova classe de ferramentas com as quais poderemos viver, trabalhar e aprender juntos.

A Web Semântica pode atuar no processo de aumento da qualidade de informação sobre saúde distribuída na Web e tornar positivo o seu impacto na saúde do público que a acessa. Bem como, servir de ferramenta para armazenamento de informação para o processo de decisão médica, que são baseadas em uma variedade de informações médicas e não-médicas obtidas de diversas fontes: histórico do paciente, status sócio-econômico, preferências do paciente, tipo e estágio da doença, e são influenciados pela experiência pessoal, conhecimento e disponibilidade, levando em consideração a possibilidade de recursos restritos [8].

7. Metadados e Ontologias

Na tomada da decisão médica, o RDF teria a função de definição da sintaxe, da gramática e da infraestrutura, aliada a ontologia médica, onde os conceitos ou vocabulários especialistas e fontes de descrição de dados permitem a busca de sinônimos e conceitos correlatos.

Na prática, isso significa a mescla dos metadados com as ontologias especializadas, pois os recursos de Web Semântica precisam ser marcados com os metadados ou indexados em um banco usando esses metadados. Assim, conceitos explícitos poderiam ser mapeados para uma ontologia, ou seja, terminologia clínica ou classificação, ou uma agregação de ontologias que permitiria agentes/sistemas de busca baseados em Web Semântica inferirem significados implícitos, os quais não foram diretamente mencionados nem nos recursos, nem nos metadados [9].

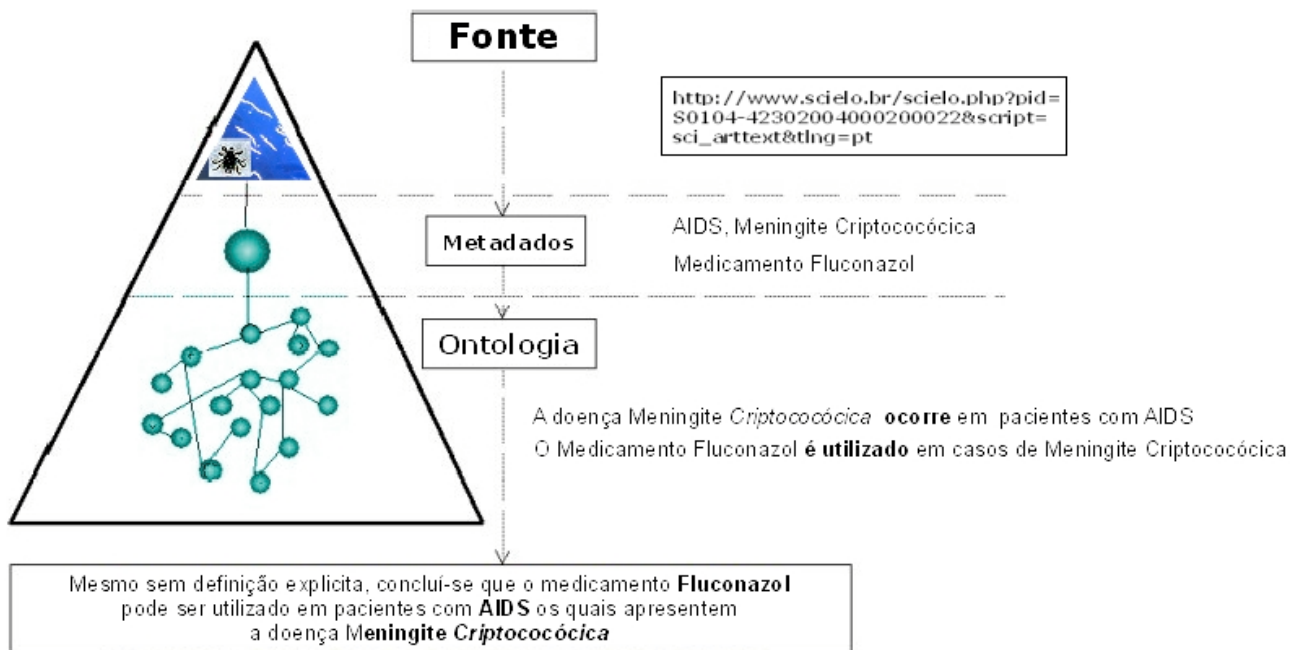


Figura 4 – Relação da fonte com metadados e ontologia

O metadado sozinho não é suficiente para a busca bem sucedida desta fonte. Neste exemplo, mesmo sem a fonte ou o metadado mencionar que o ‘medicamento Fluconazol é utilizado em pacientes com Aids e Meningite Criptocócica’, em uma busca assistida por ‘medicamento Fluconazol no tratamento da AIDS’ irá encontrar a URL informada na fonte.

8. Projetos

A busca por mais recursos da Internet atual requer soluções de maior qualidade para problemas de interação entre as fontes, chamadas de interoperabilidade, integração e colaboração. A solução escalável envolve o aperfeiçoamento das interações automáticas as quais só podem ocorrer com o acesso a dispositivos melhorados de “entendimento” da internet, através do desenvolvimento de agentes (softwares especialistas), que saibam trabalhar com essas informações.

Buscando desenvolver modelos de intercâmbio entre os dados armazenados por diferentes sistemas de tomada de decisão, o *Dublin Core Metadata Project* (<http://www.dublincore.org/>), busca promover uma ampla adoção dos padrões de metadados e vocabulários da Web Semântica, para o desenvolvimento de dados especializados, a serem utilizados por sistemas inteligentes para descoberta de informações.

Projetos de Web Semântica, como o *MedCircle* (<http://www.medcircle.info/>), procuram desenvolver e promover tecnologias capazes de permitir o acesso a informação confiável sobre saúde na *Web* e estabelecer uma rede global de informação sobre saúde com qualidade. Buscando diminuir o trabalho das autoridades para: regular, controlar, censurar ou aprovar informação, fornecedores de informação e websites. O *MedCircle* é um sistema descentralizado, mas procura atuar como integrador entre profissionais de saúde e pacientes, fornecendo informações com qualidade, através da Web Semântica [10].

O projeto MEDTAG utiliza a linguagem NLP (*natural language processing*) na aplicação de *tags* à informação, inserindo regras linguísticas para uma melhor compreensão e evolução das consultas a serem realizadas sobre a informação transformada [11].

9. Resultados

Essencialmente, a medicina baseada em informação busca a informação correta no formato correto para o indivíduo correto no tempo correto. Atualmente, pesquisadores têm se utilizado de grandes volumes de dados necessários para o desenvolvimento de tratamentos. Neste ponto, a Web Semântica surge como uma possibilidade de convergência dos avanços científicos, das melhores práticas de medicina e do gerenciamento de dados individuais do paciente, permitindo melhorar os resultados e reduzir erros.

Os esforços efetuados por organizações como a W3C (*World Wide Web Consortium*) no desenvolvimento da Web Semântica levam a aceleração da velocidade de disseminação do conhecimento, pois a busca por conceitos se torna mais simples, no momento em que se removem os resultados indesejados e fora do contexto. Nesse ambiente de conhecimento disseminado, o conceito de Inteligência Coletiva apresenta também seu crescimento, devido à farta apresentação de conteúdo.

10. Conclusão

O “mapeamento” do conhecimento e sua distribuição através de um framework comum, como o RDF, permitirão o desenvolvimento de ferramentas que podem interpretar a massa de dados externa ao conhecimento dos grupos que a utilizarem e, com isso, levar em consideração um número quase ilimitado de fatores importantes para a tomada de decisão com qualidade. Esta evolução possibilita: A aceleração da pesquisa e desenvolvimento de tratamentos e identificação de falhas mais rapidamente; compilar conhecimento do crescente número de dados clínico, mercadológico e científico; utilizar as tecnologias correntes, dados e aplicações legadas através deste *framework* aberto e fornecer dados para a tomada de decisão diagnóstica e clínica.

Como Peter Druker mencionou a Revolução da Informação é na verdade a Revolução do Conhecimento, onde o software representa a reorganização do trabalho tradicional, baseada em séculos de experiência, por meio da aplicação do conhecimento e, principalmente, da análise sistemática e lógica. O segredo não é a eletrônica, mas sim a ciência cognitiva. Cabe ao homem descobrir que o próximo passo está no compartilhamento de suas idéias e opiniões, buscando seu desenvolvimento e das comunidades virtuais que o cercam [1].

11. Referências

- [1]Drucker, P. (1998), “Além da Revolução da informação”, Revista HSM Management, n. 18, ano 3.
- [2]Lévy, P. (1995) *Intelligence Coletive*. Paris: Éditions La Découverte,.
- [3]Massad, A. (2002) “A caminho da inteligência coletiva.”
[http://www.lainsignia.org/2002/novembre/cyt_008.htm]. Acesso em 30 de novembro de 2002.
- [4]Heylighen, F. (1997) *Towards a Global Brain. Integrating Individuals into the World-Wide Electronic Network*. Free University of Brussels, Brussels, Belgium.
- [5]Crawford, R. (1994) *Na era do capital humano: o talento a inteligência e o conhecimento como forças econômicas, seu impacto nas empresas e nas decisões de investimento*. São Paulo, Atlas.
- [6]Berners-Lee, T.; Hendler, J. e Lassila, O. (2001) “The Semantic Web”.
[http://www.scientificamerican.com/print_version.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21].

- [7]Nagel, E.; Newman, J. R. (1956) *Gödel's Proof*. New York: Simon e Schuster.
- [8]Hölzer, S., Tafazzoli, A. G. e Dudeck, J. (1998) "Data Warehousing – Analytical Processing of all available information". Scherrer JR.
- [9]Network Inference Website – Technology: Semantic Web.
<http://networkinference.semanticweb.org/technology.asp?id=46&meu=2>
- [10]Eysenbach, G. Köhler, Ch., Roth-Berghofer, T., Mayer, M.A., Fiene, M. e Darmoni, S. (2003) "MedCircle – Collaboration for Internet Rating, Certification, Labeling and Evaluation of Health Information on the (Semantic) World-Wide-Web".
- [11]Ruch, P., Bouillon, P., Baud, R. H., Rassinoux, M. e Scherrer, J. (1999) "MEDTAG: Tag-like Semantics for Medical Document Indexing".