

Uma ontologia comum para a integração de bases de informações e conhecimento sobre ciência e tecnologia

Roberto Carlos dos Santos Pacheco

Doutor em Engenharia de Produção
Coordenador do Grupo STELA – desenvolvedor do sistema de Banco de Teses e Dissertações do PPGE/UFSC
E-mail: pacheco@eps.ufsc.br

Vinícius Medina Kern

Doutor em Engenharia de Produção
Coordenador do Núcleo de Mídia Científica do PPGE/UFSC e professor da Univali São José
E-mail: kern@eps.ufsc.br

Resumo

A produção de documentos científicos cresce em ritmo acelerado, da mesma forma que a demanda por busca, verificação, recuperação e análise destes documentos. Esta demanda não pode ser atendida satisfatoriamente pelas ferramentas disponíveis. Os documentos e indicadores de atividade científica requeridos por pesquisadores, bibliotecas e outros agentes só podem ser obtidos se houver a integração de sistemas de informações. Este artigo descreve uma iniciativa brasileira que potencializa a integração de sistemas de informações sobre ciência e tecnologia: a Linguagem de Marcação da Plataforma Lattes (LMPL), definida pelo consenso de peritos de várias instituições de ensino superior. Apresenta-se o problema da integração de sistemas. Discute-se também a iniciativa de criar uma ontologia comum para a informação sobre ciência e tecnologia. São aventadas possibilidades presentes e futuras para os sistemas de informações sobre ciência e tecnologia a partir da disponibilidade da LMPL.

Palavras-chave

Ontologia; Metadados; Linguagens de marcação; Integração de sistemas; Informetria; Bibliometria; Bases de conhecimento; Gestão do conhecimento.

A common ontology for information and knowledge integration in science and technology

Abstract

Scientific publication has maintained an accelerated grow rate, as well as the demand for document search, verification, retrieval, and analysis. This demand cannot be adequately satisfied using current tools. Information systems integration can realize the demands of researchers, libraries, and other agents. This paper describes a Brazilian initiative aimed at the integration of science and technology information systems: the Linguagem de Marcação da Plataforma Lattes (LMPL). LMPL was built using the consensus of experts of several educational institutions. The problem of information systems integration is presented, together with an account of the LMPL initiative. Present and future possibilities for science and technology systems are considered.

Keywords

Ontology; Metadata; Markup languages; Systems integration; Informetry; Bibliometry; Knowledgebases; Knowledge management.

INTRODUÇÃO

A expansão da Internet e a tendência mundial a ampliar a formação acadêmica têm permitido um avanço constante e acelerado no número de publicações científicas. Fazer sentido desta produção, no entanto, não é tarefa simples. Há quem considere que o excesso de informação e a multiplicidade e fugacidade de formatos digitais venham dificultar muito a manutenção de documentos em meio eletrônico. Lukesh¹ considera que um registro em papel tem mais chances de durar um século do que um registro digital.

A demanda por ferramentas de busca, verificação, recuperação e análise de documentos também aumenta. Sistemas de busca na Internet, bem-sucedidos para aplicações genéricas, inspiram pesquisadores e instituições a pleitear facilidades de acesso e ferramentas de análise até este momento inexistentes para o domínio específico da academia.

Kondo² discorre sobre as razões para desenvolver indicadores estratégicos em ciência e tecnologia. Os documentos e indicadores de atividade científica requeridos por pesquisadores, bibliotecas e outros agentes só podem ser obtidos se houver a integração de informações presentes em sistemas heterogêneos. De outra forma, a comunicação científica permanecerá como está: truncada, repleta de inconsistências, requerendo retrabalho, restringindo a colaboração entre parceiros.

Este artigo descreve uma iniciativa brasileira que potencializa a integração de sistemas de informações sobre ciência e tecnologia: a construção, pelo consenso de peritos de várias instituições, da Linguagem de Marcação da Plataforma Lattes (LMPL). Esta linguagem estabelece uma gramática e um vocabulário comuns no âmbito da Plataforma Lattes de informações em ciência e tecnologia. A Plataforma Lattes atende a pesquisadores, instituições de ensino e pesquisa e agências governamentais.

A gramática e vocabulário comuns são também conhecidos como **ontologia**. Esta conceitualização formalizada potencializa a integração de sistemas de informações sobre ciência e tecnologia, favorecendo um avanço substancial em quantidade e qualidade nas várias disciplinas de indicadores de atividade científica³: bibliometria, cienciometria e informetria.

Bibliometria é o estudo quantitativo da produção, disseminação e circulação (empréstimos) de documentos científicos, incluindo autores e usuários. **Cienciometria** é o estudo quantitativo das atividades científicas com o objetivo de identificar domínios de interesse (áreas, assuntos, disciplinas) e compreender a comunicação entre cientistas. **Informetria** é o estudo quantitativo da informação em qualquer formato, inclusive palavras e outras peças de informação com granularidade mais fina do que os objetos de interesse da bibliometria e da cienciometria. A figura 1 representa graficamente a abrangência dos objetos de estudo destas três disciplinas.

A próxima seção discute a integração de informações e sistemas, uma necessidade das três disciplinas. A seção seguinte descreve a LMPL e sua construção pelo consenso de peritos de várias instituições. As possibilidades presentes e futuras para a atividade científica a partir da disponibilidade da LMPL são aventadas.

Integração de informações e sistemas

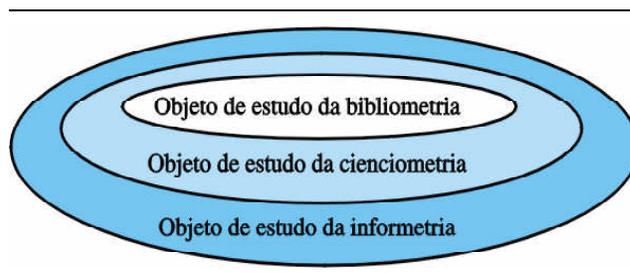
A expansão da tecnologia computacional tem acontecido em ritmo acelerado desde o surgimento e a disseminação dos computadores nas grandes empresas, seguidos pela popularização dos computadores pessoais, até o crescimento da Internet. Em poucas décadas, chegou-se a um ponto no qual as organizações mantêm sistemas legados de grande porte, utilizando tecnologia ultrapassada, ao mesmo tempo em que desenvolvem e mantêm sistemas tecnologicamente atualizados.

O motivo para a manutenção de sistemas baseados em tecnologia arcaica, com custos altos de manutenção, é que a migração para uma nova tecnologia quase sempre precisa ser feita abruptamente, sem a interrupção de serviços. No entanto, a decisão de migrar para uma nova tecnologia envolve uma manobra custosa e arriscada para a organização.

A convivência de sistemas não integrados acarreta uma série de problemas, entre eles: redundância de dados e inconsistência de dados redundantes; dispêndio de recursos que não agregam valor ao negócio para transferir informação de um sistema para outro; perda semântica na tradução, redigitação de dados. Sistemas que usam

FIGURA 1

Abrangência dos objetos de estudo da bibliometria, cienciometria e informetria



tecnologia arcaica, como sistemas de informações não baseados em sistemas de gerência de banco de dados, estão ainda sujeitos a deficiências de controle e a conseqüente perda de integridade da informação. A migração de um sistema arcaico para uma nova tecnologia é, portanto, uma necessidade que se impõe.

A integração dos sistemas legados com os sistemas novos poderia atenuar o risco de uma migração abrupta. Se os sistemas legados pudessem trocar dados com os sistemas novos, o desenvolvimento de testes de bases de dados e aplicativos para substituir as soluções arcaicas poderia ser feito de forma mais pausada, com tempo adequado para testes, diminuindo o risco da mudança abrupta. Por outro lado, sistemas legados integrados e que atendem bem às necessidades não precisariam ser reimplementados com tecnologia atualizada.

Em uma situação ideal, diversos aplicativos poderiam **interoperar**, isto é, trocar dados e chamadas a rotinas de forma transparente para o usuário. A interoperabilidade é definida como “a interconexão efetiva de diferentes sistemas de computador, bancos de dados ou redes com o fim de apoiar a computação distribuída e/ou o intercâmbio de dados”⁴.

Apesar de ser uma característica desejada nos sistemas, a interoperabilidade plena dificilmente é atingida – mesmo quando se trata de sistemas relativamente atuais. A interoperabilidade costuma ser impedida por barreiras de plataformas diferentes de *hardware* e também de *software*, incluindo sistemas operacionais, paradigmas de programação e de modelos de dados.

Uma das barreiras à interoperabilidade é a incompatibilidade dos modelos de dados subjacentes às aplicações. Na área de desenho, por exemplo, é possível representar um arco circular de nove formas distintas⁵. Dois aplicativos de desenho poderão trocar dados sobre arcos circulares somente se usarem a mesma forma de representação ou se dispuserem de uma forma explícita de

tradução entre as duas formas distintas. Kent assinalou este lapso existente entre a realidade e suas formas de representação abstrata da seguinte forma:

“Uma mensagem para os fazedores de mapas: as estradas não são pintadas de vermelho, os rios não têm linhas de limite territorial correndo pelo meio, e não se podem ver curvas de nível em uma montanha”⁶.

Na área da ciência da informação, o problema de cadastro e indexação de nomes de autores de publicações pode ser resolvido por vários aplicativos não interoperáveis e baseados em diferentes concepções: enquanto um aplicativo toma o último nome como indexador de autor, outro admite isto e mais uma extensão como Jr., Filho, etc. Outro, ainda, admite optar pela forma espanhola, na qual o “último nome” é o penúltimo (por exemplo: GARCÍA MÁRQUEZ, Gabriel), enquanto um outro aplicativo admite que o usuário escolha o modo de fazer a indexação. Uma biblioteca que dispuser destes vários aplicativos, cada um mantendo bases de dados próprias, terá enorme dificuldade em integrá-los.

A incompatibilidade de abstrações de dados é um problema crucial na área de engenharia. STEP – *standard for the exchange of product model data*, ISO 10303⁷, é o maior esforço para a construção de um modelo de dados já empreendido. Visa à interoperabilidade de aplicações de engenharia, mas vem enfrentando problemas desde o início do empreendimento, em 1984.

Um dos problemas detectados diz respeito ao conceito de **produto**. Embora seja o conceito mais importante do padrão, dois documentos de STEP tratam produto como coisas distintas: a coisa⁷ e a informação sobre a coisa⁸, o que pode impedir a interoperabilidade de aplicações baseadas nestes dois documentos – justamente o tipo de problema que o padrão se propõe a resolver. Críticos⁹ apontam a falta, em STEP, de uma modelagem **ontológica** do domínio “aplicações de engenharia”.

A construção de ontologias comuns tem sido proposta como abordagem promissora para a interoperabilidade de sistemas^{9, 10, 11, 12}. Ontologia é usualmente definida como “a especificação explícita de uma conceitualização”¹³. Uma ontologia comum é uma formalização compartilhada de um certo domínio de aplicação. Assim, por exemplo, uma formalização de conceitos sobre informações em ciência e tecnologia poderia permitir que diversos aplicativos deste domínio compartilhassem um vocabulário comum sobre o assunto. Esta é a característica da Linguagem de Marcação da Plataforma Lattes (LMPL), discutida a seguir.

A LINGUAGEM DE MARCAÇÃO DA PLATAFORMA LATTES

A Plataforma Lattes¹⁴ de sistemas de informações em ciência e tecnologia surgiu a partir da necessidade de integração de informações mantidas por CNPq, Capes, Fapesp, Finep e outros sistemas do Ministério de Ciência e Tecnologia. Atendendo a uma reivindicação antiga da comunidade científica, o sistema Currículo Lattes permitiu integrar informações de aplicativos não integrados que, apenas no CNPq, envolviam Bcurr, minicurriculo e Genos – precursor do Currículo Lattes.

O Currículo Lattes foi integrado a várias outras aplicações de ciência e tecnologia, tais como o Diretório de Grupos de Pesquisa. Esta construção de sistemas integrados resultou na Plataforma Lattes, usada por pesquisadores, grupos, instituições de ensino e pesquisa, CNPq, Capes e outros órgãos governamentais.

Alguns dos problemas de qualidade da informação que foram equacionados para que a Plataforma Lattes pudesse existir foram¹⁵ desintegração dos sistemas existentes, conteúdo fundamentado na necessidade operacional (e não estratégico-gerencial), erros nas informações (principalmente por falta de controle de preenchimento e suporte à correção pelo usuário), redundância nas informações, falta de visibilidade (relatórios não configuráveis e falta de um sumário de indicadores de produção) e o equilíbrio difícil entre tabulações excessivas e liberalidade de cadastro (imposição de muitos preenchimentos pelo usuários ou a liberação para o preenchimento voluntário).

A modificação do modelo de dados do sistema de currículo Genos, do CNPq, para compatibilizá-lo com as necessidades de informações da Capes, segundo Barcia¹⁵, representou um desafio significativo. Alguns módulos tinham interseção quanto às necessidades de informação (formação acadêmica, atividades profissionais e produção), mas com diferenças de concepção para as duas agências. Este esforço de compatibilização foi, de certa forma, precursor da LMPL, uma ontologia comum para informações de ciência e tecnologia desenvolvida com o apoio de nove universidades (PUC-PR, UFBA, UFPE, UFRGS, UFRJ, UFRN, UFSC, Unicamp e USP).

O empreendimento foi motivado pela necessidade de ampliar os horizontes da Plataforma Lattes para atender a uma demanda das instituições de ensino e pesquisa. Cada instituição participante havia contactado o CNPq e solicitado que as informações existentes na Plataforma fossem disponibilizadas para as instituições, de forma que pudessem alimentar seus bancos de dados e usá-los em funções de cadastro, análise e auditoria.

Em qualquer sistema de informações há usuários com necessidades e percepções específicas sobre as informações armazenadas. No entanto, para que possam usar o mesmo sistema, é preciso um modelo definindo as estruturas de dados e as restrições aplicáveis ao conteúdo, qualquer que seja a aplicação. Este modelo é conhecido como metabanco de dados ou banco de **metadados**. O banco de metadados estabelece uma estrutura comum, integrada, para que vários usuários possam usar o mesmo banco de dados.

A integração de dados implementada em um banco de metadados é, muitas vezes, insuficiente para suprir as necessidades de informação de uma organização. Quando há necessidade de trocar ou compartilhar informações entre vários sistemas, é preciso **traduzir** as informações conforme as **estruturas** e regras presentes nos modelos destes sistemas. Neste ponto surgem barreiras tecnológicas que podem impedir a comunicação. A incompatibilidade de metadados é um exemplo de barreira tecnológica à integração dos sistemas.

A LMPL foi escrita em XML (*extendable markup language*), uma linguagem de marcação. Bax¹⁶ recentemente delineou as principais características de linguagens de marcação. A linguagem HTML (*hypertext markup language*) foi projetada para marcar documentos com conteúdo fixo. Em contraste, a linguagem XML foi projetada para a marcação de documentos com conteúdo variável. Desta forma, pode-se encarar um banco de dados como um texto de conteúdo variável ao qual se aplica uma marcação que segue regras gramaticais bem definidas.

Em XML 1.0¹⁷, um documento XML bem formado deve ajustar-se a uma estrutura sintática bem definida. Assim é documentada a forma de uso de *tags* e restrições aplicáveis a um documento XML. Para que um documento XML seja também válido para intercâmbio, deve aderir a uma estrutura sintática definida em um documento chamado DTD (*document type definition*).

FIGURA 1 Diagrama esquemático da representação de informações sobre currículos segundo a LMPL¹⁸

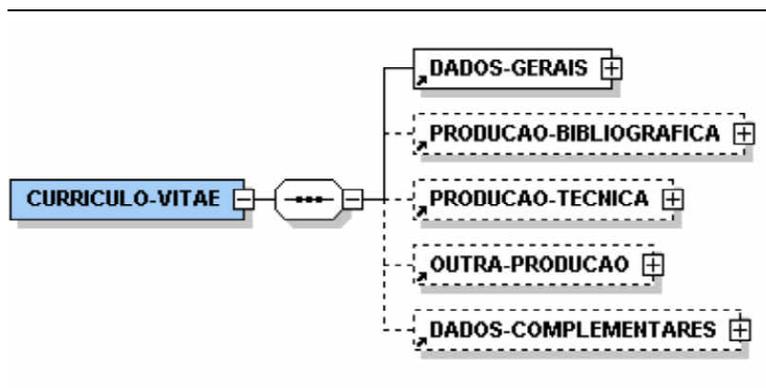
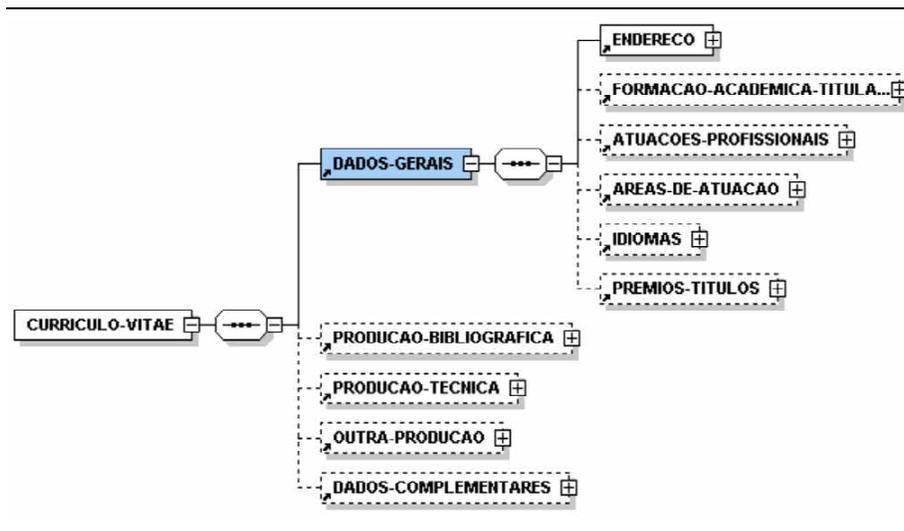


FIGURA 2 Detalhamento do diagrama da figura 1, mostrando a organização dos dados gerais de qualquer currículo¹⁸



A figura 1 ilustra, de forma esquemática, a estrutura da informação sobre currículos de pesquisadores conforme a LMPL. Segundo o esquema, as informações presentes em qualquer *curriculum vitae* são categorizadas em dados gerais, produção bibliográfica, produção técnica, outra produção e dados complementares. Esta definição é resultado do consenso dos membros da comunidade LMPL.

A figura 2 apresenta um esquema mais detalhado do que o da figura 1, mostrando como a informação curricular referente a “dados gerais” é organizada. Os dados gerais de cada pesquisador são organizados da seguinte forma: endereço, formação acadêmica, atuações profissionais, áreas de atuação, domínio de idiomas e prêmios e títulos honoríficos.

A LMPL utiliza sucessivos níveis de detalhamento como os apresentados nas figuras 1 e 2 para representar a estrutura da informação curricular de pesquisadores. Esta visão esquemática é uma abstração dos documentos efetivamente usados no intercâmbio e compartilhamento de informações: DTDs e arquivos de conteúdo XML. A figura 3 apresenta parte do código DTD que define a sintaxe específica a ser respeitada em arquivos XML que contêm informações sobre currículos. A figura 4 mostra um trecho de arquivo XML que adere à gramática da LMPL, documentada no DTD ilustrado nas figuras 1 a 3.

A especificação contida na LMPL favorece a escrita de componentes de *software* reusáveis para a geração e interpretação dos arquivos XML que contêm informações sobre ciência e tecnologia. A partir da LMPL, não apenas as agências podem acessar a informação de seu interesse alimentada por pesquisadores e gestores, mas também a comunidade de instituições de ensino e pesquisa, parceiros na construção da LMPL. A próxima seção discute este aspecto.

PRESENTE E FUTURO A PARTIR DA LMPL

No momento presente, a definição de um documento que representa o consenso de uma comunidade sobre conceitos e restrições aplicáveis à informação sobre ciência e tecnologia garante que será possível incluir a informação relevante nas bases de dados com representação e controle adequados. Os aplicativos que atualmente fazem parte da Plataforma Lattes aderem a esta definição, bem como os que vierem a ser desenvolvidos.

Aplicativos externos à Plataforma, mas desenvolvidos por participantes da comunidade que conhecem a gramática da LMPL, poderão interoperar com aplicativos pertencentes à Plataforma. Desta forma, cada participante da comunidade pode desenvolver aplicativos para atender a necessidades muito específicas, fazendo uso das informações inseridas via Plataforma Lattes. Pesquisadores poderão desenvolver investigações com base em uma série extensa de informações confiáveis. Um pesquisador que troca de instituição pode transferir de forma simples para a nova instituição toda a informação sobre sua carreira, cuja manutenção esteve sob seu cuidado. Direções, coordenadorias, reitorias e pró-reitorias, secretarias e ministérios poderão obter a informação de que necessitam e formatá-la conforme suas necessidades.

Cada membro da comunidade LMPL poderá, também, compartilhar o acervo que detém com outros membros da comunidade. Isto inclui a possibilidade de compartilhamento de conhecimento e não apenas de informação, usando tecnologia de sistemas de informações e

FIGURA 3
Trecho do documento DTD que define a gramática da LMPL¹⁸

```
<!ELEMENT CURRICULO-VITAE (
    DADOS-GERAIS,
    PRODUCAO-BIBLIOGRAFICA?,
    PRODUCAO-TECNICA?,
    OUTRA-PRODUCAO?,
    DADOS-COMPLEMENTARES?)>

<!ELEMENT DADOS-GERAIS (
    ENDERECO,
    FORMACAO-ACADEMICA-TITULACAO?,
    ATUACOES-PROFISSIONAIS?,
    AREAS-DE-ATUACAO?,
    IDIOMAS?,
    PREMIOS-TITULOS?)>
```

FIGURA 4
Trecho de documento XML (bem formado e conforme à LMPL) que contém as informações de um currículo de pesquisador¹⁸

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
- <CURRICULO-VITAE SISTEMA-ORIGEM-
XML="LATTES_OFFLINE" >
+ <DADOS-GERAIS NOME-COMPLETO="Fulano de
Tal" NOME-EM-
CITACOES-BIBLIOGRAFICAS="TAL, F."
NACIONALIDADE="B"
CPF="98765432109" PAIS-DE-
NASCIMENTO="Brasil" UF-
NASCIMENTO="DF" CIDADE-
NASCIMENTO="Brasília" DATA-
NASCIMENTO="19500716" SEXO="MASCULINO"
NUMERO-
IDENTIDADE="1234567890" ORGAO-
EMISSOR="SSP" UF-
ORGAO-EMISSOR="DF" DATA-DE-
EMISSAO="20001231" NOME-
DO-PAI="José Fulano de Tal" NOME-DA-
MAE="Sicrana
dos Anzóis de Tal" PERMISSAO-DE-
DIVULGACAO="SIM" >
+ <PRODUCAO-BIBLIOGRAFICA>
+ <PRODUCAO-TECNICA>
+ <OUTRA-PRODUCAO>
+ <DADOS-COMPLEMENTARES>
</CURRICULO-VITAE>
```

inteligência artificial, como já fazem alguns aplicativos da Plataforma Lattes através de *data warehousing*, raciocínio baseado em casos, mineração de dados etc. Um destes aplicativos é o sistema de informações do Diretório de Grupos de Pesquisa, Versão 4.1¹⁹, que usa a tecnologia de *data warehousing* para possibilitar a solicitação de consultas com nível de granularidade grosso (informações de nível estratégico-gerencial, deduzidas a partir de uma grande quantidade de informações de nível operacional). Este tipo de consulta não é possível com o uso de tecnologia convencional de bancos de dados devido ao custo excessivo de processamento.

A perspectiva de intercambiar informações sobre a produção científica e realizar consultas desde o nível operacional até os níveis gerencial e estratégico favorece o avanço na construção de indicadores de pesquisa abrangentes e de qualidade. Cientistas da informação terão, a partir da disponibilidade de aplicativos baseados na LMPL, condições de planejar e executar a constituição de seus indicadores de interesse em bibliometria, cienciométrica e informetria.

As possibilidades de intercâmbio, compartilhamento e interoperabilidade citadas até aqui beneficiam os membros da comunidade LMPL (atuais e futuros) – aqueles que conhecem sua definição. Uma questão a se colocar é a possibilidade de disponibilizar o acesso à Plataforma Lattes para qualquer ator externo, obviamente com níveis diferenciados de privacidade e propriedade de informações. Esta possibilidade pode ser considerada interessante na medida em que outras comunidades também podem estar dispostas a franquear o acesso a suas informações, com benefício mútuo.

A LMPL viabiliza o intercâmbio de dados e o acesso interoperável no que tange ao modelo de dados subjacente aos aplicativos, pois contém uma gramática que define como representar informação sobre ciência e tecnologia. Escrita em XML, permite o intercâmbio de dados na forma de texto com marcação. Por ser baseada em sintaxe, e não

FIGURA 5

Dois estruturas semanticamente compatíveis e sintaticamente incompatíveis (adaptadas a partir de Decker *et alii*¹¹)

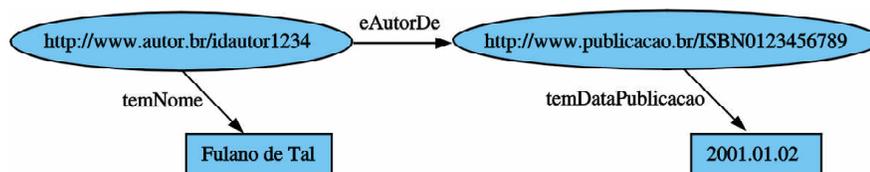
```

<class-def>
  <name>branch</name>
  <slot-constraint>
    <name>is-part-of</name>
    <has-value>tree</has-value>
  </slot-constraint>
</class-def>

<class-def>
  <class name="branch"/>
  <slot-constraint>
    <slot name="is-part-of"/>
    <has-value>
      <class name="tree"/>
    </has-value>
  </slot-constraint>
</class-def>
    
```

FIGURA 6

Grafo RDF que expressa a associação entre os objetos 'http://www.autor.br/idautor1234' (com valor "Fulano de Tal" para o atributo 'temNome') e 'http://www.publicacao.br/ISBN0123456789' (com valor '2001.01.02' para o atributo 'temDataPublicacao')



em semântica, a LMPL só permite a interoperabilidade de aplicações que a conhecem previamente. Isto atende à demanda atual, mas não prevê a interoperabilidade com aplicações externas à comunidade LMPL, o que é uma tendência na Web¹¹.

A figura 5 ilustra duas definições em XML sintaticamente distintas, embora semanticamente equivalentes: definem de formas diferentes a estrutura todo-parte "árvore-ramo". Aplicativos diferentes que manipulam dados baseados nestas ontologias só poderão interoperar se houver um mecanismo de tradução entre as duas ontologias. No entanto, as duas construções são semanticamente equivalentes e poderiam interoperar se compartilhassem uma ontologia comum baseada em semântica.

Conforme Decker *et alii*¹¹, a Web foi viabilizada até hoje por vários padrões projetados para processamento direto por seres humanos. Entretanto, a habilitação de serviços inteligentes como *brokers* (intermediadores) de informação, agentes de busca e filtros de informação requer o processamento por máquinas, com mais funcionalidade e interoperabilidade do que os serviços disponíveis atualmente.

Os autores propõem a utilização da linguagem RDF (*resource description framework*)²⁰, baseada em semântica, para a escrita de ontologias. RDF tem como bloco básico a tripla objeto-atributo-valor, ou seja, um objeto O tem um atributo A cujo valor é V. A figura 6 mostra um grafo RDF que expressa a autoria de um livro por um autor.

Na figura 6, os objetos são representados por elipses, os valores por retângulos e os atributos por setas. Objetos e valores são intercambiáveis em RDF, isto é, um valor em uma tripla pode ser objeto em outra. Isto permite o encadeamento de setas. Assim, pode-se obter o aninhamento de grafos e, desta forma, expressar dúvida ou apoio a sentenças

criadas por outros. Ainda, é possível indicar que um objeto é de um certo tipo, o que permite estabelecer uma ligação relativa à sua definição. Na figura 7, (a) mostra a ligação da publicação “ISBN0123456789” com sua definição, contida em `rdf:type publicacao`. O item (b) mostra o código equivalente ao grafo da figura 6 na forma A(O,V).

Esta construção de ontologia baseada em semântica pode ser uma direção para a evolução da LMPL, se houver interesse no incremento do acesso interoperável, incluindo usuários e aplicativos externos à comunidade LMPL. Através da interoperabilidade semântica será possível ampliar as possibilidades de uso da Plataforma Lattes, que já incluem a interoperabilidade sintática aliada ao poder de expressar qualquer tipo de dados. Esta ampliação é compatível com a chamada *Semantic Web*^{11,21}, que propõe a expansão dos serviços disponíveis na Web, com a inclusão de serviços inteligentes com processamento automático.

CONSIDERAÇÕES FINAIS

Este artigo apresentou uma breve discussão sobre o problema da integração de sistemas de informações e descreveu a criação da LMPL, uma ontologia comum para informações sobre ciência e tecnologia. A LMPL potencializa a integração de sistemas e favorece o avanço da facilidade de acesso a conteúdo e indicadores científicos. As possibilidades de desenvolvimento a partir da disponibilidade da LMPL incluem a colaboração entre parceiros que fazem parte da comunidade LMPL ou mesmo parceiros externos, concorrendo para a realização do conceito de *Semantic Web*.

A LMPL permite integrar uma gama cada vez maior de informações sobre ciência e tecnologia. Este avanço tem o potencial de colaborar para minorar distorções na avaliação da produção científica. O estabelecimento de

FIGURA 7

(a) Associação de um objeto com o esquema que contém sua definição e (b) três triplas Atributo(Objeto,Valor) equivalentes ao grafo da figura 6 e que incluem um objeto cuja definição está em (a)

(a)	(b)
<code><rdf:Description about=</code>	<code>temNome</code>
<code>"http://www.publicacao.br/ISBN0123456789"></code>	<code>('http://www.autor.br/idautor1234',</code>
<code><rdf:type resource=</code>	<code>"Fulano de Tal")</code>
<code>"http://descricao.org.br/esquema/publicacao"></code>	<code>eAutorDe</code>
<code></rdf:Description></code>	<code>('http://www.autor.br/idautor1234',</code>
	<code>('http://www.publicacao.br/ISBN0123456789',</code>
	<code>temDataPublicacao</code>
	<code>('http://www.publicacao.br/ISBN0123456789',</code>
	<code>"2001.01.02")</code>

estratégias de política científica tem sido limitado pela “falta de bases de dados que permitam perceber a produção científica em um contexto amplo”²², problema que tende a ser suavizado com a disponibilidade de aplicações baseadas na LMPL.

O critério atualmente aceito para medir a qualidade da produção (quantidade de publicações em revistas indexadas) é reconhecidamente prejudicial para pesquisadores em algumas áreas do conhecimento, como, por exemplo, a computação, pois prejudica as carreiras dos pesquisadores e lesa a área²³. Nesta área há clara preferência pela publicação em conferências. Além disso, é importante considerar outros tipos de produção, como a produção de *software* conceitualmente inovador, não contemplada nas medições de publicações em periódicos indexados.

Por outro lado, mesmo se considerarmos o critério “publique ou morra” adequado, ignorar a produção científica não indexada é desperdiçar recursos preciosos. Estar na sombra, ou seja, ter relevância considerada menor, não é o mesmo que não existir. Aplicações sobre bases de dados baseadas na LMPL permitem manter informações mais amplas sobre a produção científica, como, por exemplo, a produção de *software* e a publicação não indexada. Isto favorece a busca de novas formas para avaliar a produção mais bem alinhadas com o interesse da comunidade.

A sociedade é beneficiada também quando há acesso facilitado à produção científica. A correlação entre disponibilidade *on-line* e número de citações de um trabalho está demonstrada²⁴. A LMPL tem um papel importante na facilitação do acesso, tanto no que diz respeito às facilidades de pesquisa quanto ao acesso ao conteúdo, por meio dos repositórios de informações e conteúdo providos de forma integrada pelas instituições participantes.

REFERÊNCIAS BIBLIOGRÁFICAS

1. LUKESH, S. S. E-mail and potential loss to future archives and scholarship or the dog that didn't bark. *First Monday*, v. 4, n.9, 1999. Disponível em: <http://firstmonday.org/issues/issue4_9/luke/index.html>. Acesso em: 20 set. 2001.
2. KONDO, E. K. Desenvolvendo indicadores estratégicos em ciência e tecnologia: as principais questões. *Ciência da Informação*, Brasília, v. 27, n. 2, p. 128-133, 1998.
3. MACIAS-CHAPULA, C. A. O papel da informetria e da cienciométrica e sua perspectiva nacional e internacional. *Ciência da Informação*, Brasília, v. 27, n. 2, p. 134-140, 1998.
4. FEDERAL COORDINATING COUNCIL FOR SCIENCE (Washington, D.C.). Office of Science and Technology Policy. *High performance computing & communications: toward a national information infrastructure*. Washington D.C., 1994. 176 p. (Committee on Physical, Mathematical, and Engineering Sciences Report).
5. OWEN, J. *STEP: an introduction*. Winchester, UK, Information Geometers, 1993. 143 p.
6. KENT, W. *Data and reality*. [S. l.] : North-Holland, 1978.
7. INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. (Genève). *ISO 10303-1: product data representation and exchange - part 1: overview and fundamental principles*. Paris : Committee Draft, 1992.
8. INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. (Genève). *ISO 10303-41: product data representation and exchange - part 41: integrated resources: fundamentals of product description and support*. Paris : ISO Draft International Standard, 1994.
9. MEIS, E.; OSTERMAYER, R. *Recommendations to the STEP committee*. Document KACTUS-01-RPK-D007 v. 1.1, 1996. 24 p. Disponível em: <<http://www.swi.psy.uva.nl/projects/NewKACTUS/Reports.html>>.
10. BENJAMINS, R. *et al.* Wondertools? A comparative study of ontological engineering tools. In: INTERNATIONAL WORKSHOP ON KNOWLEDGE ACQUISITION, MODELING, AND MANAGEMENT, 12., Banff, Canada, 1999. , Banff, Canada. *Proceedings...* Banff, Canada : [s. n.], 1999.
11. DECKER, S. *et al.* The semantic web: the roles of XML and RDF. *IEEE Expert*, v. 15, n. 3, 2000.
12. STAAB, S.; MAEDCHE, A. Ontology engineering beyond the modeling of concepts and relations. In: ECAI'2000 WORKSHOP ON APPLICATION OF ONTOLOGIES AND PROBLEM-SOLVING METHODS, 2000, Amsterdam. *Proceedings ...* [S. l.] : IOS Press, 2000.
13. GRUBER, T. Towards principles for the design of ontologies used for knowledge sharing. *International Journal of Human and Computer Studies*, v. 43, n. 5/6, p. 907-928, 1994.
14. PLATAFORMA lattes. Disponível em: <<http://lattes.cnpq.br:8888/plataformalattes/index.jsp>>. Acesso em: 18 nov. 2001.
15. BARCIA, R. M. (Coord.). *Desenvolvimento de sistemas de informações gerenciais integrados ao novo modelo de gestão do CNPq*. Projeto (renovação). [Florianópolis] : Universidade Federal de Santa Catarina, Programa de Pós-Graduação em Engenharia de Produção, 1999.
16. BAX, M. P. Introdução à linguagem de marcas. *Ciência da Informação*, Brasília, v. 30, n. 1, p. 32-38, 2000.
17. BRAY, T.; PAOLI, J.; SPERBERG-McQUEEN, C. M. *Extensible Markup Language (XML) 1.0: W3C recommendation*, 1998. Disponível em: <<http://www.w3.org/TR/REC-xml>>.
18. SALM JÚNIOR., J. F. *Proposta XML lattes currículo*. [Florianópolis]: PPGEP/UFSC, 2001. (Slides).
19. DIRETÓRIO DOS GRUPOS DE PESQUISA NO BRASIL - versão 4.1. Disponível em: <<http://www.cnpq.br/plataformalattes/dgp/versao4/>>. Acesso em: 20 nov. 2001.
20. LASSILA, O.; SWICK, R. *Resource Description Framework (RDF): model and syntax specification*, W3C Recommendation, 1999. Disponível em: <<http://www.w3.org/TR/REC-rdf-syntax>>.
21. BERNERS-LEE, T. *Weaving the web*. San Francisco : Harper, 1999.
22. MENEGHINI, R. Avaliação da produção científica e o Projeto SciELO. *Ciência da Informação*, Brasília, v. 27, n. 2, p. 219-220, 1998.
23. PATTERSON, D.; SNYDER, L.; ULLMAN, J. Best practices memo: evaluating computer scientists and engineers for promotion and tenure. *Computing Research News*, p. A-B, Sept. 1999. Special insert.
24. LAWRENCE, Steve. Online or invisible? *Nature*, v. 411, n. 6837, p. 521, 2001.

AGRADECIMENTOS

Os autores gostariam de agradecer ao CNPq, às instituições participantes do esforço de definição da LMPL e a José Francisco Salm Jr., principal desenvolvedor da LMPL.

Artigo recebido em 26/11/2001.
