

# GERAÇÃO (SEMI)AUTOMÁTICA DE METADADOS: UM CONTRIBUTO PARA A RECUPERAÇÃO DE OBJECTOS DE APRENDIZAGEM

Vitor Gonçalves

*Escola Superior de Educação – Instituto Politécnico de Bragança  
Campus de Santa Apolónia, Apartado 1101, 5301-856 Bragança  
vg@ipb.pt*

Eurico Carrapatoso

*Faculdade de Engenharia da Universidade do Porto / INESC Porto  
Rua Dr. Roberto Frias, s/n 4200-465 Porto  
emc@fe.up.pt*

## RESUMO

A alteração da Lei de Bases do Sistema Educativo Português, impulsionada pelo Processo de Bolonha, abre várias oportunidades de utilização do e-Learning não só no âmbito da formação contínua, mas também no contexto da formação inicial. Tal como em muitas outras instituições, o Moodle foi a opção de b-Learning mais natural e viável para a Escola Superior de Educação de Bragança. Com os recentes cursos (Plano Bolonha) foram identificados novos requisitos: necessidade de obter informação actual sobre os programas e conteúdos das diferentes disciplinas para suportar a tomada de decisão de inscrição em novas disciplinas; exigência de uma aprendizagem rápida; e obtenção de informação para suportar processos interdisciplinares. Essa informação reside na plataforma de e-Learning. Contudo, o Moodle não possui um mecanismo de pesquisa próprio que permita localizar e recuperar informação sobre os recursos de aprendizagem, garantindo que o objecto de aprendizagem propriamente dito não seja visualizado por utilizadores ou alunos não autorizados. A geração (semi)automática de metadados para facilitar a localização e recuperação dos objectos de aprendizagem foi a solução encontrada para responder aos requisitos identificados sem afectar a estrutura global do sistema de e-Learning. Assim, este artigo tem como principal objectivo descrever as actividades de especificação e desenvolvimento da solução encontrada.

## PALAVRAS-CHAVE

e-Learning, metadados, ontologias, Web Semântica, RDF, SPARQL.

## 1. INTRODUÇÃO

A alteração da Lei de Bases do Sistema Educativo Português, impulsionada pelo Processo de Bolonha, passou a reconhecer o conhecimento como um bem universal, a evidenciar a necessidade de tornar o Ensino Superior mais atractivo e mais próximo dos interesses da sociedade; a admitir a necessidade de adaptar o processo de aprendizagem aos conceitos e perspectivas da sociedade moderna e aos meios tecnológicos disponíveis, a promover a mobilidade e a dimensão europeia do Ensino Superior, a fomentar a aprendizagem ao longo da vida e, por último, a assumir legal e formalmente o ensino a distância como uma das modalidades especiais de educação escolar. Por conseguinte, prefiguram-se vários cenários de utilização do e-Learning não só no âmbito da formação contínua e ensino recorrente, mas também no contexto da formação inicial do Ensino Superior (licenciaturas e mestrados).

Uma das plataformas de e-Learning que mais interesse tem despertado nos últimos anos é o Moodle *Course Management System*. Tal como em muitas outras instituições, esta foi a opção de e-Learning e de b-Learning mais natural e viável para a Escola Superior de Educação de Bragança (ESEB).

Um sistema de e-Learning representa um ambiente de aprendizagem, no qual a distribuição de conteúdos multimédia, a interacção social e o apoio na aprendizagem são suportados pela Internet, Intranet ou Extranet.

Ao longo dos últimos anos, a ESEB tem vindo a recorrer a vários sistemas de informação para a Web (incluindo os sistemas para suportar a aprendizagem) para disponibilizar o acesso à informação. Mas, a descoberta de conhecimento nem sempre é uma tarefa fácil, já que o processo de recuperação, análise e agregação de informação é cada vez mais complicado. A grande quantidade de informação e a discrepância entre essa informação, a variedade das fontes de informação e a falta de mecanismos de integração dessas fontes, bem como as limitações dos mecanismos de recuperação, são alguns dos motivos que afectam negativamente esta situação.

Alguns destes problemas relacionam-se com questões de interoperabilidade e com a ausência de mecanismos de pesquisa que permitam procurar num determinado contexto específico. A forma de procurar informação, disponibilizada pelos motores de busca actuais, através de palavras-chave desprovidas de significado e contexto, leva a que muitas das vezes a informação passe despercebida, seja difícil de localizar ou, simplesmente, esteja inacessível em bases de dados apenas pesquisáveis com aplicações específicas.

A Web foi criada com a visão de que seria um espaço onde a informação teria um significado bem definido, facilitando a cooperação e a comunicação entre as pessoas e os agentes de software (Berners-Lee et al., 2001). Passada mais de uma década do seu lançamento, essa visão ainda não foi alcançada. A Web continua organizada na perspectiva dos humanos (linguagem natural e ênfase na apresentação através da linguagem HTML), o que torna a busca de informação uma tarefa árdua. Há que encontrar soluções de representação, organização, integração, intercâmbio e compreensão semântica da informação tanto na óptica dos humanos, como na óptica das máquinas ou dos agentes de software. É neste contexto que surge a denominada Web Semântica (*Semantic Web*) (W3C, 2001) como um cenário desejável para a Web. A Web Semântica é uma iniciativa liderada pelo *World Wide Web Consortium* (W3C), cujo propósito principal é recomendar e especificar tecnologias e linguagens que facilitem a integração, intercâmbio e compreensão semântica da informação tanto na óptica dos humanos, como na óptica das máquinas, nomeadamente através dos metadados (dados sobre dados), das ontologias (forma de representar explicitamente a semântica dos dados de um determinado domínio) e dos agentes de software (componente de software capaz de actuar de forma a resolver tarefas em nome do seu utilizador).

A ideia da Web Semântica pode resumir-se à seguinte questão: como fazer com que os computadores entendam o conteúdo da Web? O primeiro passo será organizar e estruturar a informação e o segundo será adicionar semântica às informações da Web, de tal forma que os agentes de software as possam compreender.

Primeiramente, o recurso a metadados permitirá registar a estrutura, o contexto e o significado dos recursos de informação, facilitando a recuperação de informação. Posteriormente, para esclarecer o significado universalmente aceite nos diversos domínios do conhecimento, será necessário recorrer às ontologias para melhor orientar os agentes de software. Contudo, a geração manual de metadados e de ontologias tem limitado o almejado sucesso dos sistemas que recorrem às potencialidades destas duas tecnologias.

Neste sentido, após referir sucintamente os conceitos e o estado da arte do e-Learning e das tecnologias de metadados, este artigo apresenta uma proposta para a geração (semi)automática de metadados a partir dos objectos de aprendizagem publicados num sistema de e-Learning Moodle e explicita sucintamente os aspectos mais relevantes inerentes às actividades de especificação e de desenvolvimento dos módulos correspondentes e dos mecanismos de pesquisa baseados em metadados que corroboram a utilidade da proposta.

## 2. E-LEARNING

Genericamente, o e-Learning preocupa-se, por um lado, com a comunicação entre o professor e o aluno (interacção social e intra-pessoal) e, por outro, com os conteúdos do curso (interacção com os recursos de aprendizagem). É esta última componente que interessa no âmbito deste artigo.

Nesta última década surgiram no mercado plataformas adequadas à criação de ambientes de aprendizagem hipermédia ou sistemas de e-Learning que integram tecnologias Web para permitir uma rápida distribuição e actualização dos conteúdos, tecnologias de comunicação para promover a interacção síncrona/assíncrona e ferramentas de autor para a criação de aplicações multimédia.

Não menosprezando as vantagens de cada uma destas plataformas, a verdade é que muitas das experiências de implementação de sistemas de e-Learning não têm obtido resultados satisfatórios devido, por um lado, ao tipo de matérias e de destinatários e, por outro, aos conteúdos disponibilizados.

No primeiro caso, justifica-se que o e-Learning tenda a ser complementado com sessões presenciais (esta combinação das duas formas de ensino designa-se por b-Learning ou *Blended Learning*) (Hofmann, 2002), maximizando os proveitos e minimizando os prejuízos que ambos possam acarretar. No segundo caso, há que reconhecer definitivamente que a transposição simples e directa dos conteúdos do ensino presencial para um formato digital (por exemplo, PDF) não é a solução adequada para criar cursos de e-Learning com qualidade e susceptíveis de proporcionar verdadeiramente a aprendizagem.

As tecnologias, por si só, não são suficientes para obter sistemas de e-Learning adequados às necessidades dos professores e alunos, pelo que o modelo pedagógico inerente à organização e estruturação dos conteúdos não deve ser descurado.

O reconhecimento desta situação traduziu-se numa constante procura de soluções para a criação, representação e estruturação dos conteúdos de aprendizagem, destacando-se os esforços de consórcios, entidades ou iniciativas, tais como: ADL (*Advanced Distributed Learning*), IMS/GLC (*Instructional Management Systems Global Learning Consortium*), IEEE/LTSC (*Institute of Electrical and Electronics Engineers Learning Technology Standards Committee*), AICC (*Aviation Industry CBT (Computer-Based Training) Committee*), ARIADNE (*ARIADNE Foundation*), entre outras.

No contexto do desenvolvimento de conteúdos de aprendizagem, podemos encontrar sistemas de aprendizagem electrónica em 4 etapas distintas de evolução (Torres et al., 2006).

A **primeira etapa** caracteriza-se pela publicação de disciplinas ou cursos de formação compostos por conteúdos indivisíveis ou inseparáveis (a baixa granularidade dos cursos e dos conteúdos muito dificilmente permitia a sua combinação e reutilização).

A **segunda etapa** aposta na publicação dos conteúdos de aprendizagem através de aplicações simples (páginas de texto, páginas HTML e alguns recursos multimédia). A interacção com os conteúdos centra-se em objectos de aprendizagem (*learning objects* ou simplesmente LOs) que podiam ser anotados com metadados e distribuídos através de IMS-CP (**IMS Content Packaging**). A granularidade, a definição dos objectivos, a combinação, a reutilização e a interoperabilidade dos LOs passaram a ser as principais preocupações no desenvolvimento de e-cursos.

Os LOs baseiam-se na filosofia da programação por objectos das ciências da computação, pelo que a ideia é construir pequenas peças de instrução para serem reutilizadas em diferentes contextos de aprendizagem, como se de peças de LEGO se tratasse. Embora o LOMWG (*Learning Objects Metadata Working Group*) do IEEE/LTSC apresente uma definição mais ampla, vamos considerar um LO como um recurso digital (texto, imagem, som, vídeo, applet Java, filme flash, programa de simulação, entre outros componentes distribuídos por intermédio de plug-ins apropriados) que pode ser reutilizado para apoiar a aprendizagem.

A granularidade de um LO pode variar da simples imagem ou gráfico até ao currículo completo de uma lição ou curso (LOMWG, 2002). Não obstante, quanto maior for a dimensão do LO, menor será a sua possibilidade de reutilização. Logo, quanto maior for a granularidade de uma disciplina ou curso de formação, maior é a sua flexibilidade e, consequentemente, do sistema de e-Learning.

A **terceira etapa** tem vindo a apostar em especificações para agregar os conteúdos educativos, nomeadamente especificações simples centradas em sequências de objectos de aprendizagem, tais como, IMS-SS (*Simple Sequencing*), impulsionando a normalização através do **SCORM** (*Sharable Content Object Resource Model*). O SCORM é um conjunto de normas, especificações e orientações técnicas para o desenvolvimento de conteúdos de aprendizagem e respectivos metadados, de modo a garantir a reutilização, interoperabilidade, durabilidade e acessibilidade (ADL, 2004).

O Modelo SCORM não é mais do que um manual de boas práticas composto de 4 manuais técnicos:

- 1) *SCORM Overview Book*: disponibiliza a introdução aos conceitos chave do SCORM, entre outras informações conceptuais;
- 2) *SCORM Content Aggregation Model*: modelo de agregação de conteúdos que especifica como encontrar, combinar, agregar, descrever, sequenciar e mover recursos de aprendizagem, usando metadados na importação ou exportação entre sistemas;
- 3) *SCORM Run-Time Environment*: ambiente de execução que especifica como executar os conteúdos e como registar o percurso do aluno, tendo como objectivo a interoperabilidade entre recursos de aprendizagem e LMSs (*Learning Management Systems*);
- 4) *SCORM Sequencing and Navigation*: sequenciação e navegação que descreve como os conteúdos podem ser ordenados para o aluno.

A **quarta etapa** avança lentamente para a adopção de linguagens de modelação educativas que permitem o desenvolvimento de processos de aprendizagem com maior flexibilidade pedagógico-didáctica.

Contrariamente às etapas anteriores, é colocada ênfase na personalização dos conteúdos e promovida uma participação mais activa do aluno no processo de aprendizagem, através da modelação de unidades de aprendizagem compostas, essencialmente, por cenários ou ambientes, objectivos, actividades, recursos ou objectos de aprendizagem, serviços e perfis. Este processo de modelação é conhecido por *Learning Design* e corresponde ao plano ou projecto de aprendizagem baseado num determinado método pedagógico que deriva do *Instructional Design* (processo sistemático de tradução dos princípios gerais de ensino/aprendizagem em planos para materiais de ensino/aprendizagem, ou seja, a análise sistemática das necessidades dos alunos deve resultar na planificação das estratégias e materiais mais adequados para satisfazer essas necessidades). A proposta que mais interesse tem despertado na comunidade científica tem sido a especificação *IMS-Learning Design*, enquanto especificação que permite a planificação da aprendizagem e a concepção de recursos educativos adequados, independentemente da metodologia pedagógica adoptada.

### 3. METADADOS

Embora não sejam um conceito recente, os metadados têm vindo a assumir uma importância crescente no âmbito da gestão e recuperação de informação. Em primeira instância, metadados são “dados sobre dados”. São um conjunto de dados estruturados que descrevem, explicam e localizam a informação, i.e., tornam mais fácil recuperar, usar ou controlar um recurso de informação (DCMI, 2002).

Os metadados não se aplicam apenas a informação textual, mas também a outros recursos digitais tais como imagens, músicas ou vídeos. A adição de descrições textuais a esse tipo de recursos tem vindo a ser uma forma de os indexar e catalogar, permitindo melhores resultados na sua localização e recuperação. Os metadados fornecem pistas, rotuladas pelas tags XML, para que as máquinas possam “compreender” a informação. Fornecem as bases para que os computadores tenham acesso a camadas de conhecimento que anteriormente só poderiam existir ou ser utilizadas pelo cérebro humano através da interpretação do significado dos documentos. A adição de um maior nível semântico à camada de informação permitirá que os computadores possam realizar inferências com vista à geração de novo conhecimento ou à sua reutilização.

A criação dos metadados pode ocorrer aquando da criação do recurso de informação à qual se referem ou aquando do processo de classificação e indexação do documento. Face à falta de especialistas para classificar e indexar os recursos no contexto de um sistema de e-Learning, muitas vezes a concepção de um recurso digital e a criação dos metadados para esse recurso ocorrem ao mesmo tempo e o autor do recurso é também o autor dos metadados, ou seja, o professor.

Os metadados podem ser parte integrante do documento ao qual se referem ou estarem armazenados num documento diferente daquele que contém a informação propriamente dita. É desejável que os metadados e os dados do recurso sejam mantidos em documentos separados para permitir diferentes graus de acesso aos metadados (pode haver necessidade de disponibilizar uma maior quantidade de metadados a um professor do que a um aluno), para oferecer diferentes descrições consoante o contexto (aquando da reutilização de um recurso por diferentes autores, pode haver necessidade de proceder a alterações somente nos metadados, mantendo-se inalterável o recurso e vice-versa) ou mesmo para fornecer diferentes manifestações de um objecto conceptual (um documento textual em PDF e o mesmo documento numa versão áudio armazenada num ficheiro MP3 referidas na mesma descrição ou metadados). Esta estratégia pode não ser necessária, ou mesmo adequada, se a linguagem para a descrição dos recursos permitir a riqueza semântica suficiente, como é o caso da RDF (*Resource Description Framework*).

O esquema de metadados mais popular é o DCM ou *Dublin Core Metadata* (não esquecendo o MARC 21 *Concise Format for Bibliographic Data* no âmbito das bibliotecas e centros de documentação), enquanto que o LOM ou *Learning Object Metadata* é o mais utilizado no campo da Educação. Contudo, a RDF é a proposta que mais se destaca no contexto particular da Web Semântica, uma vez que é uma recomendação W3C e permite expressar os conjuntos de elementos de metadados dos esquemas DCM e LOM.

#### 3.1 Dublin Core Metadata

As normas de metadados NISO Z39.85-2001 e ISO 15836-2003, que dizem respeito à especificação formal do Dublin Core Metadata Element Set (DCMES), correspondem a um conjunto de especificações para a descrição dos recursos da informação de um qualquer domínio do conhecimento através de um vocabulário

de 15 elementos. O DCMES usa quinze elementos (designados *Core Elements*) para descrever qualquer recurso de informação (DCMI, 2002):

- *Title* (título ou nome atribuído ao recurso);
- *Creator* (entidade responsável pela criação ou existência do recurso);
- *Subject* (assunto e palavras-chave que caracterizam o conteúdo do recurso);
- *Description* (descrição ou resumo do conteúdo do recurso);
- *Publisher* (entidade responsável por editar, publicar ou manter acessível o recurso);
- *Contributor* (entidade responsável por qualquer contribuição para o conteúdo do recurso);
- *Date* (data inerente à criação ou publicação do recurso);
- *Type* (tipo, função, natureza ou género do recurso);
- *Format* (formato do recurso: físico ou digital, tamanho ou duração);
- *Identifier* (referência para identificar o recurso num determinado contexto: URI, ISBN, etc.);
- *Source* (referência a um recurso de onde o recurso actual deriva);
- *Language* (língua do conteúdo intelectual do recurso: pt, en, fr);
- *Relation* (referência a um recurso relacionado);
- *Coverage* (extensão, alcance ou âmbito do recurso);
- *Rights* (gestão dos direitos: direitos de propriedade intelectual, direitos de autor, entre outros).

A norma *Dublin Core Metadata* (DCM) inclui dois níveis para a descrição de recursos: simples (*simple*) e qualificado (*qualified*). Os esquemas de metadados que usam apenas os quinze elementos mencionados designam-se por *Simple Dublin Core Metadata*. Cada um destes quinze elementos pode ser refinado com recurso a um conjunto limitado de qualificadores (*Dublin Core Qualifiers*). Os esquemas de metadados que, para além dos 15 elementos que constituem o DCMES, usam outros elementos (*Audience*, *Provenance*, *RightsHolder*, *accrualMethod*, *accrualPeriodicity*, *accrualPolicy*, *instructionalMethod*) e os elementos de refinamento designam-se por *Qualified Dublin Core Metadata* (Hillmann, 2005).

Tanto o *Simple DC* como o *Qualified DC*, para além de serem expressos em XML (*eXtensible Markup Language*), também podem ser codificados em: *Expressing Simple Dublin Core in RDF/XML* e *Expressing Qualified Dublin Core in RDF/XML*.

No contexto das ferramentas para o e-Learning, o editor eXe (*eLearning XHTML editor*) é uma das ferramentas que usa a especificação DCM para anotar os objectos de aprendizagem.

### 3.2 Learning Object Metadata

Os metadados LOM (*Learning Object Metadata*) são uma norma ou especificação que permite descrever e facilitar a localização e a recuperação de LOs. Os atributos que permitem a descrição de um LO podem ser agrupados em nove categorias (LOMWG, 2002):

- Geral (*General*) - refere-se a informação geral que descreve o LO como um todo;
- Ciclo de vida (*Lifecycle*) - agrupa os atributos referentes ao ciclo de vida de um recurso: a história de um LO e respectivos contributos para essa evolução, para além dos inerentes ao estado actual desse LO;
- Meta-metadados (*Metametadata*) - reúne informações sobre a própria instância dos metadados (em vez do LO que a instância do metadado descreve). Ou seja, descreve os metadados que indexam o LO;
- Técnica (*Technical*) - agrupa os requisitos e características técnicas do LO;
- Educacional (*Educational*) - agrupa os atributos pedagógicos do LO;
- Direitos (*Rights*) - agrupa os direitos de propriedade intelectual e condições de utilização do LO;
- Relação (*Relation*) - incide nos atributos inerentes às relações entre recursos, ou seja, agrupa características que definem a relação entre o LO e outros objectos educacionais correlacionados;
- Anotação (*Annotation*) - permite comentários sobre o uso educacional do LO e provê informação de quando e por quem foram criados os comentários;
- Classificação (*Classification*) - descreve a posição do LO em relação a um sistema de classificação - UDC (*Universal Decimal Classification*) ou ARIADNE, etc.

Tal como para o DCM, têm também vindo a ser desenvolvidos esforços para de expressar o LOM em RDF (Nilsson et al., 2003): *Draft IEEE Learning Object Metadata RDF binding* (IEEE LTSC P1484.12.4).

No contexto das ferramentas para o e-Learning, o RELOAD editor, RELOAD LD editor e LomPad são algumas das ferramentas que usam a especificação LOM para anotar os objectos de aprendizagem.

### 3.3 Resource Description Framework

A *Resource Description Framework* (RDF) é uma *framework*, tecnologia ou linguagem para representar informação na Web (Lassila e Swick, 1999). A RDF é uma linguagem que permite a codificação, o intercâmbio e a reutilização de metadados estruturados (Daconta et al., 2003). Basicamente, a RDF fornece uma forma de associar propriedades a recursos, através de *statements*. Um *statement* (declaração) tem três componentes: *resource* (recurso), *property* (propriedade) e *value* (valor), através dos quais é possível descrever os recursos de informação disseminados na rede, definindo recursos e identificando as suas propriedades e valores. Mas uma propriedade pode representar também o relacionamento entre recursos. Para definir o significado, as características e as relações entre recursos recorremos ao Esquema RDF (RDF Schema ou RDFS). Podemos resumir uma declaração como “o recurso (sujeito) possui a propriedade (predicado) com o valor (objecto). Por exemplo, a declaração <“http://www.vgportal.ipb.pt”, “autor”, “Vitor Gonçalves”> teria o significado: A página http://www.vgportal.ipb.pt (recurso) tem como autor (propriedade) Vitor Gonçalves (valor), tal como se pode ver no grafo da figura 1.

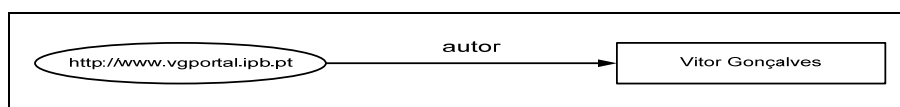


Fig. 1. Grafo de um Modelo de Dados RDF

Ou seja, um conjunto de descrições RDF descreve um recurso. Se aplicarmos a sintaxe RDF baseada em XML ao grafo acima obtemos o seguinte código:

```
<?xml:namespace ns = http://www.w3.org/RDF/RDF/" prefix="rdf" ?>
<?xml:namespace ns = http://purl.oclc.org/DC/" prefix="dc" ?>
<rdf: RDF>
  <rdf: Description about="http://www.vgportal.ipb.pt">
    <dc:Creator>Vitor Gonçalves</dc:Creator>
  </rdf: Description>
</rdf: RDF>
```

Num diagrama de modelo de dados, um objecto pode corresponder a outro recurso identificado por um URI, uma string ou outro qualquer tipo de dados definido em XML. Por exemplo, se necessitarmos fornecer mais informação sobre o autor (e-mail e instituição), o valor “Vitor Gonçalves” teria que ser substituído por um recurso (identificado por um URI) com as propriedades (nome, e-mail e instituição).

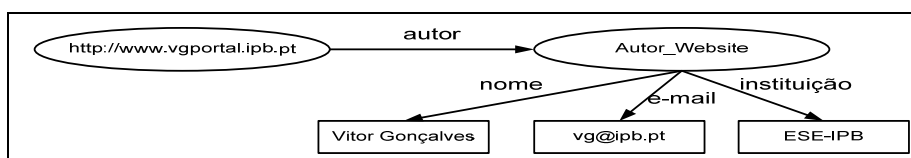


Fig. 2. Grafo de um Modelo de Dados RDF

Mas, mesmo assim, há a necessidade de especificar que deverá ser indiferente usar como identificação “Gonçalves, Vitor”, “Vitor Gonçalves” ou “Vitor M. B. Gonçalves”, pois o mecanismo de busca interpretá-los-á correctamente, já que cada propriedade corresponde a um tipo definido globalmente. Um exemplo similar é a confusão entre datas do tipo UK (dia/mês/ano) e do tipo US (mês/dia/ano) (Thompson, 2004).

A RDF não fornece a informação suficiente para que uma máquina (aplicação ou agente de software) perceba que o valor de uma propriedade que representa um autor é uma referência a uma pessoa e não a uma instituição, localidade ou veículo. Necessitamos de um esquema que valide as afirmações expressas em RDF. Necessitamos do esquema RDFS ou de outra proposta ontológica (OWL é a recomendação do W3C).

Genericamente, o RDF *Core* (termo utilizado para distinguir o núcleo RDF do esquema RDFS) define como descrever as propriedades e os valores dos recursos, enquanto que o RDF *Schema* define as propriedades que podem ser utilizadas para definir esquemas.

## 4. GERAÇÃO AUTOMÁTICA DE METADADOS

Esta secção tem por objectivo apresentar um protótipo para a geração (semi)automática de metadados a partir dos LOs publicados num sistema de e-Learning Moodle com vista a facilitar a localização e recuperação dos conteúdos educativos ou simplesmente de informações acerca desses conteúdos (i.e. metadados). Primeiramente convém referir os parâmetros principais que influenciaram a escolha da plataforma Moodle:

- É um software *Open Source*, permitindo o seu uso e modificação, respeitando a licença GNU;
- Baseia-se no Construtivismo, nomeadamente no Construcionismo Social;
- Oferece um ambiente intuitivo e agradável caracterizado pela flexibilidade e facilidade na navegação;
- Disponibiliza de um conjunto de recursos e actividades suficientes para a criação on-line de LOs, evitando que os professores se limitem a disponibilizar apenas documentos PDF, Word ou Power Point;
- Permite a importação e execução de pacotes SCORM desenvolvidos em editores (tais como: Editor eXe e Editor RELOAD e Editor Reload *Learning Design*) facilitando o intercâmbio e reutilização de LOs;
- A inclusão de *Learning Design* através de um motor de execução de projectos de aprendizagem é possível através da integração com sistema LAMS (*Learning Activity Management System*);
- É um ambiente modular, permitindo o desenvolvimento e a adição de novos módulos e blocos.

Tal como referimos, os cursos adequados ao processo de Bolonha originaram a identificação de novos requisitos. Referiramo-nos mais concretamente ao seguinte:

- Os cursos adequados ao processo de Bolonha permitem que os alunos possam escolher disciplinas de outros cursos para complementarem o seu plano de estudo. Para tal, necessitam de informação actual sobre os programas e conteúdos leccionados nas diferentes disciplinas para poderem tomar decisões.
- Reconhece-se desde há muito tempo que a prática da interdisciplinaridade é relevante para o processo de aprendizagem e de formação do aluno. Mas, o diálogo entre professores nem sempre é possível e muitas vezes a interdisciplinaridade fica apenas manifestada nos planos, por falta de acesso a informação sobre o processo de aprendizagem das restantes disciplinas do mesmo curso.

Essa informação reside na plataforma de e-Learning. Contudo, o Moodle não possui um mecanismo de pesquisa próprio que permita localizar e recuperar informação sobre os LOs, garantindo que o LO propriamente dito não seja visualizado por alunos não autorizados. Mas, para realizar esse tipo pesquisas, os LOs devem previamente ser descritos através de metadados. No entanto, o Moodle não permite a anotação dos LOs (recursos e actividades) criados on-line na plataforma. A possibilidade de importação de pacotes SCORM, IMS-CP e IMS-LD permite adicionar LOs descritos com metadados. Contudo, a observação da prática docente mostrou-nos que muito raramente os professores o faziam.

Com vista a livrar o professor do processo tedioso de descrever semanticamente os recursos e evitar eventuais erros de anotação por falta de familiarização com as especificações, exigia-se que a descrição dos LOs fosse automatizada tanto quanto possível. Neste sentido, a solução mais natural passou pela criação de dois módulos para anotação (semi)automática de metadados. Um em conformidade com a especificação DCMES e o outro de acordo com a especificação IEEE-LOM (e IMS-LRM para LOM), deixando a decisão da escolha para o professor. Embora a norma LOM seja a mais adequada para descrever LOs, alguns dos recursos e actividades podem ser anotados com DCMES, não só porque o grau de complexidade é menor, mas também porque no caso de artigos, posters, manuais ou referências documentais poderá ser a especificação mais adequada. Não obstante, posteriormente uns podem ser mapeados nos outros caso o professor decida alterar a especificação que descreve os LOs da sua disciplina.

No caso do módulo de metadados DCMES, dos 15 elementos do *Simple DC*, treze são gerados automaticamente, e dos 22 elementos e 33 qualificadores do *Qualified DC*, 30 são gerados automaticamente. Destes, os formadores do caso de estudo identificaram 11 elementos sobre os quais ocasionalmente têm necessidade de proceder a alterações. No caso do módulo de metadados LOM, dos 35 (sub)elementos que compõem o *Simple LOM* (elementos e sub-elementos seleccionados para formar um perfil ou esquema de metadados menos complexo para descrever os LOs do Moodle), 28 foram gerados automaticamente, e dos 58 elementos e sub-elementos do *Complete LOM* (versão completa da especificação IEEE-LOM 1484.12.3/D2-2004), normalmente 46 elementos são gerados automaticamente (este valor poderá ser ligeiramente inferior caso o tipo de LOs derive de um módulo de actividade novo). Destes, apenas 9 elementos foram identificados pelos formadores deste caso de estudo como elementos em que pontualmente têm necessidade de proceder a alterações. A maioria dos elementos é gerada com base na informação textual dos LOs ou outras informações existentes na base de dados do sistema Moodle, à excepção da extracção de palavras-chave que se baseia em

algoritmos que incidem na correspondência dos termos com a ontologia do sistema. Não obstante, para melhorar a fase de extracção de palavras têm vindo a ser estudadas alternativas baseadas nos trabalhos de Ferreira e Kofuji (2006), Lourenso (2005), Shi e outros (2003), Han e outros (2003); Takasu (2003) e Barbosa (2003).

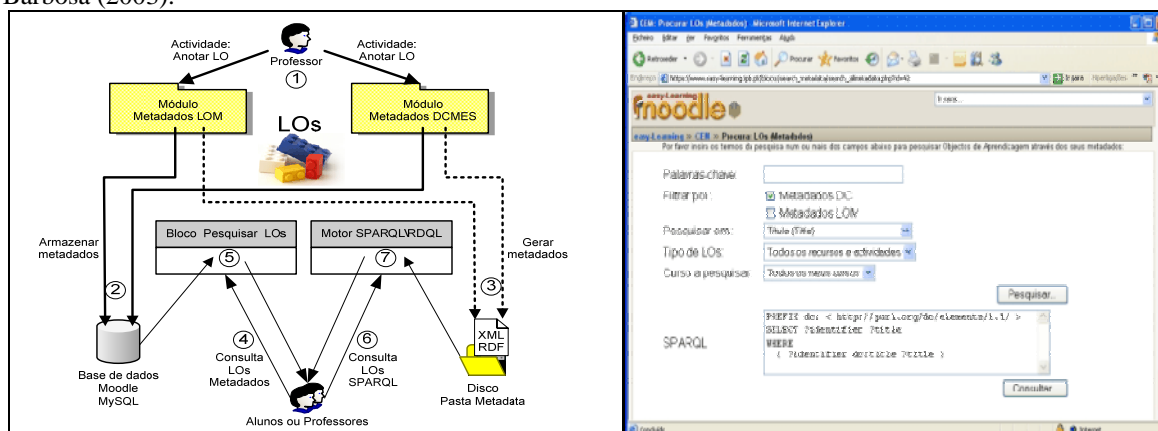


Fig. 3. Geração automática de metadados e recuperação de LOs

A proposta acima ilustra os principais componentes desenvolvidos em PHP (dois módulos que correspondem a duas actividades distintas, um bloco de pesquisa baseado nos metadados armazenados na base de dados do Moodle e um motor SPARQL/RDQL baseado nos metadados em RDF/XML), a sua integração e as principais funções:

1. No contexto de uma disciplina do sistema de e-Learning, o(s) professor(es) podem descrever as suas actividades ou recursos (LOs) adicionando a actividade “Metadados LOM” ou “Metadados DCMES”;
2. Seleccionado o LO, os metadados são gerados automaticamente e o professor é convidado a complementá-los ou alterá-los;
3. Caso proceda a alterações deverá gravar para gerar os metadados que serão armazenados na pasta “metadata” da disciplina (caso não exista será criada automaticamente) no formato RDF/XML;
4. O processo de pesquisa de LOs pode ser efectuado através dos registos de metadados existentes na base de dados do Moodle ou através dos ficheiros RDF/XML gerados automaticamente;
5. A pesquisa baseada na base de dados permite consultar todos os LOs pelo nome do criador, professor, palavras-chave, assunto ou escolhendo um dos elementos de uma das normas. É possível pesquisar em todos os cursos ou apenas nos cursos do utilizador ou num determinado curso específico. Caso o utilizador não tenha autorização para visualizar os LOs devolvidos, pode no entanto ver os metadados de acordo com as definições atribuídas pelo professor/criador dos metadados;
6. Uma vez que a disciplina pode conter pacotes SCORM/IMS recheados de metadados, para além dos ficheiros gerados pelos módulos, foi desenvolvido também um motor SPARQL/RDQL em PHP recorrendo a *RDF API for PHP* para realizar consultas directamente nos ficheiros RDF/XML.
7. O resultado devolvido em XML é formatado recorrendo a folhas de estilo em linguagem XSL (*eXtensible Style Language*). Com vista a tirar partido das potencialidades da linguagem RDQL, GRDDL-*Microformat Extration* e SQL-SPARQL *Bridge* e das tecnologias de ontologias e regras de inferência (RDFS, OWL e SWRL) este módulo continua em desenvolvimento.

A opção de desenvolver os dois tipos de pesquisa (uma baseada na base de dados e a outra baseada nos documentos RDF/XML) teve como finalidade permitir a comparação da utilização das tecnologias para a Web Semântica com a tecnologia SQL convencional. O bloco de pesquisa baseado em metadados armazenados na base de dados permitiu satisfazer os requisitos de alunos e professores. No entanto, o motor de pesquisa SPARQL/RDQL permitiu maximizar as potencialidades dos metadados ao orientar as inferências através da ontologia do sistema e, conseqüentemente, obter resultados mais precisos e navegar nas relações entre os LOs. O domínio da ontologia para o sistema de e-Learning e as principais classes e relações são ilustradas no modelo gráfico da figura 4.

Comparando com alguns trabalhos de geração (semi)automática de metadados e de ontologias que surgiram nestes últimos 3 anos, não encontramos nenhum com aplicação directa numa plataforma de e-



Learning. Aqueles que tinham aplicação em sistemas de e-Learning, embora recorressem a *frameworks* (Jena) e motores de inferência baseados em metadados e ontologias, não efectuavam geração automática de metadados ou estavam dependentes de uma única especificação. Não obstante, projectos como ELENA (2006), MOSAICA (2006) e LUISA (2006) têm vindo a desenvolver esforços para garantir este requisito.

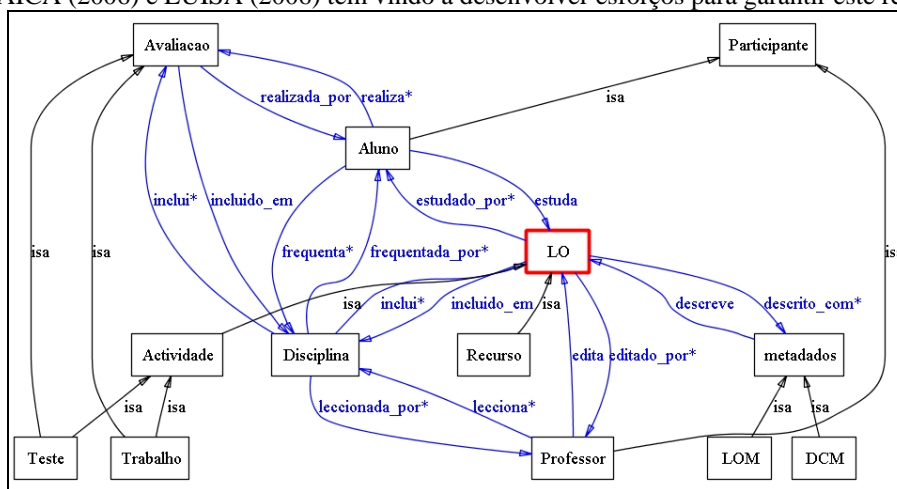


Fig. 4. Ontologia para o sistema de e-Learning

## 5. CONCLUSÃO

A primeira geração da WWW preocupou-se com a implementação da infra-estrutura tecnológica da Web (o conteúdo era disponibilizado através de páginas estáticas), a segunda incidiu na construção de aplicações Web (o conteúdo passou também a ser disponibilizado através de páginas dinâmicas geradas a partir de BD) e, actualmente, a terceira avança rumo a uma Web de Dados – a Web Semântica (Gonçalves, 2007).

A XML assume-se como uma tecnologia flexível que fornece a sintaxe normalizada para documentos estruturados, através da qual o significado pode ser comunicado independentemente da plataforma, sistema operativo ou aplicação, facultando a criação de uma estrutura arbitrária nos documentos, mas nada diz acerca do significado dessa estrutura. Esta tarefa é deixada para as tecnologias de metadados ou metalinguagens (RDF(S)/OWL) que permitem descrever a informação de forma não ambígua para depois ser processada pelas máquinas (Thompson, 2004).

A proposta apresentada neste artigo pode ser experimentada em <http://www.easy-learning.ipb.pt> e constitui mais um pequeno passo rumo à tão almejada Web Semântica. Em última instância, o motor de pesquisa poderia ser aplicado a um nível superior integrando os diversos sistemas de e-Learning Moodle distribuídos, por exemplo, pelas diversas escolas secundárias.

## REFERÊNCIAS

- ADL (2004). Sharable Content Object Reference Model (SCORM). *Advanced Distributed Learning*. Acesso em: 04 de Junho de 2004, Disponível em: <http://www.adlnet.org>
- Antoniou, G., & Harmelen, F. (2004). *A Semantic Web Primer*. [S.l.]: The MIT Press.
- Barbosa, I. C. (2003). *Prospecção de Dados na Manutenção Semi-automática de Ontologias, Dissertação de mestrado*. Universidade Federal de Minas Gerais, Minas Gerais.
- Beckett, D. (2004). RDF/XML Syntax Specification (Revised). *W3C Recommendation*. Acesso em: 10 de Agosto de 2005, Disponível em: <http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/>
- Berners-Lee, T. (2005). *Semantic Web Concepts*. *WWW Consortium, MIT Computer Science & Artificial Intelligence Laboratory*. Acesso em: 30 de Outubro de 2006, Disponível em: <http://www.w3.org/2005/Talks/0517-boit-tbl/>
- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web – a new form of the Web content that is meaningful to computer will unleash a revolution of new possibilities. *Scientific American*.
- Bourda, Y., & Bich-Liên, D. (2003). The Semantic Web for Learning Resources. *IEEE - Computer Society*. Acesso em: Agosto de 2004, Disponível em: <http://csdl.computer.org/comp/proceedings/icalt/2003/1967/00/19670322.pdf>

- Brase, J., & Nejdil, W. (2004). Ontologies and Metadata for eLearning. In S. Staab & R. Studer (Eds.), *Handbook on Ontologies* (pp. 555-574). Berlin: Springer-Verlag.
- Breitman, K. (2005). *Web Semântica: A Internet do Futuro*. Brasil: LTC.
- Brickley, D., & Guha, R. V. (2004). RDF Vocabulary Description Language 1.0: RDF Schema. *W3C Recommendation*. Acesso em: 25 de Agosto de 2005, Disponível em: <http://www.w3.org/TR/2004/REC-rdf-schema-20040210/>
- Daconta, M. et al., 2003. *The Semantic Web*. Wiley Publishing, Inc, Indianapolis, USA.
- Davies, J., Fensel, D., & Harmelen, F. v. (2003). *Towards The Semantic Web – Ontology-Driven Knowledge Management*. Chichester, England: John Wiley & Sons, Ltd.
- DCMI (2002). Dublin Core Metadata. *Dublin Core Metadata Initiative*. Disponível em: <http://www.dublincore.org>
- ELENA (2006). About ELENA: Smart Spaces for LearningTM. *Information Society Technologies*. Acesso em: 17 de Fevereiro de 2006, Disponível em: <http://www.elena-project.org>
- Fensel, D. (2001). *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce*. [S.l.]: Springer Verlag.
- Ferreira, E., & Kofugi, S. (2006). Geração Automática de Metadados. In M. Cunha & Á. Rocha (Eds.), *Actas da 1.ª Conferência Ibérica de Sistemas e Tecnologias de Informação - Sistemas e Tecnologias de Informação no Espaço Ibérico* (Vol. 2). Ofir: Escola Superior de Tecnologia - Instituto Politécnico do Cávado e do Ave.
- Gonçalves, V., 2007. A Web Semântica no Contexto Educativo: Um sistema para a recuperação de objectos de aprendizagem baseado nas tecnologias para a Web Semântica, para o e-Learning e para os agentes, Dissertação de Doutoramento, FEUP.
- Han, H. C., Giles, L., Manavoglu, E., Zha, H., Zhang, Z., & Fox, E. A. (2003). Automatic document metadata extraction using support vector machines, *Proceedings of the Third ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 37-48). New York: ACM Press.
- Harmelen, F. (2004). The Semantic Web: What, Why, How, and When. *IEEE Intelligent Systems*.
- Hayes, P. (2004). RDF Semantics. *W3C Recommendation*. Acesso em: 10 de Setembro de 2005, Disponível em: <http://www.w3.org/TR/2004/REC-rdf-nt-20040210/>
- Hillmann, D. (2005). Using Dublin Core. *Dublin Core Metadata Initiative*. Acesso em: 30 de Março de 2006, Disponível em: <http://dublincore.org/documents/usageduide/>
- Hofmann, J., 2002. Blended Learning Case Study. In ROSSETT, A. *The ASTD E-Learning Handbook*. McGraw-Hill.
- IEEE-LTSC (2002). IEEE 1484.12.1-2002 - Draft Standard for Learning Object Metadata. *Learning Technology Standards Committee of the IEEE*. Disponível em: [http://ltsc.ieee.org/wg12/files/LOM\\_1484\\_12\\_1\\_v1\\_Final\\_Draft.pdf](http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf)
- IMS (2004). IMS Learning Resource Metadata Specification. *Instructional Management Systems Project - IMS Global Learning Consortium*. Acesso em: 02 de Junho de 2004, Disponível em: <http://www.imsglobal.org/specifications.html>
- Jenkins, C., Jackson, M., Burden, P., & Wallis, J. (1999). Automatic RDF Metadata generation for Resource Discovery. In E. Tang (Ed.), *Proceedings of the WWW8 - Eighth International World Wide Web Conference* (Vol. 31, pp. 1305-1320). Toronto, Canada: Foretec Seminars.
- Klyne, G., & Carroll, J. (2004). Resource Description Framework (RDF): Concepts and Abstract Syntax. *W3C Recommendation*. Acesso em: Agosto de 2005, Disponível em: <http://www.w3.org/TR/2004/REC-rdf-primer-20040210/>
- Lassila, O., & Swick, R. (1999). Resource Description Framework (RDF) Model and Syntax Specification. *World Wide Web Consortium*. Acesso em: 12 de Janeiro de 2004, Disponível em: <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>
- LOMWG (2002). Standard for Learning Object Metadata. *IEEE-LTSC Committee*. Acesso em: 26 de Abril de 2004, Disponível em: <http://ltsc.ieee.org/wg12/index.html>
- Lourenço, R. (2005). *Segmentação De Objetos Complexos Em Um Sistema De Banco De Dados Objeto Relacional Baseado Em Grids, Dissertação de Doutoramento*. Escola Politécnica da Universidade de São Paulo.
- LUISA (2006). Learning Content Management System Using Innovative Semantic Web Services Architecture. *European Commission FP6 funded Specific Targeted Research Project*. Disponível em: <http://www.luisa-project.eu>
- Moodle. (2004). Moodle course management system. *Moodle Pty Ltd*. Disponível em: <http://moodle.org>
- MOSAICA (2006). Semantically Enhanced Multifaceted Collaborative Access to Cultural Heritage. *European Commission - IST Programme* Acesso em: 16 de Dezembro de 2006, Disponível em: <http://www.mosaica-project.eu/>
- Nilsson, M., Palmér, M., & Brase, J. (2003). The LOM RDF Binding - Principles and Implementation. *Proceedings of the Third Annual ARIADNE conference* Acesso em: 12 de Fevereiro de 2006, Disponível em: <http://kmr.nada.kth.se/papers/SemanticWeb/LOMRDFBinding-ARIADNE.pdf>
- OCDE (2006). Reviews of National Policies for Education - Tertiary Education in Portugal - Examiners' Report, *Directorate for Education - Education Committee - Organisation de Coopération et de Développement Economiques (Organisation for Economic Co-operation and Development)*. Centro Cultural de Belém, Lisbon, Portugal.
- Shi, R., Mary, K., & Zubair, M. (2003). *Automatic Metadata Discovery from non-cooperative Digital Libraries*. USA: Dep. of Computer Science, Old Dominion University.
- Takasu, A. (2003). Bibliographic attribute extraction from erroneous references based on a statistical model, *Proceedings of the Third ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 49-60). New York: ACM Press.
- Thompson, J. (2004, Maio). Features: Semantic Web – Straight Talking. *PC Pro Computing in the Real World*.
- Torres, J., Dodero, J., Diaz, P., & Ignacio, A. (2006). Diseño Instruccional en Internet y los Languages de Modelado Educativo. In L. Alonso, L. González, B. Manjón & M. Nistal (Eds.), *Proceedings of the 8th International Symposium on Computers in Education (SIE'06)* (Vol. 2). Leon: Universidad de Leon.
- W3C (2001). Semantic Web Activity. *World Wide Web Consortium*. Disponível em: <http://www.w3.org/2001/sw/>