

UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação
ISSN 0103-2569

MEGEN: UM GERADOR AUTOMÁTICO DE METADADOS
PARA A CO TEIA

Luciano T. E. Pansanato
Renata P. M. Fortes

Nº 287

RELATÓRIOS TÉCNICOS



São Carlos – SP
Dez./2006

SYSNO	1563077
DATA	/ /
ICMC - SBAB	

Sumário

1. Introdução	1
2. Um <i>Application Profile</i> para Páginas Wiki Educacionais.....	2
3. Geração Automática de Metadados	4
3.1. Geração de Metadados Baseada no Recurso	4
3.2. Geração de Metadados Baseada no Contexto.....	4
3.3. Geração de Metadados Baseada em Template	4
3.4. Geração de Metadados Baseada em Regra	4
3.5. Tratamento de Conflitos	5
4. MeGen	5
5. Trabalho Relacionado	7
6. Resultados da Aplicação do MeGen aos Repositórios CoTeia	7
7. Conclusões e Trabalhos Futuros.....	12
Referências	13

MeGen: Um Gerador Automático de Metadados para a CoTeia

Luciano T. E. Pansanato^{1,2}, Renata P. M. Fortes²

¹Universidade Tecnológica Federal do Paraná (UTFPR)
Campus de Cornélio Procopio
Av. Alberto Carazzai, 1640 – 86300-000 Cornélio Procopio, PR

²Departamento de Ciências de Computação
Instituto de Ciências Matemáticas e de Computação (ICMC/USP)
Caixa Postal 668 – 13560-970 São Carlos, SP
luciano@utfpr.edu.br, renata@icmc.usp.br

Abstract. *In this paper, we describe our experience of implementing an application profile to educational wiki pages. We discuss methods for automatic and semi-automatic generation of metadata. Furthermore, we also report the implementation of an automatic metadata generator to support the application profile and these methods. The results of the application of the metadata generator to repositories of educational wiki pages are also described.*

Resumo. *Neste artigo é descrita a experiência de implementação de um application profile para páginas wiki educacionais. Métodos para geração automática e semi-automática de metadados são discutidos e é apresentado um gerador automático de metadados para o application profile. Os resultados da aplicação do gerador para repositórios de páginas wiki educacionais também são descritos.*

1. Introdução

A função final dos registros de metadados é permitir aos usuários encontrar recursos relevantes. No entanto, antes de acontecer qualquer descoberta é necessário preencher os campos do registro de metadados associado com cada recurso. No contexto educacional, se o usuário não conseguir encontrar um recurso educacional, este não pode ser reusado e o conceito de reusabilidade de objetos de aprendizagem não irá funcionar.

Muitos padrões de esquemas de metadados (*metadata schema*) têm sido propostos para facilitar interoperabilidade e descoberta de recursos (por exemplo, IEEE LOM [IEEE LTSC 2002] e Dublin Core [DCMI 2003]); e a maioria dos elementos de metadados de alto nível definidos por esses esquemas têm de ser produzidos manualmente pelos criadores de registros de metadados [Westermann & Klas 2003]. As exceções são aqueles esquemas que definem uma quantia significativa de metadados técnicos de baixo nível que são extraídos automaticamente. Além disso, determinados elementos de metadados são opcionais em muitos esquemas; conseqüentemente, os metadados para um determinado registro podem ser insuficientes, e provavelmente as bases de dados não possuem todos os campos preenchidos [McClelland 2003].

Um sistema de autoria pode gerar valores automaticamente para determinados elementos de metadados como autor, data de criação, e palavras-chave; mas ainda existem muitos campos que os usuários devem preencher. Essa é a motivação para o desenvolvimento e uso de métodos para a geração automática e semi-automática de metadados. Assim, tempo e esforço podem ser minimizados se o usuário não quer criar os metadados manualmente.

Na Seção 2 é descrito um *application profile* para páginas wiki educacionais. Na Seção 3 são discutidos categorias de métodos para a geração automática e semi-automática de metadados. Na Seção 4 é apresentado o gerador automático de metadados desenvolvido, denominado de MeGen (*Metadata Generator*). Na Seção 5 é apresentada uma comparação deste trabalho com um trabalho relacionado. Na Seção 6 são descritos os resultados da aplicação do MeGen a repositórios¹ de páginas wiki educacionais. A Seção 7 é dedicada à apresentação das conclusões e trabalhos futuros.

2. Um *Application Profile* para Páginas Wiki Educacionais

O *IEEE Learning Technology Standards Committee*² (LTSC) define objeto de aprendizagem (*learning object*) como qualquer entidade, digital ou não digital, que pode ser usada, reusada ou referenciada durante a aprendizagem apoiada por tecnologia. Essa definição é altamente genérica e grupos diferentes a têm refinado para algo mais específico. Em geral, essas definições adicionais são amplas o suficiente para incluir todo tipo de informação que pode ser acessado publicamente na Web.

Neste trabalho, páginas wiki são consideradas objetos de aprendizagem (ou recursos educacionais) porque elas são usadas para apoiar atividades educacionais. As páginas wiki são mantidas em repositórios baseados na Web usando a CoTeia [Arruda Jr et al. 2002, Arruda Jr & Pimentel 2001], uma ferramenta de Aprendizagem Colaborativa Apoiada por Computador (*Computer Supported Collaborative Learning – CSCL*) usada no Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (ICMC/USP) para complementar as aulas presenciais com atividades colaborativas. A CoTeia é uma ferramenta colaborativa assíncrona baseada em wiki, e é análoga à CoWeb [Guzdial 1999]. Os usos de páginas wiki na CoTeia variam bastante, por exemplo, criação colaborativa de conteúdo, atividades de revisão, distribuição de informação, entre outras [Macedo et al. 2001].

O registro de metadados (*metadata record*) de cada página wiki da CoTeia é um *application profile* [Duval & Hodgins 2003] baseado no padrão *IEEE Learning Object Metadata* (LOM) [IEEE LTSC 2002] e no padrão *Dublin Core* (DC) [DCMI 2003], ou seja, o esquema de metadados usado na CoTeia contém elementos desses dois padrões.

O LOM é um padrão de metadados usado para descrever recursos educacionais baseados na Web. Os seus elementos são organizados em nove categorias: *General*, *Life Cycle*, *Meta-Metadata*, *Technical*, *Educational*, *Rights*, *Relation*, *Annotation*, e *Classification*. O modelo de dados do padrão LOM é uma hierarquia de elementos de

¹ Neste trabalho, repositórios são páginas Web dinâmicas que derivam parte ou todo o seu conteúdo a partir de arquivos ou bases de dados.

² <http://ltsc.ieee.org/wg12/>

dados, e inclui elementos agregados e elementos simples (os nós folha da hierarquia). Os elementos de dados simples têm valores individuais definidos por meio do seu espaço de valores e tipo de dados associado. O esquema de numeração dos elementos de dados representa uma hierarquia de agregação de elementos de dados e seus componentes. Por exemplo, o elemento agregado *1.1:General.Identifier* tem dois componentes: *1.1.1: General.Identifier.Catalog* e *1.1.2:General.Identifier.Entry*.

O DC é um outro padrão de metadados para descrever uma gama extensiva de recursos interligados em rede. O padrão DC tem dois níveis: Simples e Qualificado. O DC Simples abrange 15 elementos; o DC Qualificado inclui três elementos adicionais, e também um grupo de refinamentos de elemento (também chamados de qualificadores) que refinam a semântica dos elementos. Por exemplo, o elemento *Date* pode ser refinado pelos qualificadores *Available*, *Created*, *Date Accepted*, *Date Copyrighted*, *Date Submitted*, *Issued*, *Modified*, e *Valid*.

Ao application profile da CoTeia foram adicionados cinco elementos de metadados para fornecer suporte a um domínio de aplicação específico: páginas wiki educacionais no ICMC/USP. A idéia é fornecer metadados adicionais com vocabulário conhecido pela comunidade do ICMC/USP para melhor apoiar os usuários da CoTeia na descoberta de páginas wiki (recursos educacionais). Na Tabela 1 são apresentados os elementos de metadados do application profile da CoTeia. Nessa tabela também é mostrada a sobreposição da semântica dos elementos dos padrões DC e LOM, e dos novos elementos (Local). Cada registro de metadados é codificado em RDF segundo diretrizes específicas [DCMI 2002, Nilsson et al. 2003].

Tabela 1. Um application profile para páginas wiki educacionais.

LOM	Dublin Core	Local
General.Title	Simple DC: title	-
General.Language	Simple DC: language	-
General.Description	Simple DC: description	-
General.Keyword	Simple DC: subject	-
LifeCycle.Version	-	-
LifeCycle.Contribute.Entity (Role = author)	Simple DC: creator	-
LifeCycle.Contribute.Entity (Role = publisher)	Simple DC: publisher	-
LifeCycle.Contribute.Date	Simple DC: date	-
Technical.Format	Simple DC: format	-
Technical.Size	Qualified DC: extent	-
Technical.Location	-	-
Educational.LearningResourceType	Simple DC: type	Learning Material Type
Educational.Context	-	Brazilian Context
Educational.TypicalAgeRange	Qualified DC: audience	-
-	Simple DC: source	-
-	-	Knowledge Area*
-	-	Course
-	-	Learning Activity Type

*Anteriormente denominado de Topic, este elemento de metadados foi renomeado.

3. Geração Automática de Metadados

São quatro as principais categorias de métodos identificados para a geração automática de metadados [Pansanato & Fortes 2005a, 2005b]: (1) geração baseada no recurso, (2) geração baseada no contexto, (3) geração baseada em template, e (4) geração baseada em regra. Essas categorias são discutidas a seguir nas subseções 3.1 a 3.4. Na Subseção 3.5 são descritas algumas abordagens que podem ser adotadas para tratar conflitos, ou seja, quando valores diferentes são gerados para um mesmo elemento de metadados.

3.1. Geração de Metadados Baseada no Recurso

Os métodos para geração de metadados baseados no recurso (*resource-based metadata generation*) são baseados na análise do conteúdo do próprio recurso. No caso da CoTeia, os recursos são as páginas wiki. Nesse sentido, podem ser utilizadas técnicas típicas, por exemplo: classificação de idioma (*language classification*), sumarização automática (*automatic summarization*), e extração de palavras-chave (*keyword extraction*).

3.2. Geração de Metadados Baseada no Contexto

Os métodos para geração de metadados baseados no contexto (*context-based metadata generation*) são baseados na informação do contexto no qual o recurso é usado. A informação de contexto pode ser obtida a partir da estrutura dos recursos (por exemplo, estrutura de aninhamento ou composição), do gerenciamento dos usuários relacionados ao recurso, e a partir do sistema de arquivos ou outro software responsável pelo gerenciamento do armazenamento do recurso.

3.3. Geração de Metadados Baseada em Template

Os métodos para geração de metadados baseados em template (*template-based metadata generation*) são baseados na informação obtida a partir de templates ou profiles associadas ao recurso. Os métodos dessa categoria são considerados semi-automáticos porque requerem a intervenção do usuário em algum momento. As templates podem ser mantidas no nível de sistema, de grupo de recursos, e no nível de usuário. A idéia é manter um conjunto de metadados válido para todos os recursos criados (sistema), um conjunto válido para um grupo de recursos, e um conjunto válido para todos os recursos criados por um mesmo usuário. No caso de templates no nível de usuário, por exemplo, um usuário pode evitar a entrada manual de alguns metadados para cada página wiki criada por ele inserindo essas informações na sua template.

3.4. Geração de Metadados Baseada em Regra

Os métodos para geração de metadados baseados em regra (*rule-based metadata generation*) são baseados em regras que permitem explorar relacionamentos entre os metadados do recurso. Por exemplo, determinados metadados podem ser derivados de outros metadados. Isso é especialmente indicado quando é necessário mapear um elemento de metadados local para algum elemento de um esquema padronizado ou quando existe uma regra heurística conhecida que pode ser aplicada.

3.5. Tratamento de Conflitos

O uso de métodos diferentes pode melhorar o processo de geração automática, mas valores diferentes podem ser gerados para um mesmo elemento de metadados. Nesse caso, ocorre um conflito entre os geradores que implementam os métodos que precisa ser resolvido. Existem diversas abordagens para resolver conflitos; dependendo do elemento, uma abordagem pode funcionar melhor do que outra [Cardinaels et al. 2005]:

1. Incluir todos os valores gerados no conjunto resultante de metadados;
2. Propor as opções para o usuário decidir qual valor utilizar;
3. Tentar encontrar os geradores mais prováveis a estarem corretos e utilizar seus valores no resultado;
4. Aplicar heurísticas para decidir qual valor seria correto.

A seguir é apresentada a implementação de um gerador de metadados para as páginas wiki da CoTeia.

4. MeGen

O gerador de metadados, denominado de *Metadata Generator* (MeGen), é um sistema para a geração de metadados que implementa diversos métodos de acordo com as categorias descritas na Seção 3. A arquitetura do MeGen é apresentada na Figura 1. Um *crawler* (*spider* ou *robot*) automaticamente busca as páginas wiki nos repositórios (*sites*) especificados em um arquivo de configuração e armazena a referência (*link*) e uma cópia de cada página wiki identificada no processo. O MeGen utiliza as referências e as cópias das páginas wiki no processo de geração automática dos metadados. No arquivo de configuração do MeGen são especificadas as diretivas necessárias para especializar os métodos para um determinado domínio. Por exemplo, podem ser especificados os arquivos com as tabelas de palavras-chave (códigos, nomes, etc.) utilizadas pelo método baseado no recurso, as informações de contexto utilizadas pelo método baseado no contexto, e as regras de mapeamento utilizadas pelo método baseado em regra. Essas informações são chamadas na Figura 1 de “Dados de domínio”. Os valores gerados para os elementos de metadados são armazenados em um “Modelo RDF” no formato de triplas. Esse modelo pode ser manipulado posteriormente por ferramentas para diversas finalidades, por exemplo, para a descoberta de recursos relevantes.

Uma hierarquia de classes simplificada para o MeGen é mostrada na Figura 2. O diagrama de classe fornece uma visão geral do gerador de metadados apresentando a classe *MetadataGenerator* como uma superclasse das classes que correspondem às categorias de métodos descritos na Seção 3. A partir desse diagrama pode-se notar que alguns elementos de metadados podem ser gerados por dois ou mais métodos (operações), por exemplo, o tamanho (*Extent*) e o título (*Title*) de uma página wiki. Nesse caso o método definido para manipular conflitos deve ser capaz de resolver quais valores devem ser considerados. O modelo de classes mostrado na Figura 2 pode ser estendido facilmente com novas subclasses para a geração de metadados como também adaptado para outro tipo de objeto de aprendizagem.

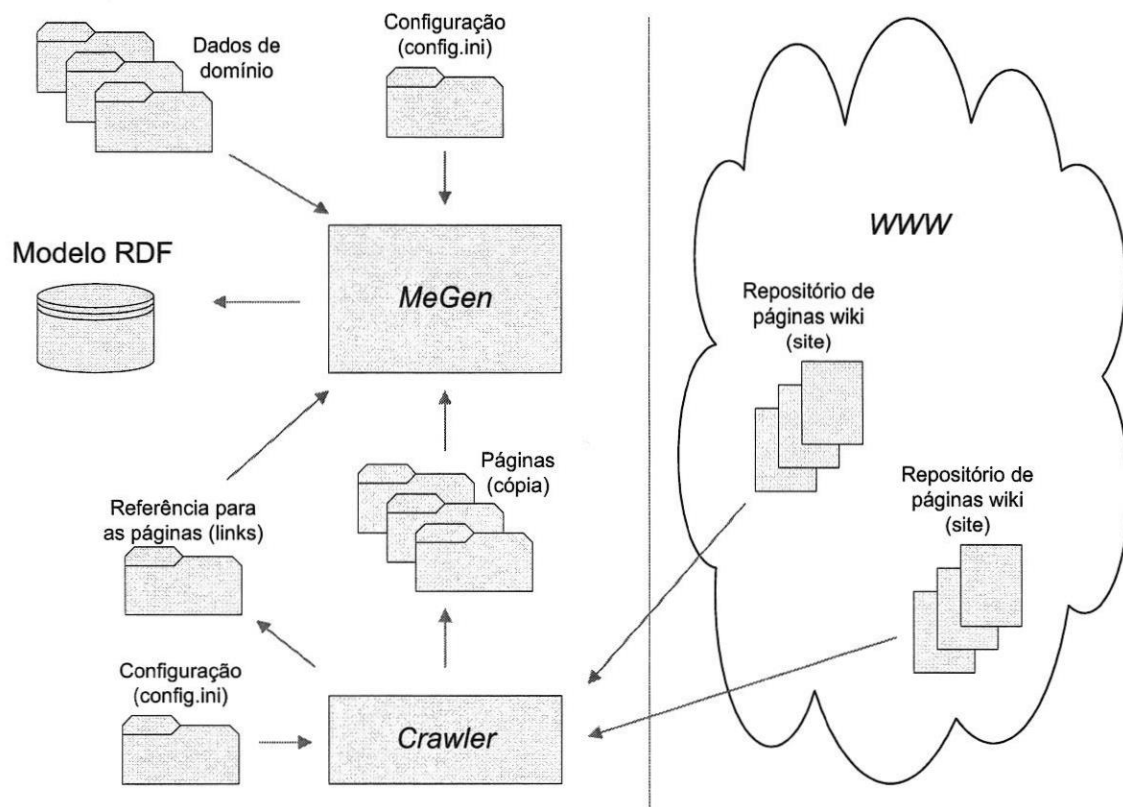


Figura 1. Arquitetura do MeGen.

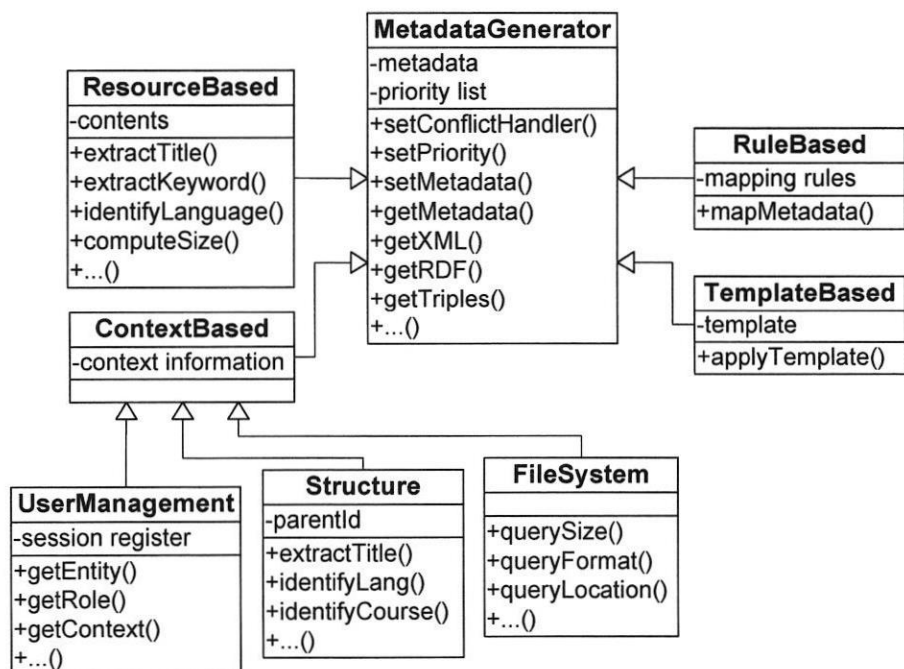


Figura 2. Uma hierarquia de classes simplificada para o MeGen.

Na versão atual do MeGen o tratamento de conflitos foi implementado usando prioridades. Um valor de prioridade pode ser atribuído para cada par <elemento de metadados, gerador>. As abordagens 1 e 3 para resolver conflitos, descritas na

Subseção 3.5, podem ser facilmente implementadas por meio de prioridades. A primeira abordagem – incluir todos os valores gerados no conjunto resultante de metadados – é obtida quando as prioridades são iguais para um mesmo elemento de metadados. A terceira abordagem – tentar encontrar os geradores mais prováveis a estarem corretos e utilizar seus valores no resultado – é obtida quando são estabelecidas prioridades diferentes para um mesmo elemento de metadados. Na verdade, o MeGen não “tenta encontrar”, mas segue a definição contida no arquivo de configuração de qual seria o gerador mais preciso para aquele elemento de metadados.

5. Trabalho Relacionado

O *ARIADNE Knowledge Pool System (KPS)* [Duval et al. 2001] utiliza três métodos combinados para a geração automática e semi-automática de metadados: *profile* (template), análise de conteúdo, e similaridade com outros objetos. Em outro trabalho [Cardinaels et al. 2005] os autores também exploram análise de contexto e apresentam um *framework* para geração automática de metadados denominado de Automating Metadata Generation (AMG).

A classe *MetadataGenerator* do MeGen corresponde à classe *Indexer* no AMG. Além disso, as subclasses *ResourceBased* e *ContextBased* são similares às subclasses *ObjectBasedIndexer* e *ContextBasedIndexer*. Entretanto, no AMG não são explorados métodos baseados em regra. No MeGen é usado um método simples baseado em regras de mapeamento: um elemento de metadados, geralmente relativo ao domínio local, é mapeado para gerar algum elemento de um esquema padronizado. Por exemplo, o caso do elemento *Course* que é mapeado para *Context* e *Typical Age Range* do padrão LOM.

Regras heurísticas também podem ser exploradas. Esse tipo de regra pode ser aplicado quando heurísticas sobre os elementos de metadados são conhecidas. Por exemplo, o valor de *3.4:Meta-Metadata.Language* pode ser um valor fixo, geralmente “pt” (Português, no caso de páginas wiki no domínio educacional do ICMC/USP) e pode ser obtido a partir da template do sistema. Esse elemento representa o idioma do registro de metadados. Uma heurística para gerar automaticamente esse elemento é classificar o idioma utilizado em *1.4:General.Description* (ou outro elemento de texto livre).

6. Resultados da Aplicação do MeGen aos Repositórios CoTeia

A coleção utilizada para avaliar o MeGen consistiu de 4.234 páginas wiki, divididas em dois repositórios. Na Tabela 2 são mostrados dados referentes a esses repositórios: o nome pelo qual o repositório é conhecido, o período principal que as páginas foram criadas, a URL de acesso, e a quantidade de páginas wiki disponíveis. O repositório denominado CoWeb³ também armazena algumas páginas wiki relacionadas a projetos.

³ O nome CoWeb faz referência ao software desenvolvido por Mark Guzdial do Geórgia Tech e utilizado no ICMC/USP antes do desenvolvimento da CoTeia. Atualmente, o software utilizado para acesso e edição das páginas são versões daquele disponível em <http://incubadora.fapesp.br/projects/coteia/>.

Tabela 2. Repositórios utilizados para avaliar o MeGen.

Nome	Período	URL	Quantidade de páginas
CoTeia	2004-2006	http://coteia.icmc.usp.br/coteia/	1.731
CoWeb	2001-2003	http://safedevl.icmc.usp.br/coweb/	2.503
Total:			4.234

Na Figura 3 são mostradas as porcentagens de valores gerados pelo MeGen para os metadados associados às páginas wiki do repositório CoTeia (2004-2006). Na Figura 4 são mostradas as porcentagens com relação ao repositório CoWeb (2001-2003). Na Figura 5 são mostradas as porcentagens totais, considerando os dois repositórios.

Aproximadamente 84,9% de todos os elementos de metadados do *application profile* da CoTeia foram gerados pela versão do MeGen utilizada. Entretanto, como esperado, alguns elementos de metadados que são gerados automaticamente por meio de uma abordagem algorítmica tiveram altas porcentagens (100%) de valores gerados (veja Figura 5, com as porcentagens relativas aos dois repositórios), por exemplo: *Title* e os metadados técnicos (*Format*, *Extent*, e *Location*). Em contraste, outros elementos de metadados que tipicamente requerem conhecimento do domínio para serem gerados tiveram porcentagens menores, por exemplo, aqueles que descrevem informação educacional sobre o conteúdo das páginas: *Learning Material Type* (60,7) e *Learning Activity Type* (44,6%).

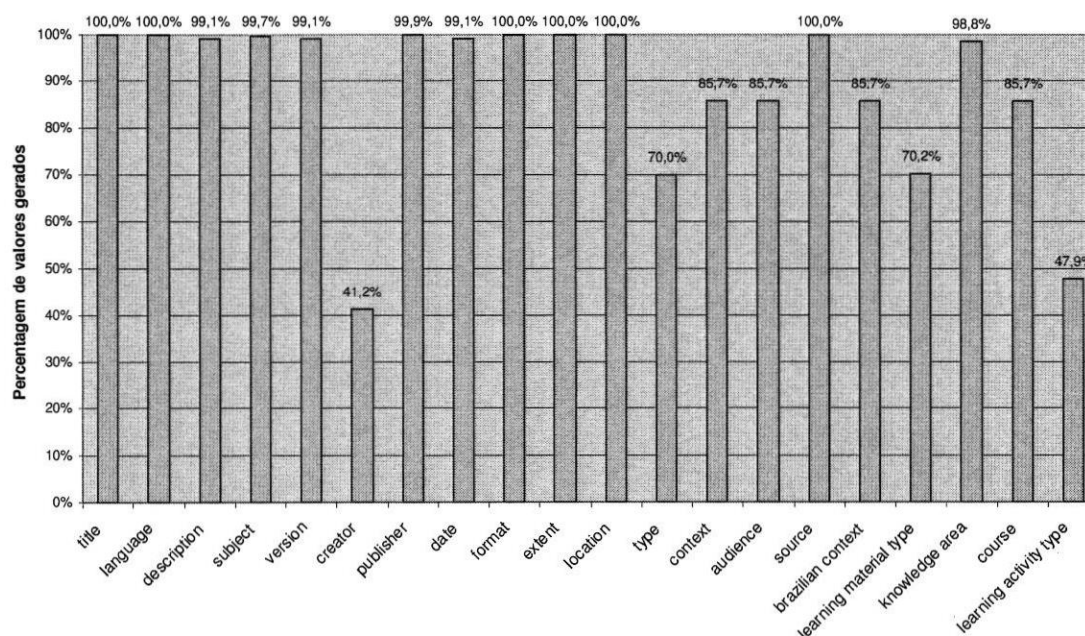


Figura 3. Porcentagem de valores gerados para os metadados associados às páginas wiki do repositório CoTeia (2004-2006).

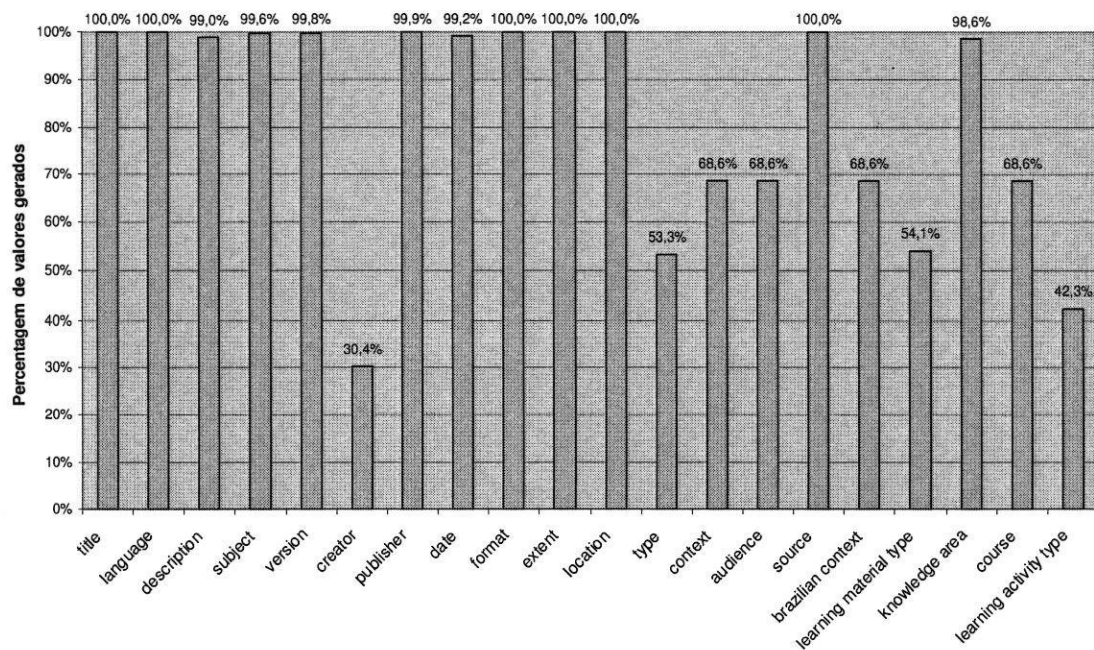


Figura 4. Porcentagem de valores gerados para os metadados associados às páginas wiki do repositório CoWeb (2001-2003).

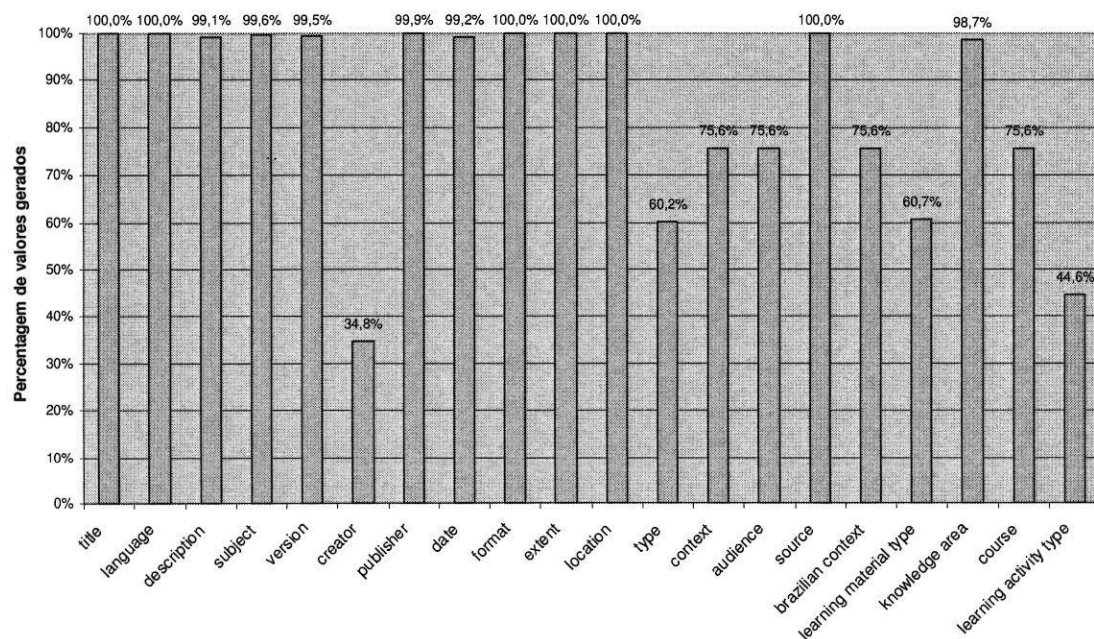


Figura 5. Porcentagem de valores gerados para os metadados associados às páginas wiki dos repositórios CoTeia (2004-2006) e CoWeb (2001-2003).

O comportamento complementar do MeGen com relação a determinados elementos de metadados é um aspecto importante observado nos resultados obtidos. Por exemplo, o elemento de metadados *Course* pode ser gerado automaticamente usando os métodos baseado no recurso e baseado no contexto. No caso do método baseado no recurso o valor é identificado por meio da localização do código ou nome da disciplina no conteúdo da página wiki. No caso do método baseado no contexto o valor é obtido a partir da estrutura de aninhamento ou composição das páginas wiki. Na CoTeia, as

páginas de uma mesma disciplina são organizadas hierarquicamente a partir de uma mesma entrada (link na página inicial). Para o elemento de metadados *Course*, o contexto utilizado consiste na lista de páginas ascendentes a partir da página em análise. No caso de *Course*, o MeGen foi configurado para considerar os valores gerados pelo método baseado no recurso (considerado mais preciso) como de maior prioridade do que aqueles gerados pelo método baseado no contexto. Um conflito ocorria quando valores diferentes eram obtidos pelos métodos, e o valor gerado pelo método baseado no contexto (prioridade menor) era sempre descartado. O aspecto interessante observado é que os valores para *Course* foram gerados parte pelo método baseado no recurso e parte pelo método baseado no contexto, e ocorreram poucos conflitos. Na Figura 6 são mostrados os dados referentes à geração dos valores para o elemento de metadados *Course* com relação ao total de valores que poderiam ser gerados. Na Figura 7 é mostrada a contribuição real de cada um dos métodos com relação ao total de valores efetivamente gerados. O método baseado no recurso contribuiu com 45,3% dos valores (9,1% mais 36,2% porque este método tem prioridade maior) e o método baseado no contexto contribuiu com 54,7%. O método baseado no contexto também conseguiu gerar valores iguais aos gerados pelo método baseado no recurso para 36,2% das páginas. No caso de *Course* pode-se dizer que os valores gerados pelo método baseado no recurso foram complementados pelo método baseado no contexto. Portanto, ao decidir pela implementação de uma geração automática de metadados é importante considerar métodos diferentes para um mesmo elemento de metadados.

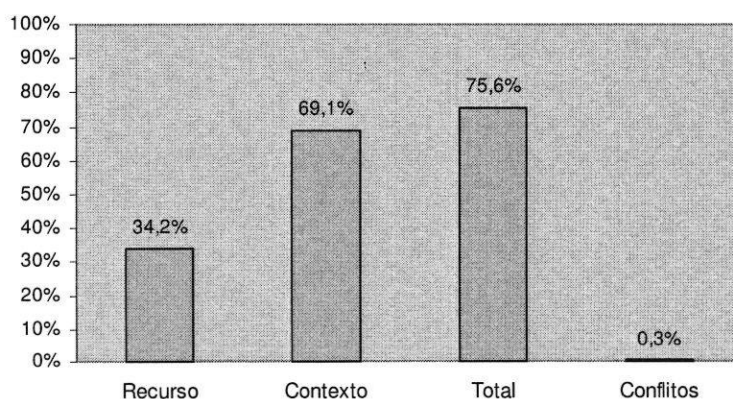


Figura 6. Porcentagens referentes à geração de valores para o elemento de metadados *Course*.

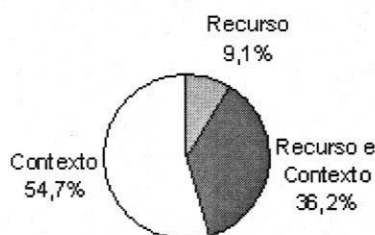


Figura 7. Porcentagens referentes à contribuição de cada um dos métodos na geração do total de valores para o elemento de metadados *Course*.

Esse comportamento do MeGen com relação a *Course* não foi um caso isolado. Na Figura 8 são mostrados os dados referentes à geração do elemento de metadados *Knowledge Area*. No caso de *Knowledge Area*, ocorreram muitos conflitos em comparação com *Course*. No entanto, a porcentagem total de valores gerados para este elemento de metadados é maior do que as porcentagens individuais dos métodos baseado no recurso e baseado no contexto (veja Figura 8). Na Figura 9 é mostrada a contribuição real de cada um dos métodos com relação ao total de valores efetivamente gerados.

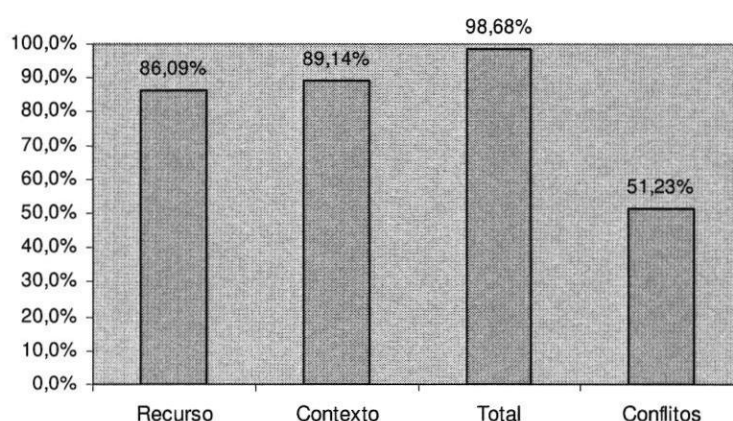


Figura 8. Porcentagens referentes à geração de valores para o elemento de metadados Knowledge Area.



Figura 9. Porcentagens referentes à contribuição de cada um dos métodos na geração do total de valores para o elemento de metadados Knowledge Area.

O comportamento do método baseado em regra implementado também é um aspecto interessante observado nos resultados. Os elementos de metadados mapeados mantêm a mesma porcentagem do elemento utilizado como entrada para o mapeamento (por exemplo, veja as porcentagens de *Context*, *Audience*, *Brazilian Context* e *Course* na Figura 5). O método de mapeamento implementado permite o mapeamento de um ou mais elementos de metadados para outros elementos de metadados. Na Figura 10 são mostrados os mapeamentos explorados para o *application profile* da Tabela 1. Os mapeamentos de metadados, também conhecidos como *crosswalks*⁴, são comuns na literatura sobre bibliotecas digitais. Um *crosswalk* fornece um mapeamento dos elementos de metadados de um padrão para um outro padrão, por exemplo, de Dublin

⁴ <http://libraries.mit.edu/guides/subjects/metadata/mappings.html>

Core para IEEE LOM. Neste trabalho são explorados relacionamentos entre um elemento de metadados do domínio local e um elemento de um esquema padronizado definido no *application profile*. Por exemplo, Course (Local) é mapeado para Brazilian Context (Local) que, por sua vez, é mapeado para Context (LOM) e Audience (Dublin Core) (veja Figura 10). A vantagem dessa abordagem é permitir que o usuário tenha acesso aos recursos usando visões diferentes, por exemplo, audience em vez de course.

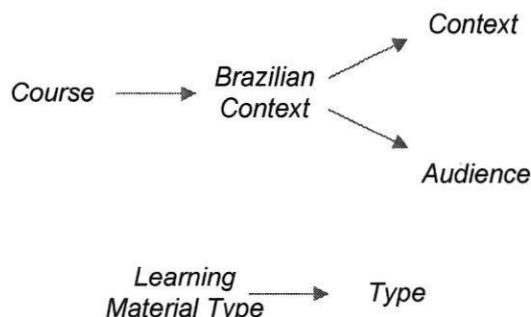


Figura 10. Mapeamentos entre elementos de metadados explorados por um método baseado em regra implementado no MeGen.

A qualidade dos metadados obtidos a partir do uso do MeGen foi verificada por amostragem. É importante examinar os valores obtidos para ajustar melhor as prioridades dos métodos diferentes que podem trabalhar separadamente para gerar o mesmo elemento de metadados. Os elementos de metadados considerados com baixa porcentagem de geração também foram examinados para verificar se os valores gerados são representativos, isto é, se eles correspondem a uma representação correta da página wiki. Nesse caso, os metadados gerados para páginas wiki com pouco conteúdo tendem a ser incorretos ou insuficientes. Isso acontece quando esses metadados são somente gerados usando métodos baseados no recurso.

7. Conclusões e Trabalhos Futuros

A geração automática de metadados é um recurso importante na ausência de motivação ou tempo para o usuário dedicar-se ao trabalho de preenchimento dos campos de metadados. Além disso, é uma alternativa que pode ser considerada para repositórios de recursos nos quais não existem interfaces disponíveis para a entrada manual de metadados. Com a disponibilidade de metadados sobre os recursos, ferramentas que manipulam esses metadados com o objetivo de permitir a descoberta de recursos relevantes para o usuário podem ser desenvolvidas.

O reuso dos métodos implementados no MeGen é variável e é necessário um estudo da capacidade de adaptação e/ou extensão para outros domínios. Por um lado, os métodos genéricos podem ser facilmente reusados, por exemplo: classificação de idioma, extração de palavras-chave, exploração do contexto do sistema de arquivos, mecanismo de mapeamento. Por outro lado, os métodos dependentes do domínio (páginas wiki educacionais) são mais difíceis de serem reusados. No entanto, alguns métodos podem ser generalizados para páginas Web em geral, por exemplo, o método que extrai metadados a partir da *tag* <META> (título, autor, etc.).

A sequência deste trabalho envolve o desenvolvimento de um ambiente para navegação através das páginas wiki de diferentes repositórios. Esse ambiente deve

integrar diversas ferramentas que utilizam metadados em seu mecanismo e permitir ao usuário escolher aquelas mais adequadas ou a melhor combinação de ferramentas (isto é, a melhor estratégia) de acordo com seus diferentes níveis de habilidades e conhecimento (*background*), suas preferências, e o tipo de informação que estão procurando no momento.

Referências

- Arruda Jr, C. R. E., Izeki, C. A., Pimentel, M. G. C. CoTeia: Uma ferramenta colaborativa de edição baseada na Web. In *Proceedings of the 8th Brazilian Symposium on Multimedia and Hypermedia Systems*, Fortaleza, 2002, 371-374.
- Arruda Jr, C. R. E., Pimentel, M. G. C. Projeto e implementação de um sistema colaborativo de edição. *Revista Eletrônica de Iniciação Científica*, 1(2), 2001, <http://www.sbc.org.br/index.php?language=1&subject=101>.
- Cardinaels, K., Meire, M., Duval, E. Automating Metadata Generation: the Simple Indexing Interface. In *Proceedings of the 14th international conference on World Wide Web*, 2005, 548-556.
- DCMI, Dublin Core Metadata Initiative. *Expressing Qualified Dublin Core in RDF / XML*, 2002. <http://dublincore.org/documents/dcq-rdf-xml/>.
- DCMI, Dublin Core Metadata Initiative. *Using Dublin Core*, 2003, <http://www.dublincore.org/documents/usageguide>.
- Duval, E., Forte, E., Cardinaels, K., Verhoeven, B., Van Durm, R., Hendrikx, K., Forte, M., W., Ebel, N., Macowicz, M., Warkentyne, K., Haenni, F. The ARIADNE Knowledge Pool System. *Communications of the ACM*, 44(5), 2001, 73-78.
- Duval, E., Hodgins, W. A LOM Research Agenda. In *Alternate Paper Tracks Proc. 12th International Conference on World Wide Web*, 2003.
- Guzdial, M. Supporting learners as users. *The Journal of Computer Documentation*, 23(2), 1999, 3-13.
- IEEE LTSC, Learning Technology Standards Committee. *IEEE 1484.12.1-2002 Draft Standard for Learning Object Metadata*, 2002.
- Macedo, A.; Bulcão Neto, R.F.; Pimentel, M.G.P. Autoria Colaborativa na Web: Experiências e Reflexões sobre a CoWeb. *Revista Brasileira de Informática na Educação (RBIE)*, September 2001.
- McClelland, M. Metadata standards for educational resources. *Computer*, 36(11), 2003, 107-109.
- Nilsson, M., Palmér, M., Brase, J. The LOM RDF Binding - Principles and Implementation. In *Proceedings of the Third Annual ARIADNE conference*, 2003.
- Olivié, H., Cardinaels, K., Duval, E. Issues in Automatic Learning Object Indexation, In *Proceedings of the World Conference on Educational Multimedia, Hypermedia and Telecommunications*, 2002(1), 2002, 239-240.

- Pansanato, L. T. E., Fortes, R. P. M. Strategies for automatic LOM metadata generating in a web-based CSCL tool. In *Proceedings of the 11th Brazilian Symposium on Multimedia and the Web*, 2005, 1-8.
- Pansanato, L. T. E., Fortes, R. P. M. Strategies for Filling Out LOM Metadata Fields in a Web-Based CSCL Tool. In *Proceedings of the Third Latin American Web Congress*, 2005, 187-190.
- Westermann, U., Klas, W. An Analysis of XML Database Solutions for the Management of MPEG-7 Media Descriptions. *ACM Computing Surveys*, 35(4), 2003, 331-373.