

EMPACOTAMENTO DE EXPERIMENTOS CONTROLADOS COM ABORDAGEM EVOLUTIVA BASEADA EM ONTOLOGIA

SCATALON, Lilian Passos
Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP)
lilian.scatalon@gmail.com

GARCIA, Rogério Eduardo
Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP)
rogerio@fct.unesp.br

RESUMO: Um corpo de conhecimento em Engenharia de Software requer a replicação de experimentos controlados. O conhecimento gerado por um experimento é registrado no chamado pacote de laboratório, o qual deve ser revisado por um eventual grupo de pesquisa com a intenção de replicá-lo. Entretanto, pesquisadores enfrentam dificuldades ao revisar pacotes de laboratório, o que leva a problemas de compartilhamento de conhecimento entre grupos de pesquisa. Além disso, a falta de padronização é um obstáculo à integração de um estudo isolado em um corpo de conhecimento comum. Nesse contexto, é proposto um *workflow* para gerar pacotes de laboratório baseados na *ExperOntology*, uma ontologia do domínio de experimentos controlados, visando a contribuir na revisão e no entendimento de pacotes de laboratório.

PALAVRAS-CHAVE: Experimento Controlado, Ontologia, Engenharia de Software Experimental, Representação de Conhecimento

ABSTRACT: *A body of knowledge in Software Engineering requires experiments replications. The knowledge generated by a study is registered in the so-called lab package, which must be reviewed by an eventual research group with the intention to replicate it. However, researchers face difficulties reviewing lab packages, what leads to problems in share knowledge among research groups. Besides that, the lack of standardization is an obstacle to the integration of the knowledge from an isolated study in a common body of knowledge. In this context, we propose a workflow to generate lab packages based on ExperOntology, an ontology of controlled experiments domain, aiming to facilitate the reviewing and understanding of lab packages.*

KEYWORDS: *Controlled Experiment, Ontology, Experimental Software Engineering, Knowledge Representation*

INTRODUÇÃO

A Engenharia de Software Experimental tem por objetivo a realização de estudos que avaliem métodos, técnicas e ferramentas envolvidos no desenvolvimento de software. Por meio da realização de um estudo, são gerados resultados e conclusões relativas à aplicação da respectiva tecnologia sob investigação (SHULL et al., 2002; WOHLIN et al., 1999).

A ISERN (International Software Engineering Research Network) é uma comunidade composta de grupos de pesquisa que acreditam que a pesquisa na área deve ser realizada em um contexto experimental. Segundo o manifesto da ISERN (ISERN, 1993), cada um desses grupos produziram

e continuarão produzindo modelos válidos de Engenharia de Software dentro do ambiente de seus laboratórios. Porém, no intuito de construir os componentes e modelos básicos da disciplina de Engenharia de Software, é necessário abstrai-los das características de ambientes específicos. Nenhum grupo de pesquisa isolado é capaz de prover o ambiente de laboratório necessário para aprender sobre variações nos efeitos das tecnologias através de múltiplos ambientes e fatores de influência (ISERN, 1993). Logo, por meio de replicações de estudos é possível tratar os diferentes conjuntos de condições a que a aplicação de uma técnica, método ou ferramenta está sujeita, para que as conclusões obtidas atinjam um contexto abrangente e, então, se consolide

conhecimento em Engenharia de Software.

Replicar um estudo depende fortemente da revisão efetiva de sua execução original, para compreender os procedimentos adotados e garantir que a replicação tenha conformidade de processo com o experimento anterior (BASILI et al., 1999; SHULL et al., 2002), isto é, que sejam comparáveis para que se viabilize a integração em um corpo comum de conhecimento. O conhecimento gerado por um estudo, incluindo a descrição dos procedimentos adotados, seus resultados e suas conclusões, é registrado no pacote de laboratório (SHULL et al., 2002). É a partir de sua revisão que pesquisadores, internos a um grupo de pesquisa ou não, têm acesso a como o experimento original foi conduzido (SHULL et al., 2004).

Contudo, mesmo que estejam especificados, os pesquisadores enfrentam dificuldades em entender e revisar pacotes de laboratório (SHULL et al., 2002). Essas complicações na transferência de conhecimento entre grupos de pesquisa, visto que o pacote de laboratório é o meio de transferi-lo, dificultam a realização de replicações. Com relação a esses problemas, vale ressaltar que compartilhar conhecimento de um domínio de interesse é uma das aplicações recorrentes a que ontologias se destinam (GUARINO et al., 2009; USCHOLD e GRUNINGER, 1996).

Ontologias vêm sendo utilizadas em Engenharia de Software com a finalidade de padronização de conceitos de um domínio, de compartilhamento de conhecimento e de reuso (BARBOSA et al., 2006; BARBOSA et al., 2008; BIOLCHINI et al., 2007; FALBO et al., 2005). Com esses mesmos objetivos, GARCIA et al. (2008) apresentaram a *ExperOntology*, uma ontologia do domínio de experimentos controlados, visando a descrever formalmente os conceitos que compõem um pacote de laboratório. Assim, as informações relativas a um estudo podem ser instanciadas nos conceitos da ontologia.

De acordo com AMARAL E TRAVASSOS (2003), o empacotamento deve ser feito fase a fase, ao longo do processo experimental (WOHLIN et al., 1999). Então, nesse contexto, é proposto um *workflow*

que incorpora o empacotamento baseado nos conceitos definidos na *ExperOntology*. E, uma vez que suas definições não são necessariamente estáticas, a ontologia pode absorver novos conceitos, conforme pacotes de laboratório com diferentes conjuntos de informação são instanciados.

O restante do artigo está organizado como segue. Na Seção 2 é dada uma visão geral de processo de experimentação. Na Seção 3 é apresentado o problema na transferência de conhecimento entre grupos de pesquisa. Com o objetivo de resolver tal problema, na Seção 4 argumenta-se a vantagem de se aplicar uma ontologia para representar o conhecimento gerado por um estudo. Na Seção 5, as definições da *ExperOntology* são apresentadas. Na Seção 6 é apresentado um *workflow* de empacotamento de experimentos controlados com base na *ExperOntology*. Por fim, na Seção 7 são dadas as considerações finais.

1. PROCESSO DE EXPERIMENTAÇÃO

O experimento controlado é um tipo de estudo em que acontece a investigação de uma hipótese, supondo uma relação causal entre características de interesse do fenômeno sob estudo. Essa relação causal que se supõe em teoria é mapeada às variáveis do estudo para representá-la em sua execução. As variáveis independentes, representando a causa, influenciam nos valores observados nas variáveis dependentes, o efeito. Variáveis independentes são aquelas que medem as características escolhidas para representar os fatores que influenciam o fenômeno. As variáveis dependentes recebem os valores dos resultados, cuja análise permite verificar a validade da hipótese.

Uma abordagem sistemática deve ser adotada com a finalidade de alcançar o nível de controle que um experimento requer. A seleção e manipulação de variáveis devem ser cautelosas a fim de refletirem resultados que levem a conclusões corretas. Apenas pelo fato de definir um experimento de maneira mais formal e explícita, a compreensão do fenômeno sob estudo aumenta e, assim, diferentes aspectos podem emergir a partir de uma melhor introspec-

ção (BASILI, 2006). O processo sugerido por WOHLIN et al. (1999) é composto pelas seguintes fases (vide Figura 1):

- Definição: a fase de definição fornece a direção geral do experimento, o seu escopo, contexto e objetivos.
- Planejamento: nessa fase as hipóteses que expressam o fenômeno sob estudo são formuladas. Em seguida são selecionadas as variáveis independentes e dependentes. Também é estabelecido o projeto experimental, em que são definidos os valores das variáveis independentes (tra-

tamentos) atribuídos a cada participante.

- Operação: a operação consiste basicamente na aplicação de um conjunto de tarefas para executar o que ficou definido no projeto experimental. Abrange a execução propriamente dita e a coleta de seus resultados.
- Análise: os resultados coletados da execução são preparados e testes estatísticos são aplicados com a finalidade de investigar a validade das hipóteses e obter conclusões sobre o fenômeno em estudo.
- Empacotamento: durante o em-

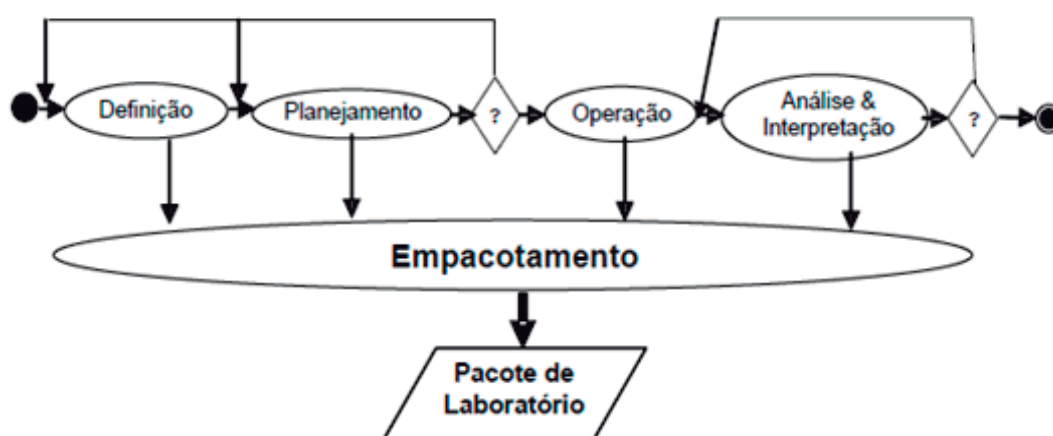


Figura 1: Processo de experimentação

pacotamento, a maneira como o experimento foi conduzido, os resultados e as conclusões devem ser registrados para uso futuro, gerando o pacote de laboratório.

O pacote de laboratório mantém o conhecimento gerado pelo estudo e, por isso, a sua disponibilidade é fundamental para que replicadores possam entender os procedimentos adotados ao longo do processo de experimentação do estudo original. Esse entendimento é fundamental para a formação de um corpo de conhecimento, pois permite não só a replicação do experimento propriamente dita, mas também identificar potenciais variações que contribuam para a generalização de resultados e aprimoramentos no experimento original.

2. REPLICAÇÕES E TRANSFERÊNCIA DE CONHECIMENTO

No estudo de uma técnica, um método ou uma ferramenta em um experimento controlado, são selecionados participantes de uma população, que os aplicam sob condições controladas, de acordo com o que foi definido no planejamento pelos pesquisadores responsáveis. A partir da análise dos resultados são tiradas conclusões sobre o uso da tecnologia sob investigação pela população da qual os participantes são considerados representativos.

Em várias partes do processo (descrito na seção anterior) há fontes de variação que influenciam nos resultados (SHULL et al., 2002; SHULL et al., 2004). Os participantes do estudo podem ser de diversos ambientes culturais ou podem estar submetidos a um diferente conjunto de condições durante a execução do experimento (MENDONÇA et al., 2008; MILLER, 2005). E, deste modo, a população à qual são apli-

cadadas as conclusões fica limitada por esses fatores.

Portanto, a fim de generalizar as conclusões de maneira abrangente, essas variações devem ser exploradas e tratadas em replicações. Assim, os resultados podem ser confirmados ou a influência dessas variações pode ser identificada e produzir outras conclusões relevantes ao assunto. Por meio de replicações baseadas em um experimento que investiga determinada tecnologia, o conhecimento sobre ela se consolida.

Empreender uma replicação requer o acesso a informações contidas no pacote de laboratório. Contudo, SHULL et

al. (2002) apontam que, apesar de especificados, a revisão de pacotes de laboratório apresenta dificuldades.

Ao notar problemas de transferência de conhecimento como entraves à realização de replicações, MENDONÇA et al. (2008) propuseram o FIRE (Framework for Improving the Replication of Experiments), o qual é composto pelos dois ciclos de atividades ilustrados na Figura 2. O ciclo interno representa a execução de um estudo dentro de um grupo de pesquisa. No ciclo externo, inter-grupos, estão as atividades que se encarregam de integrar o conhecimento gerado pelo estudo do ciclo interno num corpo comum, de tal maneira que seu pacote de



Figura 2: Ciclos do FIRE (MENDONÇA et al., 2008)

laboratório possa ser efetivamente revisado por eventuais replicadores de outro grupo de pesquisa.

Criar/evoluir o pacote de laboratório influencia na atividade externa de compartilhar conhecimento. Da mesma maneira, no outro ponto em que os ciclos se interceptam, entender os pacotes de laboratório é fundamental para se estabelecer as metas de uma nova replicação.

Pode-se perceber, portanto, que o pacote de laboratório representa a saída do ciclo interno portando consigo as informações de um estudo e também a entrada do ciclo interno de uma possível replicação,

uma vez que os pesquisadores devem revisá-lo e entender a execução do experimento original. Assim, de maneira a compartilhar conhecimento e favorecer um melhor entendimento dos pacotes de laboratório (no ciclo externo), o modo que as suas informações devem ser representadas é de extrema importância, o que se confirma também pela atividade de padronização de pacotes.

Disponibilizar pacotes de laboratório ampara a realização de mais replicações, mas pacotes bem projetados são cruciais para melhores replicações (SHULL et al., 2004). Portanto, deve ser promovida uma melhor representação ao gerar o seu

conteúdo. Nesse sentido, a aplicação de uma ontologia como modelo para o pacote de laboratório pode ser proveitosa para lidar com o problema na transferência de conhecimento.

3. EXPERONTOLOGY

Uma ontologia define o domínio de interesse explicitamente em termos de conceitos, suas relações e restrições de interpretação expressas por meio de axiomas (GRUBER, 1995). Assim, a informação modelada pela ontologia carrega consigo significado embutido, permitindo que agentes de diferentes contextos a interpretem de maneira consistente.

A *ExperOntology* (GARCIA et al.,

2008) foi elaborada para facilitar o reuso e o compartilhamento de pacotes de laboratório. Basicamente, a ontologia esclarece os conceitos associados a experimentos controlados e suas relações. São definidos conceitos de todo o processo de experimentação, desde a definição até a análise dos resultados. Assim, o empacotamento de um estudo é dado pelos conceitos da ontologia instanciados.

A captura da conceituação da *ExperOntology*, contida parcialmente nas figuras 3 e 4, é apresentada por GARCIA et al. (2008) em dois níveis de refinamento. O primeiro nível (Figura 3) se refere aos principais conceitos de um experimento controlado. O segundo (Figura 4) se refere ao refinamento do conceito de pacote de

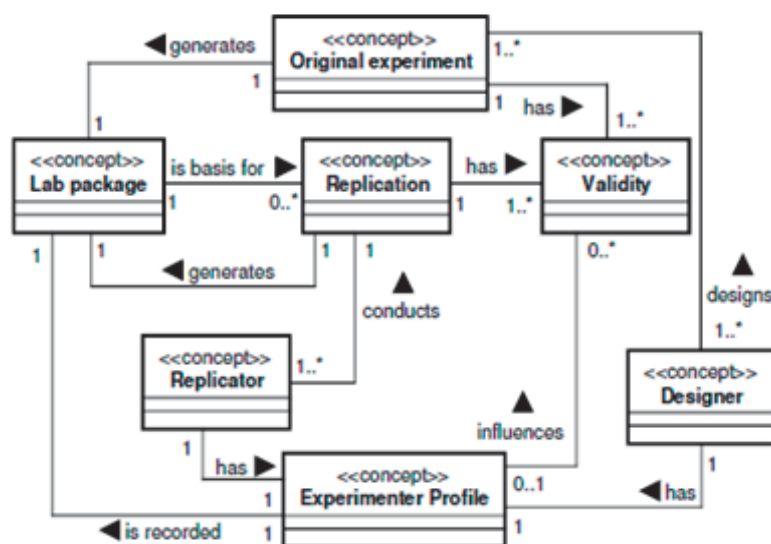


Figura 3: Conceitos de um experimento controlado (GARCIA et al., 2008)

laboratório. A captura dos conceitos foi representada usando UML (Unified Modeling Language).

Para formalizar a ontologia, foram usados axiomas de lógica de primeira ordem. O predicado *Design*, que associa tratamentos a cada participante do experimento, é dado por:

$$Design(s, SetOfTreat)$$

em que *s* representa os participantes e

$$\begin{aligned} &Factor(f_1, \dots, f_n) \\ &\forall f, \exists Treatment(f) = \{v_1, \dots, v_n\}, n \geq 2 \\ &dom(Treatment) = Artifact \cap Technology \\ &SetOfTreat = (vf_1, \dots, vf_n) \mid \forall f, vf_n \in Treatment(f_n) \end{aligned}$$

em que *f* define um fator do experimento que *f* define um fator de experimento, e *vf_n* são tratamentos atribuídos para cada *f*. A partir do predicado *Design* é estabelecido o plano para a condução do experimento, o qual é detalhado em um conjunto de tarefas:

Como ilustrado na Figura 5, as entradas do *workflow* são o pacote de laboratório a ser instanciado e a *ExperOntology*. A primeira atividade do *workflow* se repete

para cada fase do processo de experimentação: deve ser feita a seleção de informações do pacote de laboratório correspondentes a cada fase.

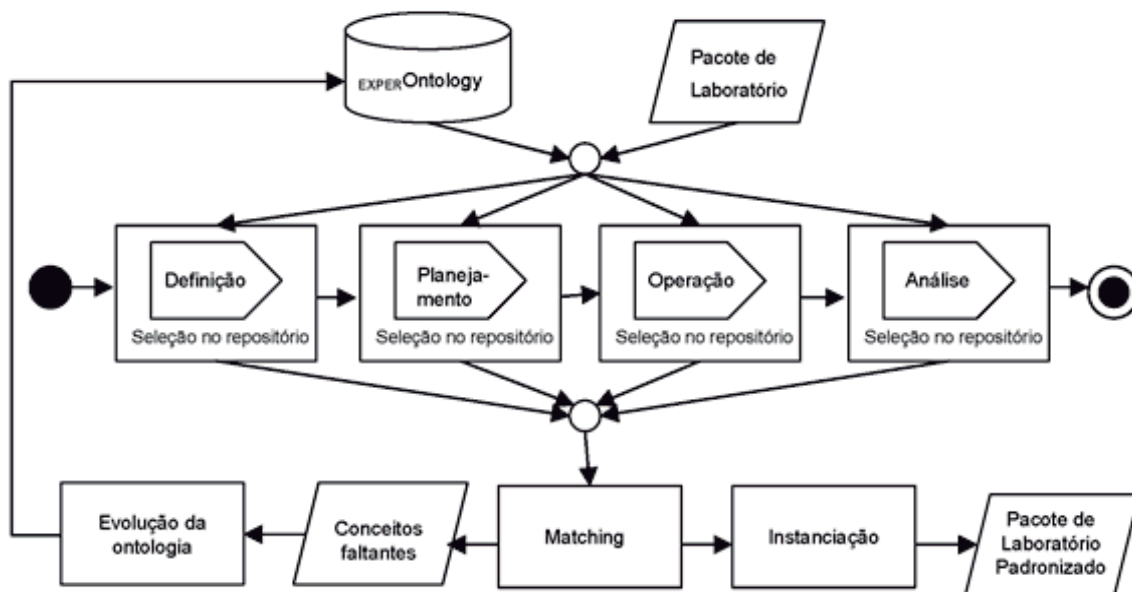


Figura 5: Workflow de empacotamento (SCATALON et al., 2011)

Em seguida, dados os conceitos definidos na *ExperOntology* e as informações selecionadas para uma fase específica do processo, equipara-se (match) conceitos e informações, por meio da aplicação técnicas de matching (EUZENAT e SHVAIKO, 2007). A atividade de matching visa a identificar (i) as informações às quais a *ExperOntology* pode fornecer conceitos definidos para instanciação e (ii) as informações sem conceito correspondente. Tais informações, portanto, indicam os conceitos que devem ser incorporados à ontologia, na atividade de evolução (STOJANOVIC et al., 2002), para possibilitar a instanciação completa, gerando um pacote de laboratório padronizado.

Dessa maneira, a execução iterativa do *workflow* caracteriza um processo de evolução da ontologia, que se aproxima de um padrão semântico do conjunto de informações que um pacote de laboratório deve conter.

Para alcançar os objetivos do trabalho, é preciso identificar qual conceito definido na *ExperOntology* corresponde a

cada elemento de informação do pacote de laboratório a ser padronizado. Cada elemento é instanciado em seu conceito correspondente definido na ontologia, conforme pode ser visto na Figura 6.

Uma vez que suas definições não são necessariamente estáticas, a ontologia pode absorver novos conceitos, conforme pacotes de laboratório contendo diferentes conjuntos de informações são instanciados. Por meio desse mecanismo, é possível evoluir a *ExperOntology* em direção a um padrão semântico que lide com o problema da heterogeneidade de pacotes de laboratório.

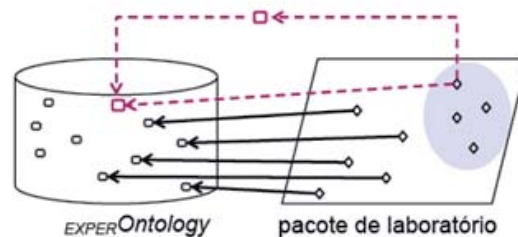


Figura 6: Padronização de um pacote de laboratório

CONSIDERAÇÕES FINAIS

Neste artigo é apresentado um *workflow* de empacotamento de experimentos controlados com base na *ExperOntology*, para padronizar os seus respectivos pacotes de laboratório e para que seja possível acomodar variações nos conjuntos de informação que um pacote de laboratório pode apresentar. Nesse sentido, a *ExperOntology* pode ser vista como um padrão unificador, colaborando, assim, à integração de estudos isolados num corpo comum.

A maneira que o pacote de laboratório fica armazenado na ontologia facilita a transferência do conhecimento nele contido. Por se tratar de uma especificação formal, é possível compartilhar um entendimento comum da estrutura da informação entre os diferentes grupos de pesquisa (NOY e MCGUINNESS, 2001). Dessa maneira, a ontologia de instâncias de um pacote de laboratório gerado por um determinado grupo de pesquisa, pode ser mais facilmente processado em outro grupo, de um contexto diferente do grupo de origem. Os pacotes de laboratório não seriam tão portáteis se fossem instanciados em bancos de dados relacionais, por exemplo.

Ao implementar o *workflow*, deve-se considerar os diversos formatos em que um pacote de laboratório de entrada pode estar representado. Entre os formatos possíveis, pode-se citar os estruturados (tais como as bases de dados), os semi-estruturados (os documentos XML, por exemplo) e os documentos não estruturados (informações descritas em linguagem natural).

A fim de estabelecer uma solução que abranja desde formatos estruturados até não estruturados, pode-se adotar como formato padrão a linguagem XML. Assim, todos os pacotes de laboratório de entrada devem ser mapeados para um documento XML a fim de se tornar uma entrada do *workflow*. No caso de uma base de dados, por exemplo, pode ser feito um mapeamento direto para o formato XML, já que as informações se encontram segmentadas em elementos de informação (tabelas com seus respectivos atributos).

No contexto do FIRE, a atividade criar/evoluir pacote sugere que a ontologia

deve evoluir, considerando que a *ExperOntology* é tomada como modelo para o pacote de laboratório. A proposta deste trabalho também apóia as atividades do ciclo externo do FIRE, pois busca um padrão que acomode os diferentes conjuntos de informação visando à integração do conhecimento de um estudo isolado em um corpo comum. E, conforme pacotes de laboratório são instanciados usando o *workflow* proposto, a *ExperOntology* se aproxima de um padrão semântico que permite acomodar variações de pacotes de laboratório. Desse modo, a evolução da *ExperOntology* também influencia na sua aplicação em empacotar experimentos.

REFERÊNCIAS BIBLIOGRÁFICAS

- AMARAL, E. A. G. G.; TRAVASSOS, G. H. A package model for software engineering experiments. In: Proceedings of ISESE 2003 – International Symposium on Empirical Software Engineering, 2003. p. 21-22.
- BARBOSA, E. F.; NAKAGAWA, E. Y.; MALDONADO, J. C. Towards the establishment of an ontology of software testing. In: International Conference on Software Engineering and Knowledge Engineering (SEKE 2006). 2006. p. 522-525.
- BARBOSA, E. F. et al. Ontology-based development of testing related tools. In: International Conference on Software Engineering and Knowledge Engineering (SEKE 2008). 2008. p. 697-702.
- BASIL, V. R. The role of controlled experiments in software engineering research. In: Empirical Software Engineering Issues: Critical assessment and future directions. 2006. p. 33-37.
- BASIL, V. R.; SHULL, F.; LANUBILE, F. Building knowledge through families of experiments. IEEE Transactions on Software Engineering, 1999. v. 25, n. 4, p. 456-473, 1999.
- BIOLCHINI, J. et al. Scientific research ontology to support systematic review in software engineering. Advanced Engineering Informatics, 2007. v. 21, n. 2, p. 133-151, 2007.
- EUZENAT, J.; SHVAIKO, P. Ontology Matching: Springer-Verlag, 2007.
- FALBO, R. de A.; RUY, F. B.; MORO, R. D. Using ontologies to add semantics to a software engineering environment. In: SEKE (Software Engineering & Knowledge Engineering). 2005. p. 151-156.
- GARCIA, R. E. et al. An ontology for controlled experiments on software engineering. In: SEKE (Software Engineering & Knowledge Engineering).

GRUBER, T. R. Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies - Special issue: the role of formal ontology in the information technology*, 1995. Academic Press, Inc., Duluth, MN, USA, v. 43, p. 907-928, December 1995.

GUARINO, N.; OBERLE, D.; STAAB, S. *Handbook on ontologies*. Springer Verlag, 2009. cap. What is an ontology?, p. 201-221.

HORRIDGE, M.; BECHHOFFER, S. The OWL API: A Java API for working with OWL 2 ontologies. In: *OWL: Experiences and Directions (OWLED)*. 2009. ISERN. ISERN Manifesto. 1993. Available at <http://isern.iese.de/>.

MENDONÇA, M. G. et al. A framework for software engineering experimental replications. In: *ICECCS*. 2008. p. 203-212.

NOY, N. F.; MCGUINNESS, D. L. *Ontology Development 101: A Guide to Creating Your First Ontology*. 2001.

SCATALON, L. P.; GARCIA, R. E.; CORREIA, R. C. M. Packaging controlled experiments using an evolu-

tionary approach based on ontology. In: *International Conference on Software Engineering and Knowledge Engineering (SEKE 2011)*. 2011. p. 408-413.

SHULL, F. et al. Replicating Software Engineering Experiments: Addressing the Tacit Knowledge Problem. 2002. p. 7-16.

SHULL, F. et al. Knowledge-sharing issues in experimental software engineering. *Empirical Softw. Engg.*, 2004. Kluwer Academic Publishers, Hingham, MA, USA, v. 9, n. 1-2, p. 111-137, mar. 2004.

STOJANOVIC, L.; STOJANOVIC, N.; HANDSCHUH, S. Evolution of the metadata in the ontology-based knowledge management systems. In: *Proceedings of the 1st German Workshop on on Experience Management: Sharing Experiences about the Sharing of Experience*. 2002. p. 65-77.

USCHOLD, M.; GRUNINGER, M. *Ontologies: Principles, methods and applications*. *Knowledge Engineering Review*, 1996. v. 11, p. 93-136, 1996.

WOHLIN, C. et al. *Experimentation in Software Engineering: An introduction*. Boston, USA: Kluwer Academic Publishers, 1999.

Lilian Passos Scatalon é bacharel em Ciência da Computação pela Universidade Estadual Paulista “Júlio de Mesquita Filho” (2010). É aluna regular do mestrado em Ciência da Computação na Unesp, atuando principalmente nos seguintes temas: Engenharia de Software Experimental, Representação de Conhecimento e Ontologia.

Rogério Eduardo Garcia concluiu o doutorado em Ciência da Computação e Matemática Computacional pela Universidade de São Paulo (ICMC-USP-São Carlos) em 2006. Atualmente é Professor Assistente da Universidade Estadual Paulista “Júlio de Mesquita Filho”, lotado no Departamento de Matemática e Computação da Faculdade de Ciência e Tecnologia de Presidente Prudente. Atua na área de Ciência da Computação, com ênfase em Metodologia e Técnicas da Computação, principalmente nos seguintes temas: Engenharia de Software e Visualização de Informação e Científica.