

IACS – Identificação Automática de Componentes de Software

Marcílio Oliveira¹⁻², José Vahl¹⁻², Kleber Bacili¹

¹ Laboratório de Inovação– DigitalAssets / Ci&T
Inovasoft - Unicamp

² Instituto de Computação – Universidade Estadual de Campinas (UNICAMP)
Caixa Postal 6176 – 13.084-971 – Campinas – SP – Brasil

{marcilio.oliveira};{jose.vahl};{kleber.bacili}@digitalassets.com.br

***Abstract.** Software reuse is seen as one of the main alternatives to increase productivity in the development of new applications. The reuse of legacy assets plays a vital role anticipating the ROI (Return on Investment) on SOA (Service Oriented Architecture) and reuse enterprise programs. This work presents a tool that implements an Automatic Identification of Software Components (AISC). AISC is an approach that brings to light what companies have already developed by applying reuse indicators with sophisticated mechanisms to identify artifacts that can be considered as reusable assets. Thus, they will have the potential of being reused in new applications, avoiding redevelopment of already existing features, enabling savings and increasing agility.*

***Resumo.** O reuso de software é uma das principais alternativas para aumentar a produtividade no desenvolvimento de aplicações. O reuso de ativos legados tem papel fundamental na antecipação do ROI (Retorno n Investiment) dentro dos programas de SOA (Service Oriented Architecture) e reuso. Este trabalho apresenta uma ferramenta que implementa a Identificação Automática de Componentes de Software (IACS). IACS é uma abordagem que destaca o que as companhias já desenvolveram, aplicando indicadores de reuso aliados a mecanismos sofisticados para identificar artefatos que podem ser considerados ativos reutilizáveis. Com isso, esses ativos terão potencial para serem reutilizados em novas aplicações, evitando desenvolvimento duplicado de funcionalidades já existentes, aumentando a agilidade e proporcionando economia. O processo de varredura e análise de novos ativos é feito visualmente por meio de um gráfico interativo dos resultados e de um mecanismo de exportação dos ativos identificados em um Modelo de Representação de Metadados largamente utilizado.*

1. Introdução

O reuso de software em desenvolvimento de sistemas é uma estratégia que proporciona melhoria na produtividade e na qualidade dos produtos [Boehm99]. O Desenvolvimento Baseado em Componentes (DBC) é um paradigma alinhado com esta estratégia, uma vez que ela é voltada para reuso de partes previamente construídas (componentes), construção de componentes e suporte à manutenção e melhoria dos produtos mediante substituição e adaptação destes componentes.

São notáveis as pesquisas nos últimos anos em ferramentas que promovam o reuso de tais ativos de software. Considerando em termos intra-organizacionais, temos

exemplos de repositórios de ativos digitais como o DigitalAssets Manager [Bacili06] e ambientes de desenvolvimento integrados a estes repositórios. Já em termos interorganizacionais, temos o padrão XML Web Services e redes Peer-to-Peer de compartilhamento de ativos como o RCCS [Oliveira05].

Em meio a essa tendência de maximizar o potencial de reuso, sistemas legados são aplicações já existentes, normalmente de missão crítica e ainda em funcionamento, representando um aspecto crucial a ser considerado. As principais características dos sistemas legados são o tamanho, em média, de milhões de linhas de código; encapsulam lógicas de negócio significativas; e provavelmente apresentam uma queda ou desvio acentuado de qualidade do projeto original diante de modificações em atividades de correção ou evolução do sistema [Zou03].

O Projeto IACS aborda tecnologias de identificação automática de componentes de software nesses sistemas já existentes, bem como mecanismos e padrões para a representação destes componentes e outros ativos digitais de forma flexível e interoperável em meta-modelos de dados. A identificação automática de componentes tem três aplicações principais: Reengenharia de sistemas legados, Convergência de aplicações para um novo paradigma de orientação a serviços (SOA) e Reuso de componentes.

O foco central do projeto consistiu em montar uma heurística de identificação através de um mecanismo capaz de reconhecer, nos parques de aplicações já desenvolvidas, grupos de artefatos que componham componentes. Visando, sobretudo, a extração destes componentes para catalogação e reutilização, gerando economia de investimento e evolução tecnológica/arquitetural.

Por meio de pesquisas desenvolvidas em laboratório, foram montados mecanismos capazes de identificar o agrupamento de artefatos relacionados, sugerindo ativos de software (componentes, serviços, etc.) candidatos à reutilização. Conforme ilustrado na figura a seguir, a ferramenta realiza diferentes etapas desde a leitura das aplicações legadas até a catalogação dos ativos reutilizáveis identificados em repositórios de componentes.

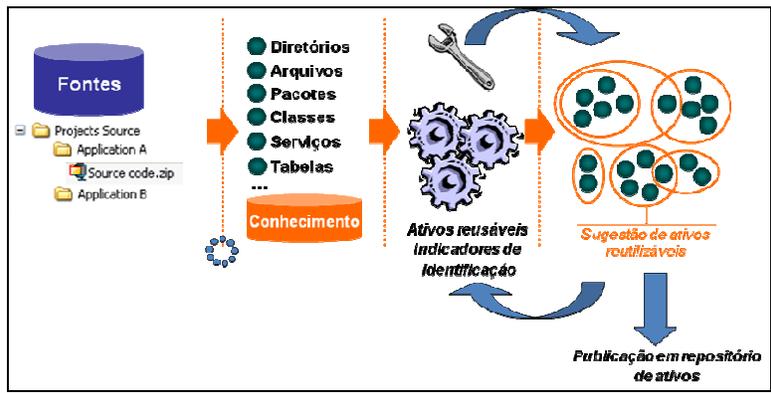


Figura 1: Mecanismo de mineração de ativos do IACS

O mecanismo de identificação de componentes de software permite realizar a reengenharia dos sistemas legados em sistemas baseados em componentes, identificando partes reutilizáveis do sistema legado e permitindo reestruturá-lo em torno destas peças.

2. Objetivo e justificativa

O foco central do projeto IACS consiste no **desenvolvimento de um mecanismo capaz de reconhecer, em parques de aplicações já desenvolvidas, grupos de artefatos que componham ativos reutilizáveis de software, serviços web e componentes. Visando, sobretudo, a extração destes componentes para catalogação e reutilização.**

O principal objetivo é evitar o re-trabalho no desenvolvimento de novos projeto, capitalizando trabalhos anteriores, fazendo com que as soluções já desenvolvidas sejam imediatamente implementadas em novos contextos. Desta forma, tem-se melhores produtos em um menor intervalo de tempo, com redução nos custos de manutenção pois as partes do sistema são independentes e facilitando a inclusão de novas funcionalidades. Além disso, a qualidade aumenta devido ao reuso de componentes previamente bem testados [D'Souza 99] [Jacobson97].

Para que seja possível reusar algo, é necessário identificar o que pode ser reutilizado. É sob esta perspectiva que este trabalho foi desenvolvido. O principal desafio foi desenvolver uma arquitetura baseada em padrões que permitam a fácil integração com outras ferramentas como a RCCS (Rede de Compartilhamento de Componentes de Software). A RCCS é o projeto de uma rede Peer-to-Peer, baseada em padrões abertos e que pode ser utilizada por qualquer pessoa ou entidade interessada em buscar ou compartilhar componentes. A RCCS foi contemplada com o prêmio Dorgival Brandão Júnior Qualidade de Software, no ciclo 2006 durante o Encontro de Qualidade e Produtividade de Software (EPQS).

3. Metodologia de execução

A identificação de ativos reutilizáveis a partir de aplicações existentes usando o DigitalAssets Discoverer ocorre em quatro estágios, como demonstrado na Figura 1.

1. **Varredura de aplicações existentes:** Aplicações existentes são selecionadas, e o processo de varredura nas fontes especificadas seleciona os artefatos primários (arquivos-fonte, bibliotecas, etc.) que integram a aplicação. Estes artefatos serão os insumos para o estágio de criação da base de conhecimento.
2. **Criação da Base de Conhecimento:** A partir dos arquivos-fonte da aplicação encontrada na varredura, analisadores estáticos de código geram a base de conhecimento das aplicações cadastradas no estágio anterior.
3. **Configuração e execução dos indicadores:** Indicadores são as heurísticas que analisam a base de conhecimento e através de critérios apoiados em práticas de programação como por exemplo modularização; nomenclatura; padrões de projeto; arquitetura de software; coesão e acoplamento; e desenvolvimento baseado em componentes; identificam partes de código com potencial para reutilização. Em um processo iterativo, calibrações dos indicadores e execuções podem ser executadas com o objetivo de alcançar resultados otimizados.
4. **Colheita:** Os ativos sugeridos pelos indicadores como candidatos a componentes é apresentado. O analista pode navegar pelos resultados, decidir quais sugestões são relevantes, possivelmente adaptar as sugestões às suas necessidades.

4. Resultados

4.1. Pedidos de patente

O projeto IACS deu origem a um produto comercial batizado de DA Discoverer absorvido no portfólio de produtos da empresa DigitalAssets. O DigitalAssets Discoverer está em processo de encaminhamento do Pedido de Registro de Programa de Computador através do INPI (Instituto Nacional de Propriedade Intelectual) com os direitos patrimoniais atribuídos a DigitalAssets.

4.2. Produtos de software gerados

O DigitalAssets Discoverer é uma ferramenta de apoio a analistas e consultores dentro de iniciativas de modernização de aplicações, reuso e SOA. Tendo isso em mente, o fluxo de trabalho apresentado na seção *Metodologia de execução* foi implementado e está mapeado na interface gráfica do DigitalAssets Discoverer. Outras funcionalidades não relacionadas com o referido fluxo, mas importantes no aspecto qualitativo do software também estão presentes.

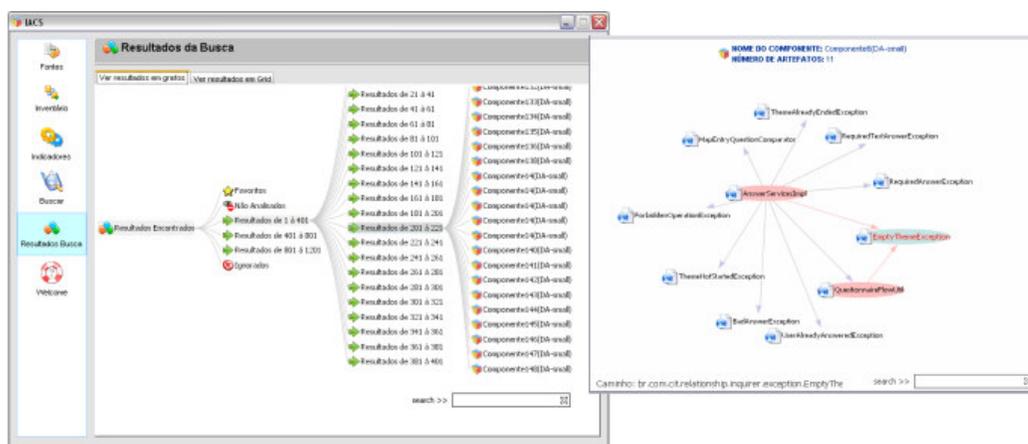


Figura 2: DigitalAssets Discoverer (interfaces de navegação nos componentes)

4.3. Outros produtos gerados (que foram disponibilizados para o mercado)

Um dos subprodutos gerados pelo projeto e eu agrega grande valor é o módulo de integração com repositório de metadados via padrão de mercado RAS. Esse módulo foi utilizado para integrar o IACS ao DigitalAssets Manager – principal ferramenta de repositório de metadados do mercado nacional.

A integração com um produto comercial deu grande visibilidade ao projeto, além de agregar valor ao negócio dos clientes do DigitalAssets Manager.

4.4. Métodos e/ou algoritmos desenvolvidos

No contexto desse projeto, *indicadores* são heurísticas capazes de identificar ativos de software com potencial para a reutilização. Portanto, os indicadores são partes fundamentais da ferramenta proposta. Muitas abordagens para identificação de componentes e outras atividades na área de engenharia reversa ou reengenharia de sistemas legados são encontrados na literatura. Foram estudadas e validadas diferentes

abordagens já publicadas de identificação de grupos de artefatos que sugerem sua utilização e reutilização conjunta.

Os indicadores selecionados para serem implementados no projeto foram:

- **Coesão e acoplamento:** Definimos a metodologia de Coesão e Acoplamento como sendo a análise estrutural (análise de artefatos e interdependências) a fim de encontrar módulos de software que compõem o sistema e caracterizados como grupos significativos (alta coesão) e independentes (baixo acoplamento). Este tipo especial de *clustering* foi explorado em [Mitchell06] com o intuito de encontrar automaticamente abstrações, como subsistemas, a fim de auxiliar o especialista na tarefa de entender a estruturação do código.
- **Análise de dominância:** Análise de Dominância é um percurso em grafos dirigidos de forma a capturar a relação de dominação entre nós, por exemplo, se um nó N é dito dominar outro nó M em um grafo dirigido G, todo caminho da raiz de G até M contém N. A árvore de dominância é uma possível representação dos relacionamentos entre um nó e seu dominador imediato. A aplicação da análise de dominância em estruturas de sistemas e código-fonte existentes já foi tema de metodologias e implementações. Em [Buschsbaum01] podemos encontrar uma proposta de ferramenta que usa esta análise tanto para avaliar qualitativamente a complexidade do sistema quanto para avaliar o impacto de evoluções nos demais subsistemas envolvidos.
- **Matriz de Estrutura das Dependências (Dependency Structure Matrix – DSM):** A Matriz de Dependências é usada para modelagem e análise de sistemas que envolvam entidades como tarefas, pessoas, etc., e relacioná-las informando o fluxo de informações entre as entidades [Yassine04]. Além de aplicações em várias áreas da engenharia de processos, DSM é útil em análises de sistemas complexos e para representação das interdependências entre os elementos de software. Devido ao grande volume em potencial de informações extraídas destes sistemas, ferramentas são empregadas para que o analista possa especificar um projeto, modelá-lo em DSM e executar automaticamente algoritmos que determinam agrupamentos ou seqüências dos elementos do DSM que melhor caracterizam o fluxo de informações.

4.5. Artigos publicados

O projeto IACS foi mencionado em diversos eventos como SBCARS 2007 (Simpósio Brasileiro de Componentes e Reuso de Software), SBQS 2007 (Simp. Brasileiro de Qualidade de Software), Wire 2007 (Workshop de Introdução de reuso em empresas), 2º Minds-ON (2º Fórum de Inovação DigitalAssets), dentre outros. Além desses, o trabalho originou publicações internacionais bastante relevantes (IEEE e ACM):

- ✓ A ferramenta foi apresentada e demonstrada em um dos principais eventos de reuso de software do mundo, o IRI (*The 2007 IEEE International Conference on Information Reuse and Integration – IEEE - Las Vegas/USA*). Sob o título: “Automatic Identification of Reusable Software Development Assets: Methodology and Tool”, [Goncalves07].
- ✓ A ferramenta também foi apresentada no evento internacional da ACM, o OOPSLA (*22st Annual ACM Conference on Object- Oriented Programming*,

Systems, Languages, and Applications, Montréal, Canada), sob o título: “DigitalAssets Discoverer: Automatic Identification of Reusable Software Components”, [Oliveira07a]. Este evento é considerado um dos mais importantes do mundo na área de orientação de objetos e aplicações. Foi possível realizar exibições práticas do mecanismo do IACS realizando identificação de componentes em sistemas.

- ✓ Além de citações e publicações relacionadas, como em [Oliveira07b], “DA Manager®, gerência e avaliação da reutilização de ativos digitais”.

O perfil prático do projeto pôde ser afirmado através das publicações, que foram, em sua maioria, em salões de demonstração. Nestes salões foi possível instalar e permitir a iteração dos participantes do evento com a ferramenta.

4.6. Recursos humanos capacitados (especialistas, mestres, doutores, etc.)

O projeto contou com pesquisadores do Instituto de Computação da Unicamp, consultores e gerentes de projeto da DigitalAssets e da Ci&T e com alunos de graduação e mestrado. Um dos principais resultados do projeto foi a capacitação de pessoal, gerando especialistas em nas tecnologias relacionadas ao projeto, sobretudo em SOA (*Service Oriented Architecture*), arquiteturas Peer-to-Peer, além de um mestre em ciência da computação, que teve a RCCS como tema de projeto de mestrado.

4.7. Eventuais parcerias ou programas de transferência de tecnologia efetuados

O Laboratório tem parceria com diversas instituições de pesquisa, e vêm realizando projetos em conjunto com C.E.S.A.R., UFPE, UNICAMP, UFPB, IBTA, CenPRA e UFV. O projeto da IACS foi desenvolvido juntamente com a Unicamp, e algumas ferramentas (como o mecanismo de visualização) foram desenvolvidas em laboratórios da Ci&T Labs.

5. Aplicabilidade dos resultados

Atualmente o setor de tecnologia sofre uma pressão que força a divisão de investimentos e o alinhamento de iniciativas de TI com as estratégias de negócio. Para lidar com esses desafios as organizações de tecnologia estão buscando uma variedade de iniciativas de desenvolvimento de software para atingir estandarização, componetização, reutilizações de ativos de software, estratégias de arquitetura orientada a serviços (SOA) e redução de custos. Esse é um mercado de novos paradigmas no desenvolvimento de software e abordagens de Arquiteturas Orientadas a Serviços (SOA) e Reutilização de Software serão formas primordiais para se atingir elevados patamares de produtividade e agilidade frente a mudanças.

Michael Blechar, analista do Gartner, declarou que “as empresas que tiram vantagem da reutilização de software conseguem aumentar produtividade, qualidade e entrega ao mercado em um fator de 5 para 1”. Mas isso só é possível quando times de projeto conseguem localizar e reutilizar facilmente os ativos digitais já previamente desenvolvidos e testados.

O conjunto de aplicativos legado no mercado geral de TI é extremamente significativo e de suma importância para qualquer projeto que envolva um conjunto

maior de ações. O projeto IACS, aplicado a esse contexto, permite a antecipação do retorno de investimento e trazendo luz aos investimentos já realizados.

Ainda no escopo do projeto IACS, também foi implementado um mecanismo de empacotamento de ativos em modelo RAS (Export-RAS). Os principais gerenciadores de ativos internacionais utilizam este padrão para intercâmbio de componentes, e essa funcionalidade alavanca as ferramentas nacionais a um padrão internacional de representatividade de ativos, que muitas vezes é requisito obrigatório para escolha de ferramentas.

Esse mecanismo de empacotamento permitiu integrar o IACS ao DigitalAssets Manager, a principal ferramenta nacional de gerenciamento de ativos digitais (prêmio B2B de Qualidade em 2005 e 2006 e prêmio IBM Best Choice em 2004 e 2005) [Bacili06]. Essa integração promove uma maior visibilidade do acervo já existente envolvendo e viabilizando a cooperação entre os diversos papéis envolvidos no processo de reuso, permitindo a definição dos processos de reutilização de código assim como assim como definição das políticas de acompanhamento e seus indicadores que irão, por fim, culminar em serviços com qualidade superior e agilidade suficiente para que empresas possam acompanhar as exigências do mercado atual.

6. Características inovadoras

Os principais tópicos de pesquisa relacionados são:

- ✓ Metodologias de Representação de Ativos Digitais e Componentes de Software
- ✓ Mecanismos de identificação e classificação de componentes de software
- ✓ Algoritmos de identificação automática de componentes.
- ✓ Ferramentas escaláveis para exibição gráfica da dependência entre artefatos.

Relacionado às características inovadoras, podemos citar:

- ✓ Implementação de um mecanismo totalmente inovador para mineração de componentes de software em parque de aplicações desenvolvida de maneira não componentizada.
- ✓ Criação de um mecanismo de análise arquitetural. O IACS possui um mecanismo para captação da complexidade arquitetural e visualização gráfica da árvore de relacionamento entre os artefatos do projeto, possibilitando uma avaliação da qualidade arquitetural. Em momentos de evolução tecnológica, isso torna-se de grande valor, inclusive em empresas usuárias do IACS.
- ✓ Integração de diversos algoritmos identificadores de possibilidade de reuso. Alguns algoritmos já existiam, mas de forma isolada em ferramentas de engenharia reversa e análise ciclomática. Estes algoritmos foram combinados para uma análise moderna para identificação de conjuntos de artefatos e componentes potencialmente reusáveis.
- ✓ Geração de uma ferramenta prática e com fluxo definido para identificação de componentes. Mesmo nos principais concorrentes internacionais, não existe ferramenta com características semelhantes (isso está, inclusive, disparando um processo de registro de patente - propriedade intelectual - da solução no Brasil e

nos EUA). Vale destacar que a DigitalAssets foi selecionada como uma das 25 empresas que mais inovam no Brasil, resultado da pesquisa “O Brasil que inova”, realizado pelo Monitor Group para a revista Exame (<http://pesquisa-inovacao-exame.monitor.com/resultados.html>). Boa parte das características inovadoras citadas da empresa estão atribuídas ao IACS

7. Conclusão e Perspectivas futuras

O Projeto IACS aborda tecnologias de identificação automática de componentes de software em sistemas já existentes, bem como mecanismos e padrões para a representação destes componentes e outros ativos digitais de forma flexível e interoperável em meta-modelos de dados.

Após o processo de registro de patente a ferramenta tende a ser total diferencial competitivo internacional da empresa. Nacionalmente, as POVs (Provas de Valor) da ferramenta têm tido boa receptividade e está sendo trabalhada uma estratégia de trabalho e serviço para atuação dentro de organizações maiores, visando agregar maior valor à solução e às empresas com grande volume de sistemas legados.

Referências

- [Boehm99] Boehm, B., *Managing Software Productivity and Reuse*, Computer, vol. 32, issue 9, Sept 1999, pp. 111-113.
- [Bacili06] Bacili, K., Oliveira, M., *DigitalAssets Manager: sharing and managing software development assets*, OOPSLA'06 Demo Session, ACM, NY, 2006, pp. 700-701.
- [Buschsbaum01] Buschsbaum, A., Chen, Y.F., Huang, H., Koutsofios, E., Mocenigo, J., Rogers, A. *Visualizing and Analyzing Software Infrastructures*, IEEE Software, Sept./Oct. 2001.
- [D'Souza99] D. F. D'Souza and A. C. Wills. *Objects, Components, and Frameworks With UML: The Catalysis Approach*. Addison Wesley. USA., 1999.
- [Goncalves07] Gonçalves, E. M.; Oliveira, M. S.; Bacili, K. R.; *DigitalAssets Discoverer: Automatic Identification of Reusable Software Components*. 22st Annual ACM Conference on Object- Oriented Programming, Systems, Languages, and Applications OOPSLA- October 21-25, 2007 – Montréal, Canada. Demonstration Tools.
- [Jacobson97] L. Jacobson, M. Griss, and P. Jonsson. *Software Reuse: Architecture, Process and Organization for Business Success*. Addison Wesley. USA., 1997.
- [Oliveira07a] Oliveira, M. S.; Gonçalves, E. M. ; Bacili, K. R.; *Automatic Identification of Reusable Software Development Assets: Methodology and Tool*. The 2007 IEEE International Conference on Information Reuse and Integration IRI2007 - August 13 - 15, 2007 - Las Vegas, USA.
- [Oliveira07b] Oliveira, M. S.; Bacili, K. R.; Vahl JR., J. C.; *DA Manager®, gerência e avaliação da reutilização de ativos digitais*. SBCARS 2007 - Brazilian Symposium on Software Components, Architectures and Reuse, August 29 – 31, 2007. Campinas, SP – Brazil.
- [Mitchell06] Mitchell, B.S., Mancoridis, S., *On the Automatic Modularization of Software Systems Using the Bunch Tool*, IEEE Trans. on Soft. Eng., vol. 32, no. 3, March 2006.
- [Oliveira05] M. S. Oliveira, I. Garcia, e A. Nunes. *RCCS, Rede de Compartilhamento de Componentes de Software*. X Salão de Ferramentas - XXIII Simpósio Brasileiro de Redes de Computadores - Fortaleza, CE - maio de 2005, 2005.
- [Yassine04] Yassine, A. *An Introduction to Modeling and Analyzing Complex Product Development Processes Using the Design Structure Matrix (DSM) Method*, Quaderni di Management, no. 9, 2004, english version.
- [Zou03] Zou, Y., *Techniques and Methodologies for the Migration of Legacy Systems to Network Centric Environments*, Ph.D. Thesis, Department of Electrical & Computer Engineering, University of Waterloo, Ontario, Canada, Sept 2003.