



Universidade do Minho
Escola de Engenharia
Departamento de Informática

SIMPÓSIO DOUTORAL 2006

Desenvolvimento e aplicações de um cálculo de
programas baseado em conexões de Galois

Identificação

Doutorando

Paulo Filipe Araújo da Silva (paufil@di.uminho.pt)

Título da Tese

Desenvolvimento e aplicações de um cálculo de programas baseado em conexões de Galois

Orientadores

Prof. Doutor José Nuno Oliveira (Universidade do Minho, Departamento de Informática, jno@di.uminho.pt)

Data Início

3 de Janeiro de 2005

Data Término

3 de Janeiro de 2008

Resumo

Área de Investigação e Desenvolvimento (I&D)

Fundamentos da Computação

Resumo

Os métodos que permitem raciocinar de uma forma precisa sobre as implementações são de grande importância, em engenharia de software. A interligação entre os dois mundos (especificação/implementação) nem sempre é evidente, levando a um afastamento, por parte dos programadores, do cálculo de programas. Este projecto procura fazer essa ligação utilizando uma abordagem de grande tradição em matemática: as conexões de Galois. Estas têm suscitado bastante interesse, quer nas ciências da computação, quer em diversas outras áreas da engenharia. Procura-se também a sua utilização em termos mais práticos na área da engenharia reversa e da renovação de software.

Objectivos Estratégicos

Neste projecto pretende-se explorar sistematicamente a algebrização da lógica da programação (pré/pós condições, invariantes, *business logic*, etc.) com base na utilização explícita de conexões de Galois. Para este efeito, deverá ser construída uma metodologia e um ambiente de trabalho/ferramenta (Calculator) que tome o par de adjuntos de Galois como a sua principal primitiva de programação, desenvolvendo sobre ele uma álgebra de combinadores que permita construir novas conexões a partir daquelas já conhecidas, raciocinar sobre elas, etc.

Uma área de aplicação e teste deste projecto será o da engenharia reversa, enquadrável no projecto PRe Project (Program Understanding and Re-engineering: Calculi and Applications) financiado pela FCT (POSI/ICHS/ 44304/2002). Pretende-se aplicar a metodologia a desenvolver à reconstituição da lógica de negócio (*business logic*) de aplicações legadas.

Contribuições

Principais contribuições técnico-científicas

Prevê-se que deste projecto resulte uma ferramenta (*Calculator*) que permita a utilização de conexões de Galois como primitiva de programação. A importância desta ferramenta relaciona-se com a sistematização do uso de conexões de Galois, permitindo a construção de estruturas complexas a partir de um repositório base. Espera-se que este ambiente beneficie o estudo da aplicação de conexões de Galois, sobretudo no âmbito do cálculo relacional de programas.

Adicionalmente, uma ferramenta para o cálculo de um modelo relacional a partir de uma especificação de dados algébrica em VDM-SL (*VooDooM*¹) foi desenvolvida e, presentemente, está a ser aperfeiçoada. Esta ferramenta demonstra a efectividade dos métodos de refinamento de dados do cálculo relacional, e a sua aplicação prática a problemas reais. A sua importância relaciona-se com a possibilidade de especificar as estruturas de dados num modelo conceptual e, automaticamente, obter-se a sua representação no modelo relacional, assim como um *script* SQL que implementa a base de dados correspondente. Actualmente, está a ser desenvolvida uma evolução desta ferramenta, utilizando as possibilidades do *type-level programming* da linguagem *Haskell*, permitindo a verificação da correcção das transformações aplicadas em tempo de compilação.

Publicações

- T.L. Alves, P.F. Silva, J. Visser, and J.N. Oliveira. Strategic term rewriting and its application to a VdmSL to SQL conversion. In John Fitzgerald, Ian J. Hayes, and Andrzej Tarlecki, editors, *FM 2005: Formal Methods: International Symposium of Formal Methods Europe, Newcastle, UK, July 18-22*, volume 3582 / 2005 of *Lecture Notes in Computer Science*. Springer-Verlag, 2005.

Abstract. We constructed a tool, called VooDooM, which converts datatypes in VDM-SL into SQL relational data models. The conversion involves

¹<http://voodooom.sourceforge.net>

transformation of algebraic types to maps and products, and pointer introduction. The conversion is specified as a theory of refinement by calculation. The implementation technology is strategic term rewriting in Haskell, as supported by the Strafunski bundle. Due to these choices of theory and technology, the road from theory to practise is straightforward.

Keywords: Strategic term rewriting, program calculation, VDM, SQL.

Enquadramento

Enquadramento Científico

A comunidade informática foi abalada quando, há cerca de quatro décadas atrás, devido há falta de técnicas e metodologias, foi declarada uma “crise de software”. Assim, foram propostas diversas abordagens para tentar ultrapassar esta situação. Uma das correntes emergentes defendeu um avanço para a crescente teorização das ciências de programação, tentando reutilizar as teorias matemáticas já existentes.

Um marco na utilização de técnicas algébricas para análise e cálculo de programas foi a publicação de *Algebra of Programming*. Este livro baseia-se na utilização da álgebra das relações binárias dita *pointfree*, derivadas da teoria de categorias.

Contudo, a complexidade lógica de um programa ou aplicação informática pode ser arbitrariamente complexa, exigindo recursos ao cálculo das relações binárias que ele ainda não pode fornecer. Em particular, o cálculo não oferece ainda dispositivos que agilizem a sua aplicação em grande escala.

Uma das descobertas dos últimos anos tem sido a estruturação do cálculo relacional com base na teoria que emerge quando os seus principais operadores são vistos como conexões de Galois. O cálculo ganha não só em estrutura como em escalabilidade, e os espectro de aplicações desta visão não pára de crescer - vejam-se exemplos como a interpretação abstracta, as leis de refinamento, a construção de estruturas de dados tão populares como as tabelas de hashing, e até, exemplos na área da engenharia civil.

Motivação

Nos últimos anos têm-se verificado avanços significativos na algebrização da programação e no desenvolvimento de programas por cálculo, cada vez mais seguindo a (boa e consagrada) tradição em engenharia de, perante um problema, formular um modelo, raciocinar sobre ele e encontrar uma solução por resolução de um ou mais sistemas de (in)equações. A ênfase em estruturas matemáticas de “sabor” polinomial e o sistemático recurso a transformadas que exprimem as estruturas de

dados e os próprios programas como soluções de equações polinomiais faz lembrar técnicas da engenharia clássica como, por exemplo, a transformada de Laplace.

No cálculo de programas verifica-se que é mais simples o raciocínio sobre as relações quando estas se apresentam na sua forma *pointfree*, isto é, quando as variáveis não aparecem explicitamente representadas. No entanto, as implementações dos programas utilizam variáveis de uma forma explícita (*pointwise*). O uso de conexões de Galois permite fazer uma “ponte” entre os dois domínios, ou seja, actua como uma transformada entre o formalismo de raciocínio em *point-free* e a sua implementação em *pointwise*. Neste projecto pretende-se explorar esta potencialidade das conexões de Galois.

A engenharia reversa é uma área onde se prevê a aplicação do trabalho a desenvolver. Apesar das vantagens apresentadas pelas metodologias formais para a produção de software, estas continuam a ser quase completamente ignoradas pela comunidade de programadores. Contudo, na área da engenharia reversa de software a situação parece ser mais promissora, uma vez que a sua grande complexidade obriga à utilização de técnicas avançadas e com uma boa base teórica.

Um dos ramos da engenharia reversa é a especificação reversa, em que se tenta inferir a especificação a partir do código fonte de um software existente. Esta disciplina é de extrema importância para empresas e instituições com grandes patrimónios de software legado, que pretendem conhecer melhor a sua funcionalidade. Este conhecimento é vital, não só para efeitos de manutenção, mas também para a renovação de partes do sistema.

A análise das estruturas de dados é um problema que suscitou bastante interesse e estudo. No entanto, a análise da lógica de negócio (*business logic*) de aplicações legadas continua a ser um campo por explorar. Neste projecto pretende-se utilizar o poder das conexões de Galois para a inferência da especificação de código legado, incluindo todo o conhecimento (lógico) nele implícito.

Objectivos Detalhados

Para atingir os objectivos globais propostos neste trabalho definiram-se vários objectivos parciais a atingir. Com o decorrer dos trabalhos outros poderão vir a ser adicionados à listagem. Em seguida, apresenta-se um resumo desses objectivos:

- Aquisição de bases científicas na área de investigação que permitam sustentar o desenvolvimento subsequente.
- Estudo e descrição das propriedades universais das construções da teoria de categorias utilizadas na álgebra de programação como conexões de Galois. Isto deverá permitir a obtenção das leis derivadas de uma construção de uma forma imediata.

- Estudo da aplicação de conexões de Galois ao refinamento de estruturas de dados. Este trabalho deverá ligar-se ao desenvolvimento do *VooDooM* que permite o cálculo de uma forma automática.
- Construção de um repositório de conexões de Galois conhecidas que sejam importantes no contexto das Ciências da Computação. Deverão ser procuradas propriedades genéricas que permitam a classificação dessas conexões segundo vários aspectos.
- Desenvolvimento de uma álgebra de combinadores de conexões de Galois. Estes combinadores deverão permitir a estruturação e a construção de conexões arbitrariamente complexas a partir de conexões já existentes.
- Estudo do teorema de abstracção de Reynold para inferir conexões de Galois complexa a partir dos tipos de funções polimórficas.
- Reconhecimento de estruturas das linguagens de programação como conexões de Galois e a utilização desse mapeamento para o cálculo de implementações a partir das especificações formais. Por outro lado, utilizar esse conhecimento para a engenharia reversa de *software*.
- Desenvolvimento do *Calculator*, ferramenta que utiliza conexões de Galois como primitiva de programação e que permite a sua utilização sistemática e o estudo da sua aplicação a problemas concretos.

Trabalhos Alternativos

O trabalho que serve de ponto de partida para esta tese consta no artigo *Safety of abstract interpretation for free, via logical relations and Galois connections*. Este é, provavelmente, o desenvolvimento mais avançado no âmbito da aplicação de conexões de Galois ao cálculo de programas. Este trabalho pretende partir dessa base, continuar o seu desenvolvimento e criar uma implementação para a sua utilização prática.

Não se conhecem actualmente trabalhos alternativos ou concorrenciais que tenham os mesmos objectivos que o presente trabalho. A maioria da investigação relacionada com conexões de Galois, em Ciências da Computação, incide sobre o seu uso em interpretação abstracta.

Bibliografia Principal

1. Kevin Backhouse and Roland Carl Backhouse. Safety of abstract interpretations for free, via logical relations and galois connections. *Sci. Comput. Program.*, 51(1-2):153–196, 2004.

2. Roland Backhouse. Mathematics of program construction. Draft, June 2004.
3. Roland Backhouse, Roy Crole, and Jeremy Gibbons, editors. *Algebraic and Coalgebraic Methods in the Mathematics of Program Construction*. Springer-Verlag, April 2000.
4. J.N. Oliveira and C.J. Rodrigues. Transposing relations: from *Maybe* functions to hash tables. In *MPC'04 : Seventh International Conference on Mathematics of Program Construction, 12-14 July, 2004, Stirling, Scotland, UK (Organized in conjunction with AMAST'04)*, volume 3125 of *Lecture Notes in Computer Science*, pages 334–356. Springer, 2004.

Desenvolvimento

Macro-planeamento das Actividades

O plano original teve uma evolução em relação à sua configuração inicial, devido a várias circunstâncias e ao surgimento de novas ideias. As diferenças referem-se à inclusão de actividades não planeadas originalmente no primeiro ano da tese, com um conseqüente alargamento temporal de algumas actividades previstas para esse ano.

Plano original

Ano	Actividades
1º Ano	Estudo da base teórica Estudo do estado de arte Definição dos objectivos estratégicos
2º Ano	Desenvolvimento de actividades de investigação originais Escrita e submissão de artigos
3º Ano	Escrita da tese de doutoramento Escrita e submissão de artigos Desenvolvimento de software (Calculator)

Plano actual

Ano	Actividades
1º Ano (concluído)	Estudo da base teórica Estudo do estado de arte Definição dos objectivos estratégicos Desenvolvimento de software (VooDooM) Escrita e submissão de artigos
2º Ano (previsto)	Desenvolvimento de actividades de investigação originais Escrita e submissão de artigos Estudo do estado de arte Desenvolvimento de software (VooDooM e protótipo do Galculator)
3º Ano (previsto)	Escrita da tese de doutoramento Escrita e submissão de artigos Desenvolvimento de software (Galculator)

Recursos Necessários

Para a execução deste trabalho não há a necessidade de recursos para além dos habituais em trabalhos em Ciências de Computação (computador pessoal, sistema de edição de texto, compiladores, etc.). Em termos bibliográficos, apresenta-se em seguida uma lista adicional de referências já consultadas ou cuja consulta futura será necessária. Eventualmente, no decorrer dos trabalhos, surgirão outras referências de consulta necessária.

Bibliografia

1. Chritiene Aarts, Roland Backhouse, Paul Hoogendijk, Ed Voermans, and Jaap van der Woude. A relational theory of datatypes. Available from www.cs.nott.ac.uk/~rcb/papers, December 1992.
2. Ralph-Johan Back and Joakim von Wright. *Refinement Calculus — A Systematic Introduction*. Graduate Texts in Computer Science. Springer, 1998.
3. Richard Bird and Oege de Moor. *Algebra of Programming*. Series in Computer Science, C.A.R. Hoare Series Editor. Prentice Hall, 1997.
4. Radim Bölohlavek. Fuzzy galois connections and fuzzy concept lattices: from binary relations to conceptual structures. pages 462–494, 2000.
5. B. A. Davey and H. A. Priestley. *Introduction to Lattices and Order*. Mathematical Textbooks. Cambridge University Press, 1990.

6. Augusto Franco de Oliveira. *Lógica e Aritmética*. Trajectos Ciência. Gradiva, 1991.
7. Bernhard Ganter and Rudolf Wille. *Formal Concept Analysis—Mathematical Foundations*. Springer-Verlag, 1999.
8. Lars Gårding and Torbjörn Tambour. *Algebra for Computer Science*. Universitext. Springer-Verlag, 1988.
9. F. William Lawvere and Stephen H. Schanuel. *Conceptual Mathematics—A first introduction to categories*. Cambridge University Press, 1997.
10. Austin Melton, Bernd S. W. Schröder, and George E. Strecker. Lagois connections - a counterpart to galois connections. *Theor. Comput. Sci.*, 136(1):79–107, 1994.
 J.N. Oliveira. "Bagatelle in C arranged for VDM SoLo". *Journal of Universal Computer Science*, 7(8):754–781, 2001. Special Issue on *Formal Aspects of Software Engineering* (Colloquium in Honor of Peter Lucas , Institute for Software Technology, Graz University of Technology, May 18-19, 2001).
11. J.N. Oliveira. First steps in pointfree functional dependency theory, 2005.
12. David Pichardie. *Interprétation abstraite en logique intuitionniste: extraction d'analyseurs Java certifiés*. PhD thesis, Université de Rennes, December 2005.
13. D.E. Rydeheard and R.M. Burstall. *Computational Category Theory*. Series in Computer Science, C.A.R Hoare Series Editor. Prentice Hall, 1988.
14. Ian Stewart. *Galois Theory*. Chapman & Hall, second edition, 1989.
15. Jean-Pierre Tignol. *Galois' Theory of Algebraic Equations*. World Scientific, 2001.
16. R.F.C. Walters. *Categories and Computer Science*. Cambridge Computer Science Texts. Cambridge University Press, 1991.

Recursos Disponibilizados

Dada a pouca exigência de recursos requerida por este trabalho, a única dificuldade relaciona-se com a obtenção de algumas referências bibliográficas que, dado a sua especificidade, não se encontram disponíveis na biblioteca nem nos repositórios digitais.

Avaliação

Análise Comparativa

Este trabalho pretende ser inovador em relação a outros trabalhos existentes na mesma área. A utilização de conexões de Galois em ferramentas está normalmente restrita à interpretação abstracta, e apenas como um meio de garantia da correcção das representações e abstracções utilizadas. Frequentemente, chegar a uma conexão de Galois que permita a interpretação abstracta de determinado problema é um trabalho que apela à criatividade dos intervenientes, sendo o cálculo utilizado apenas para a prova que a solução encontrada é adequada e correcta. *Theorem Provers* começam a ser utilizados para automatizar este processo.

Neste trabalho pretende-se que as conexões de Galois possam ser utilizadas com uma primitiva de programação numa ferramenta criada para o efeito. Actualmente, não existe nenhuma ferramenta do género disponível.

Auto-avaliação do Trabalho Realizado

O plano original reservava o primeiro ano da tese para o estudo das referências bibliográficas e aquisição do conhecimento base necessário para o desenvolvimento do resto do trabalho. Neste momento, esse objectivo está quase cumprido, faltando porém, ainda o estudo mais detalhado de algumas matérias específicas.

Actualmente, está a ser iniciado o trabalho planeado para o segundo ano da tese, nomeadamente o desenvolvimento de combinadores de conexões de Galois, e a construção de um protótipo funcional que implemente as ideias da área teórica.

Adicionalmente, no seguimento de um projecto anterior, foi desenvolvida uma ferramenta para o cálculo de um modelo relacional a partir de uma especificação de dados algébrica em VDM-SL, o *VooDooM*. Apesar deste trabalho não estar directamente relacionado com o tema da tese, espera-se poder fazer uma unificação de conceitos no futuro.

Auto-avaliação da Documentação Produzida

A documentação produzida até ao momento resume-se a uma publicação, já referenciada, sobre a ferramenta *VooDooM*. Este documento foi aceite e apresentado numa conferência internacional.

Actualmente, está a ser redigido um documento que sistematiza a teoria base necessária à compreensão do conceito de conexão de Galois, reflectindo o estudo realizado. Inclui ainda um inventário de conexões de Galois conhecidas úteis no nosso contexto, e que poderá servir de base para uma metodologia para a construção de novas conexões.

Dificuldades

Dificuldades Técnico-Científicas

As principais dificuldades sentidas durante o primeiro ano deste trabalho estão relacionadas com a falta de bases científicas em algumas áreas matemáticas necessárias para a compreensão dos conceitos envolvidos. Devido ao cariz muito teórico desta tese, isso obrigou a um trabalho suplementar de estudo, com um conseqüente progresso mais lento nos resultados produzidos.

Outras Dificuldades

A quantidade de procedimentos administrativos e burocráticos inerentes à realização da tese consomem demasiado tempo que seria mais útil se utilizado em actividades concretas de investigação. Mais grave ainda, muitos desse procedimentos são “obscuros” em termos de informação, obrigando a um esforço suplementar numa tentativa da sua compreensão e, conseqüentemente, a um dispêndio extra de tempo e paciência.