

Sérgio Queiroz de Medeiros

**Correspondência entre PEGs e Classes de
Gramáticas Livres de Contexto**

Tese de Doutorado

Tese apresentada ao Programa de Pós-graduação em Informática do Departamento de Informática da PUC-Rio como requisito parcial para obtenção do título de Doutor em Informática

Orientador: Prof. Roberto Ierusalimsky

Rio de Janeiro
Agosto de 2010



Sérgio Queiroz de Medeiros

Correspondência entre PEGs e Classes de Gramáticas Livres de Contexto

Tese apresentada ao Programa de Pós-graduação em Informática do Departamento de Informática do Centro Técnico Científico da PUC-Rio como requisito parcial para obtenção do título de Doutor em Informática. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Roberto Ierusalimschy

Orientador

Departamento de Informática — PUC-Rio

Prof. Edward Hermann Haeusler

Departamento de Informática — PUC-Rio

Prof. Noemi de La Rocque Rodriguez

Departamento de Informática — PUC-Rio

Prof. Alex de Vasconcellos Garcia

Instituto Militar de Engenharia — IME

Prof. Anamaria Martins Moreira

Universidade Federal do Rio Grande do Norte — UFRN

Prof. José Eugenio Leal

Coordenador Setorial do Centro Técnico Científico — PUC-Rio

Rio de Janeiro, 18 de Agosto de 2010

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Sérgio Queiroz de Medeiros

Graduou-se em Ciência da Computação na Universidade Federal do Rio Grande Norte (UFRN), onde também fez mestrado em Sistemas e Computação. Durante o doutorado, na PUC-Rio, trabalhou na área de Linguagens de Programação.

Ficha Catalográfica

Medeiros, Sérgio Queiroz de

Correspondência entre PEGs e Classes de Gramáticas Livres de Contexto / Sérgio Queiroz de Medeiros; orientador: Roberto Ierusalimschy. — 2010.

86 f. ; 30 cm

Tese (doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática, 2010.

Inclui bibliografia.

1. Informática – Teses. 2. Gramáticas de Expressões de Parsing. 3. Expressões Regulares. 4. Gramáticas Livres de Contexto. 5. Semântica Natural. 6. $LL(1)$. 7. $LL(k)$ -Forte. I. Ierusalimschy, Roberto. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

Agradecimentos

Primeiro, gostaria de agradecer a Roberto pelos anos de orientação e por me motivar a tentar fazer um trabalho melhor.

Agradeço também a todo mundo do LabLua. Em especial, gostaria de agradecer a Fabio que esteve sempre por aqui e ajudou em vários momentos da pesquisa.

Gostaria de agradecer aos meu pais e às minhas irmãs, que mesmo longe estão comigo.

Agradeço a todos os companheiros de república, em especial a Bruno.

Gostaria de agradecer a todos os amigos, aos antigos e aos que fiz nesses anos de Rio.

Agradeço a Marcia pelo apoio, pelo carinho e pelo sorriso.

Finalmente, gostaria de agradecer ao CNPq, ao Tecgraf e à PUC-Rio pela ajuda financeira e pelo bom ambiente de trabalho.

Resumo

Medeiros, Sérgio Queiroz de; Ierusalimschy, Roberto. **Correspondência entre PEGs e Classes de Gramáticas Livres de Contexto**. Rio de Janeiro, 2010. 86p. Tese de Doutorado — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Gramáticas de Expressões de Parsing (PEGs) são um formalismo que permite descrever linguagens e que possui como característica distintiva o uso de um operador de escolha ordenada. A classe de linguagens descrita por PEGs contém propriamente todas as linguagens livres de contexto determinísticas. Nesta tese discutimos a correspondência de PEGs com dois outros formalismos usados para descrever linguagens: expressões regulares e Gramáticas Livres de Contexto (CFGs). Apresentamos uma formalização de expressões regulares usando semântica natural e mostramos uma transformação para converter expressões regulares em PEGs que descrevem a mesma linguagem; essa transformação pode ser facilmente adaptada para acomodar diversas extensões usadas por bibliotecas de expressões regulares (e.g., repetição preguiçosa e subpadrões independentes). Também apresentamos uma nova formalização de CFGs usando semântica natural e mostramos a correspondência entre CFGs lineares à direita e PEGs equivalentes. Além disso, mostramos que gramáticas $LL(1)$ com uma pequena restrição descrevem a mesma linguagem quando interpretadas como CFGs e quando interpretadas como PEGs. Por fim, mostramos como transformar CFGs $LL(k)$ -forte em PEGs equivalentes.

Palavras-chave

Gramáticas de Expressões de Parsing; Expressões Regulares; Gramáticas Livres de Contexto; Semântica Natural; $LL(1)$; $LL(k)$ -Forte.

Abstract

Medeiros, Sérgio Queiroz de; Ierusalimschy, Roberto. **Correspondence between PEGs and Classes of Context-Free Grammars**. Rio de Janeiro, 2010. 86p. DSc Thesis — Department of Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Parsing Expression Grammars (PEGs) are a formalism that allow us to describe languages and that has as its distinguishing feature the use of an ordered choice operator. The class of languages described by PEGs properly contains all deterministic context-free languages. In this thesis we discuss the correspondence between PEGs and two other formalisms used to describe languages: regular expressions and Context-Free Grammars (CFGs). We present a new formalization of regular expressions that uses natural semantics and we show a transformation to convert a regular expression into a PEG that describes the same language; this transformation can be easily adapted to accommodate several extensions used by regular expression libraries (e.g., lazy repetition and independent subpatterns). We also present a new formalization of CFGs that uses natural semantics and we show the correspondence between right linear CFGs and equivalent PEGs. Moreover, we show that $LL(1)$ grammars with a minor restriction define the same language when interpreted as a CFG and when interpreted as a PEG. Finally, we show how to transform strong- $LL(k)$ CFGs into PEGs that are equivalent.

Keywords

Parsing Expression Grammars; Regular Expressions; Context-Free Grammars; Natural Semantics; $LL(1)$; Strong- $LL(k)$.

Sumário

| | | |
|-----|---|----|
| 1 | Introdução | 11 |
| 1.1 | Visão Geral de PEGs | 13 |
| 1.2 | Organização da Tese | 19 |
| 2 | Gramáticas de Expressões de Parsing | 20 |
| 2.1 | Definição de PEGs | 20 |
| 2.2 | Interpretação de PEGs Usando a Relação \Rightarrow_G | 21 |
| 2.3 | Formalização de PEGs Usando Semântica Natural | 23 |
| 3 | Expressões Regulares e PEGs | 32 |
| 3.1 | Expressões Regulares | 32 |
| 3.2 | Definição de Expressões Regulares Usando Semântica Natural | 33 |
| 3.3 | Equivalência Entre Expressões Regulares e PEGs | 37 |
| 3.4 | Transformação de uma Expressão Regular em uma PEG Equivalente | 39 |
| 3.5 | Transformação de Repetições e_1^* onde e_1 Casa a Cadeia Vazia | 42 |
| 3.6 | Corretude da Transformação II | 46 |
| 4 | Gramáticas Livres de Contexto e PEGs | 52 |
| 4.1 | Definição Usual de CFGs | 52 |
| 4.2 | Uma Nova Formalização de Linguagens Livres de Contexto usando Semântica Natural | 53 |
| 4.3 | Correspondência entre $\overset{\text{CFG}}{\rightsquigarrow}$ e $\overset{\text{PEG}}{\rightsquigarrow}$ | 59 |
| 4.4 | Correspondência entre CFGs Lineares à Direita e PEGs | 61 |
| 4.5 | Correspondência entre CFGs $LL(1)$ e PEGs | 63 |
| 4.6 | Correspondência entre CFGs $LL(k)$ -Forte e PEGs | 72 |
| 5 | Conclusão | 80 |
| 5.1 | Trabalhos Relacionados | 80 |
| 5.2 | Contribuições | 82 |

Lista de figuras

| | | |
|-----|--|----|
| 1.1 | PEG que Descreve a Sintaxe de PEGs | 14 |
| 2.1 | Definição da Relação $\overset{PEG}{\rightsquigarrow}$ Usando Semântica Natural | 24 |
| 2.2 | Exemplo de Árvore de Prova Usando a Relação $\overset{PEG}{\rightsquigarrow}$ | 25 |
| 2.3 | Exemplo de Casamento Quando uma PEG não é Completa | 30 |
| 3.1 | Definição da Relação $\overset{FE}{\rightsquigarrow}$ Usando Semântica Natural | 34 |
| 3.2 | Definição da Função Π , onde $G_k = (V_k, T, P_k, p_k)$ | 40 |
| 3.3 | Definição da Função $isNull$ | 43 |
| 3.4 | Definição da Função $hasEmpty$ | 44 |
| 3.5 | Definição da Função f_{out} | 44 |
| 3.6 | Definição da Função $f_{in}(e_0)$, onde $\neg isNull(e_0)$ e $hasEmpty(e_0)$ | 45 |
| 4.1 | Definição da Relação $\overset{CFG}{\rightsquigarrow}$ Usando Semântica Natural | 56 |
| 4.2 | Forma Geral da Árvore de Prova Quando a Gramática G Possui Estrutura BNF | 67 |
| 4.3 | Algoritmo para Computar $FOLLOW^G$ dada uma Gramática $LL(1)$ G com Estrutura BNF | 68 |
| 4.4 | Definição da Relação $\overset{LL(1)}{\rightsquigarrow}$ Usando Semântica Natural | 69 |
| 4.5 | Algoritmo para Computar $FOLLOW_k^G$ dada uma Gramática $LL(k)$ -Forte G com Estrutura BNF | 73 |
| 4.6 | Definição da Relação $\overset{LL(k)}{\rightsquigarrow}$ Usando Semântica Natural | 76 |

Lista de tabelas

| | | |
|-----|--|----|
| 3.1 | Expressões Regulares e suas Linguagens Correspondentes | 33 |
| 4.1 | Equivalência Entre Classes de CFGs e PEGs | 79 |

Soneto do dismantelo azul

*Então, pinte de azul os meus sapatos
por não poder de azul pintar as ruas,
depois, vesti meus gestos insensatos
e colori as minhas mãos e as tuas.*

*Para extinguir em nós o azul ausente
e aprisionar no azul as coisas gratas,
enfim, nós derramamos simplesmente
azul sobre os vestidos e as gravatas.*

*E afogados em nós, nem nos lembramos
que no excesso que havia em nosso espaço
pudesse haver de azul também cansaço.*

*E perdidos de azul nos contemplamos
e vimos que entre nós nascia um sul
vertiginosamente azul. Azul.*

Carlos Pena Filho, Livro Geral.