

UNIVERSIDADE FEDERAL DE GOIÁS – UFG
CAMPUS CATALÃO – CaC
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO – DCC

Bacharelado em Ciência da Computação

Projeto Final de Curso

**Estudo de um compilador sob a perspectiva de um
sistema online de programação**

Autor: Nélcio Carneiro Júnior

Orientador: Dr. Thiago Jabur

Nélio Carneiro Júnior

**Estudo de um compilador sob a perspectiva de um sistema online de
programação**

Monografia apresentada ao Curso de
Bacharelado em Ciência da Computação da
Universidade Federal de Goiás Campus Catalão
como requisito parcial para obtenção do título de
Bacharel em Ciência da Computação

Área de Concentração: Compiladores

Orientador: Dr. Thiago Jabur

Carneiro, Nélío

Estudo de um compilador sob a perspectiva de um sistema online de programação/Dr. Thiago Jabur- Catalão - 2011

Número de paginas: 13

Projeto Final de Curso (Bacharelado) Universidade Federal de Goiás, Campus Catalão, Curso de Bacharelado em Ciência da Computação, 2011.

Palavras-Chave: 1. Compilador. 2. Sistemas Web. 3. Linguagem de Programação

Nélio Carneiro Júnior

**Estudo de um compilador sob a perspectiva de um sistema online de
programação**

Monografia apresentada e aprovada em _____ de _____
Pela Banca Examinadora constituída pelos professores.

Dr. Thiago Jabur – Presidente da Banca

Professor 1

Professor 2

Dedico este trabalho

AGRADECIMENTOS

Por fim, agradeço a todos que de forma direta ou indireta que cotribuíram para a minha formação.

“Os desafios.

Sulamita

Sumário

Introdução	1
1 Compiladores	3
1.1 Introdução	3
1.2 Modelo de Compilação de Análise e Síntese	4
1.3 Análise	5
2 Portugol	7
2.1 Introdução	7
2.2 Estrutura da Linguagem	7
2.2.1 Comentários	7
2.2.2 Tipos básicos	7
2.2.3 Variáveis e Identificadores	9
2.2.4 Declaração de variáveis	9
2.2.5 Atribuição	9
2.2.6 Operadores	10
2.2.7 Estrutura Sequencial	11
2.2.8 Estruturas Condicionais	11
2.2.9 Estruturas de Repetição	12
2.2.10 Procedimentos e Funções	13

Lista de Figuras

Introdução

É evidente o crescente número de usuários conectados e dependentes da internet ao longo do tempo. Percebe-se também que estes usuários, estão cada vez mais necessitados de que suas ferramentas, antes usadas em seus desktops, fiquem disponíveis online, de prontidão, sempre que se fizer necessário, mesmo estando conectados longe de casa. Fez-se então necessário que o desenvolvimento de softwares caminhasse neste mesmo sentido. [carece de fontes]

Um aspecto interessante deste processo a ser analisado são as aplicações que dependem de um compilador, ou seja, programas que quando instalados na máquina sempre fazem uso de um compilador ou então, são eles mesmos, compiladores. Um exemplo deste tipo de software seria um compilador para Portugol. Como iria se comportar tal compilador ao ser usado na internet? Como se daria a implementação deste compilador? Quais aspectos a serem tratados? E como seria o sistema na web que necessitasse deste compilador?

O objetivo principal deste trabalho é analisar o comportamento de um compilador feito para a linguagem Portugol[fontes], trabalhando sob um sistema de programação na Web. A proposta é fazer um estudo detalhado da implementação deste compilador, tendo assim uma base sólida para o desenvolvimento do mesmo, visando na prática tirar conclusões à respeito das técnicas utilizadas.

Este trabalho também tem como objetivo desenvolver uma aplicação Web que irá utilizar o compilador. A aplicação será um editor o qual o usuário fará seus programas em Portugol, tendo de imediato a resposta do compilador. A intenção é analisar como irá se comportar o compilador, bem como, verificar a viabilidade de uso de tal aplicação, ao se implementar utilizando as técnicas descritas ao longo deste trabalho. A aplicação terá um caráter educativo, no qual o usuário fará programas e, ele e seus trabalhos, poderão ser acompanhados por um instrutor, visando assim verificar os passos realizados pelo usuário, bem como a maneira pela qual o compilador tem sido usado.

No capítulo 1 são abordados os conceitos de compiladores, as técnicas utilizadas na criação do compilador e como se dará o seu funcionamento diante uma aplicação Web.

No capítulo 2 é feita uma apresentação da linguagem do compilador: o Portugol. Suas características e o embasamento para a criação do compilador baseado nesta linguagem também fará parte deste capítulo

No capítulo 3 pode-se observar um levantamento sobre o estado da arte dos compiladores e como tem sido o uso destes em aplicações na Web. Neste também irei citar casos em que editores/compiladores foram colocados na Web.

No capítulo 4 é discorrido sobre a metodologia proposta para o desenvolvimento do protótipo de editor/compilador Web.

No capítulo 5 é abordado todo o processo de testes, além de ser realizado uma análise dos resultados. Por fim há uma conclusão do trabalho realizado, levantando os objetivos alcançados, pontos positivos, pontos negativos, dificuldades e trabalhos futuros.

Ainda é possível encontrar o código fonte do compilador e da aplicação Web criada, além de toda a documentação de testes elaborada no apêndice desta monografia.

Capítulo 1

Compiladores

1.1 Introdução

Criado por volta dos anos 50, o nome Compilador se refere ao processo de composição de um programa através da reunião de várias rotinas de bibliotecas. O processo de tradução (de uma linguagem fonte para uma linguagem objeto), considerado hoje a função central de um compilador, era então conhecido como programação automática[Rangel, 1999]

Definido em [AHO, 1995], um compilador é um programa que lê outro programa escrito em uma linguagem — a linguagem de origem — e o traduz em um programa equivalente em outra linguagem — a linguagem de destino. Como uma importante parte no processo de tradução, o compilador reporta ao seu usuário a presença de erros no programa origem.

Ao longo dos anos 50, os compiladores foram considerados programas notoriamente difíceis de escrever. O primeiro compilador Fortran, por exemplo, consumiu 18-homens ano para implementar[Backus, 1957]. Desde então, foram descobertas técnicas sistemáticas para o tratamento de muitas das mais importantes tarefas desenvolvidas por um compilador.

A variedade de compiladores nos dias de hoje é muito grande. Existem inúmeras linguagens fontes, as quais poderiam ser citadas em várias páginas deste trabalho. Isso se deve principalmente ao fato de que com o aumento do uso dos computadores, aumentou também, as necessidades de cada indivíduo, sendo estas específicas, exigindo por sua vez linguagens de programação diferentes. Este processo — juntamente com a evolução da tecnologia de desenvolvimento de compiladores — levou à criação de várias técnicas diferentes para a construção de um compilador, ou seja, passou a existir diferentes maneiras de se implementar um compilador. No entanto, a despeito dessa aparente complexidade, as tarefas básicas que qualquer compilador precisa realizar são essencialmente as mesmas.

A grande maioria dos compiladores de hoje fazem uso da técnica chamada: *tradução*

dirigida pela sintaxe. Nesta técnica as regras de contrução do programa fonte são utilizadas para guiar todo o processo de compilação. Algumas das técnicas mais antigas utilizadas na contrução dos primeiros compiladores (da linguagem Fortran) pode ser obtido em [Rosen, 1967].

1.2 Modelo de Compilação de Análise e Síntese

Ainda segundo [Rangel, 1999], existem duas tarefas triviais a serem executadas por um compilador nesse processo de tradução:

- *análise*, em que o texto de entrada (na linguagem fonte) é examinado, verificado e compreendido
- *síntese*, ou *geração de código*, em que o texto de saída (na linguagem objeto) é gerado, de forma a corresponder ao texto de entrada.

Em [Aho, 1995], *análise* é colocada como uma tarefa que divide o programa fonte nas partes constituintes e cria uma representação intermediária do mesmo. E *síntese* constrói o programa alvo desejado, a partir da representação intermediária.

Geralmente, pensamos nessas tarefas como fases que ocorram durante o processo de compilação. No entanto, não se faz totalmente necessário que a análise de todo o programa seja realizada antes que o primeiro trecho de código objeto seja gerado. Ou seja, estas duas fases podem ser intercaladas. Por exemplo, o compilador pode analisar cada comando do programa de entrada e então gerar de imediato o código de saída correspondente ao respectivo comando. Ou ainda, o compilador pode esperar pelo fim da análise de cada bloco de comando — ou unidade de rotina (rotina, procedimentos, funções) — para então gerar o código correspondente ao bloco. Para aproveitar melhor a memória durante a execução, compiladores costumavam ser divididos em várias etapas, executados em sequência. Cada etapa constitui uma parte do processo de tradução, transformando assim o código fonte em alguma estrutura intermediária adequada, cada vez mais próxima do código objeto final.

É natural que a análise retorne como resultado uma representação do programa fonte que contenha informação necessária para a geração do programa objeto que o corresponda. Quase sempre, essa representação (conhecida como *representação intermediária* [Rangel, 1999]) tem como complemento tabelas que contêm informações adicionais sobre o programa fonte. Pode ter casos em que a representação intermediária toma a forma de um programa em uma *linguagem intermediária*, deixando assim mais fácil a tradução para a linguagem objeto desejada.

Não importando a maneira pela qual se toma a representação intermediária, ela tem de conter necessariamente toda a informação para a geração do código objeto. Uma das características da representação intermediária é que as estruturas de dados implementadas devem dar garantia de acesso eficiente as informações.

Imagem - Rangel[1999]

Segundo [Rangel, 1999], uma das formas mais comuns de tabela utilizada nessa representação intermediária é a *tabela de símbolos*, em que se guarda para cada identificador(*símbolo*) usado no programa as informações correspondentes.

Há também um modelo possível em [Ullman, 1977], o qual se faz a separação total entre o *front-end*, encarregado da fase de análise, e o *back-end*, encarregado pela geração de código. Com isso tem-se que:

- front-end e back-end se comunicam apenas da representação intermediária;
- o front-end depende exclusivamente da linguagem fonte
- o back-end depende exclusivamente da linguagem objeto.

Essa idéia tem como objetivo simplificar a implementação de diferentes linguagens de programação para diferentes máquinas. Basta-se então escrever um front-end para cada linguagem e um back-end para cada máquina. Ou seja, se deseja implementar x linguagens para y máquinas, precisa-se fazer x front-ends e y back-ends. Este esquema se torna mais fácil de aplicar quando há semelhança entre as máquinas e o mesmo acontece com as linguagens.

1.3 Análise

É normal associar a *sintaxe* a idéia de forma, em oposição a *semântica* que se associa a significado, conteúdo. Tem-se então que a sintaxe de uma linguagem de programação deve descrever todos os aspectos relativos à forma de construção de programas corretos na linguagem, enquanto a semântica deve descrever o que acontece quando o programa é executado. Portanto, toda análise está relacionada com sintaxe, e a semântica deveria corresponder apenas à geração de código, que deve preservar o significado do programa fonte, contruindo um programa objeto com o mesmo significado[Rangel, 1999]

É necessário ressaltar uma diferença existente entre a teoria e a prática. Quando falamos em teoria, somente os programas corretos pertencem à linguagem, não havendo interesse nos programas incorretos. O fato é que um programa ou é da linguagem (está correto) ou não é da linguagem (está incorreto). No entanto, em se tratando de prática, no momento em que decide-se que um programa está incorreto, um bom compilador deverá

ser capaz de avisar sobre tal erro e de alguma forma, ajudar o usuário a corrigi-lo. Se faz necessário que o tratamento de erros inclua mensagens informativas e uma recuperação, para que a análise possa continuar e assim outros erros sejam sinalizados.

Em [AHO, 1995] vemos que a análise se constitui em 3 fases:

- *Análise Linear*, na qual um fluxo de caracteres constituindo um programa é lido da esquerda para a direita e agrupado em *tokens*, que são sequências de caracteres tendo um significado coletivo.
- *Análise Hierárquica*, na qual os caracteres ou *tokens* são agrupados hierarquicamente em coleções aninhadas com significado coletivo.
- *Análise Semântica*, na qual certas verificações são realizadas a fim de se assegurar que os componentes de um programa se combinam de forma significativa.

Sabe-se da possibilidade de total representação da sintaxe de uma linguagem de programação através de uma gramática sensível ao contexto [Wijngaarden, 1969]. No entanto, não há algoritmos práticos para tratar estas gramáticas, fazendo com que haja preferência em usar gramáticas livres de contexto. Sendo assim, fica claro que a separação entre análise sintática e análise semântica é dependente da implementação.

Sendo assim, a análise léxica tem como finalidade separar e identificar os elementos componentes do programa fonte, o qual estes geralmente, são especificados através de expressões regulares. A análise sintática deve reconhecer a estrutura global do programa, descrita através de gramáticas livre de contexto. A análise semântica se encarrega da verificação das regras restantes. Essas regras tratam quase sempre da verificação de que os objetos são usados no programa da maneira prevista em suas declarações, por exemplo verificando que não há erros de tipos [Rangel, 1999].

Capítulo 2

Portugol

2.1 Introdução

O Portugol como linguagem não é tão bem definida, ficando muito mais sujeita a uma linguagem de aprendizado, voltada para o ensino e não para projetos de grande porte. Algumas vezes é somente uma tradução simplificada de Pascal para o português.

O Portugol aqui definido será baseado em [Farrer,]. O Livro utilizado para o estudo de algoritmos e de introdução a programação, é perfeito para este trabalho. O compilador desenvolvido, bem como o sistema Web, visa estudar como o processo de compilação online se dará, sendo necessário uma linguagem simples que não tome o foco do objetivo principal.

Nas seções subsequentes veremos como o portugol foi proposto para este compilador. Suas características, sua sintaxe e a estrutura da linguagem.

2.2 Estrutura da Linguagem

2.2.1 Comentários

Mais de uma forma de comentar o código será aceito, estimulando o seu uso em todo o código. O `//` usado em linguagens como C e Java e também a `#` (cerquilha), utilizada em linguagens como o Ruby são aceitos como comentários de uma só linha. Os `{}` (colchetes) são a única forma aceita de comentários para várias linhas ou comentários de bloco.

2.2.2 Tipos básicos

Os tipos básicos são simples e restritos, pois visam o ensino da lógica de programação. Isso promove também a independência de linguagem e de máquina.

Numérico

O tipo numérico resume todos os tipos para o cálculo aritmético. Não existe distinção entre números inteiros ou números reais (ponto flutuante). Caso haja a necessidade de que algum algoritmo utilize uma propriedade específica de algum destes tipos, deverá esta ser obtida através de funções. A vírgula é o símbolo utilizado para a separação da parte decimal. Não se escrevem os separadores de milhar.

Literal

Este tipo é responsável por armazenar sequências de letras e símbolos — em muitas linguagens é o tipo *String*.

Lógico

Responsável pelo armazenamento das constantes verdadeiro e falso. Este tipo é muito importante em projetos de linguagens de programação, visto que ele é bastante utilizado, sendo o tipo de retorno de várias operações.

Vetores e Matrizes

É permitida a criação de vetores e matrizes de quaisquer tipos básicos. As operações realizadas com matrizes e vetores deverão ser realizadas de forma condizente com a linguagem Portugol. As operações básicas deste tipo seguem as mesmas permitidas em seus tipos básicos. A palavra-chave *matriz* é reservada para este tipo. O número inicial dos índices é sempre 1.

Exemplo:

```
1 a matriz numérico[30] // Cria uma matriz 30x1
2 b matriz lógico[3][3] // Cria uma matriz 3x3
```

Registros

É sempre definido pela palavra reservada **registro**. Representa um bloco composto por declarações de variáveis. Se encerra com a palavra **fimregistro**.

Exemplo:

```
1 registro
2   num1, num2 numérico
3   estaPresente lógico
4   endereco literal
5 fimregistro
```

2.2.3 Variáveis e Identificadores

As variáveis são parte importante de uma linguagem. Qualquer identificador, sendo este válido, tem como primeiro símbolo uma letra. Números e _ (sublinha) são aceites após o primeiro caractere. O tamanho máximo válido é definido em 32 caracteres.

É necessário lembrar que alguns símbolos, normalmente ignorados ou proibidos em outras linguagens, são aceites em Portugol. Símbolos estes presentes na língua portuguesa, como por exemplo, os acentos: á, ê, í, à, ã, ..., ç.

Vale ressaltar da diferença existente entre variáveis declaradas com acento e variáveis sem acento. Ou seja, a variável 'pé' é diferente da variável 'pe'.

2.2.4 Declaração de variáveis

Não existe um local específico para declaração de variáveis, mas fica a recomendação de que esta seja feita no início de uma seção, seja o programa principal ou um procedimento/função. A palavra-chave **declare** será utilizada para declarar variáveis em Portugol. É possível declarar variáveis de mesmo tipo em uma mesma linha, nunca variáveis de tipos diferentes.

Exemplo:

```
1 declare num1, num2, num3 numérico
2 declare palavra literal
3 declare a, b, c lógico
4 declare r registro
5     dia, mes numérico
6     nome literal
7     estaPresente lógico
8     fimregistro
```

Exemplo de declaração inválida:

```
1 declare num1 numérico, palavra literal
```

A visibilidade de uma palavra segue o mesmo padrão de outras conhecidas linguagens, ou seja, é dependente do escopo a qual foi definida.

2.2.5 Atribuição

A atribuição de valores a uma variável será denotada pelo símbolo ":=". Será permitido apenas a atribuição de um valor por vez. Uma atribuição será válida quando possuir um identificador válido do lado esquerdo e uma expressão válida e, do mesmo tipo, do lado direito.

Exemplo:

```

1 declare num1, num2 numérico
2 declare palavra literal
3
4 num1 := 3
5 num2 := 1,5
6 palavra := "Olá Mundo"

```

2.2.6 Operadores

Operadores Aritméticos

Os operadores são:

Tabela 2.1 - Operadores Aritméticos

Operação	Símbolo
Adição	+
Subtração	-
Multiplicação	*
Divisão	/

Operadores Lógicos

Os Operadores são:

Tabela 2.2 - Operadores Lógicos

Operação	Símbolo
inversão - não	não
junção	e
disjunção	ou

Operadores Relacionais

Os Operadores são:

Tabela 2.3 - Operadores Relacionais

Operação	Símbolo
Igualdade	=
Maior	>
Menor	<
Maior ou Igual	>=
Menor ou Igual	<=
Diferença	!= ou <>

Operadores Relacionais

Os Operadores são:

Tabela 2.4 - Operadores Literais

Operação	Símbolo
Concatenação	+
Igualdade	=
Maior	>
Menor	<
Maior ou Igual	>=
Menor ou Igual	<=
Diferença	!= ou <>

2.2.7 Estrutura Sequencial

O programa tem início com a palavra **programa** e tem fim com a palavra **fimprograma**. Caso queira importar outros módulos, estes devem ser feitos através de comando, após a declaração do programa.

Em Portugol não se faz necessária a utilização de algum símbolo especial ou visual para marcar o fim de linha. Basta que o "Retorno" — mais conhecido pela tecla "ENTER" — seja pressionado. Este símbolo é definido na posição 13 da tabela ASCII.

Qualquer que seja o símbolo ou comando declarado fora do escopo do programa, será considerado erro. Funções e procedimentos devem ser declarados externamente e então, importados pelo programa que necessite utilizá-los. Isto fará com que se estimule a modularização de código, bem como a criação de arquivos de menor tamanho e mais fáceis de entender.

2.2.8 Estruturas Condicionais

São definidas duas: **se** e o **caso**.

O **se** é utilizado quando se quer avaliar uma condição ou expressão lógica e retorna obrigatoriamente um valor lógico. Conforme o retorno a execução passa para o bloco **então** — caso seja verdadeiro — ou para o bloco **senão** — caso seja falso.

É definido também em Portugol o **fimentão** e o **senão**, responsáveis pelo fechamento dos blocos **então** e **senão** respectivamente.

Exemplo:

```
1 se num1 > num2
2     então num1 = 1
3     fimentão
```

```
4      senão num2 = 2
5      fimsenão
6  fimse
```

A estrutura **caso** permite avaliar vários valores e admite variável dos tipos lógicos, numérico e literal.

```
1  caso tamanho
2      10: i := 10
3      11: i := 11
4      12: i := 12
5  fimcaso
```

2.2.9 Estruturas de Repetição

Três estruturas são definidas em Portugol: **faça**, **enquanto**, **repita**.

O **faça** é similar ao *for* de outras linguagens. Sua construção é demonstrada abaixo:

```
1  faça i de 1 até 10
2      <instruções>
3  fimfaça
```

O bloco acima irá executar as instruções dez vezes, incrementando o valor do "i" de 1 até 10. É recomendável como boas práticas de programação que a variável de controle — no bloco acima é o "i" — não tenha seu valor alterado pelas instruções internas. O valor do "i" é incrementado em 1 sempre quando se chega ao "fimfaça". O valor do "i" após o "fimfaça" é igual ao valor máximo estipulado no início, no caso 10.

O **enquanto** substitui o *while* de outras linguagens de programação. Exemplo:

```
1  enquanto num1 > num2
2      <instruções>
3  fimenquanto
```

A estrutura de repetição **enquanto** executa o bloco de código até que a condição seja falsa. Pode acontecer das instruções nem serem executadas, caso a condição seja falsa logo de início.

A estrutura **repita** tem como característica o fato de ser básica. A verificação e a interrupção do bloco de código é livre, sendo especificada pelo comando **interrompa**. Veja:

```
1  repita
2      <comandos>
3      interrompa
4  fimrepita
```

Após a leitura do interrompa, a execução do programa se dará para o trecho de código abaixo do "fimrepita".

Como pode ser observado, a maioria das estruturas definidas são fechadas sempre utilizando o "*fimestrutura*". A não utilização de tal comando acarretará em um erro de código, podendo ser este léxico ou sintático.

2.2.10 Procedimentos e Funções