

LFApp: Um Aplicativo Móvel para o Ensino de Linguagens Formais e Autômatos

Juventino Neto, Ricardo Terra

Departamento de Ciência da Computação,
Universidade Federal de Lavras (UFLA), Brasil

jneto@computacao.ufla.br, terra@dcc.ufla.br

***Resumo.** Linguagens Formais e Autômatos (LFA) é uma importante área da computação que aborda modelos matemáticos que possibilitam a especificação e reconhecimento de linguagens, suas propriedades e características. Embora o conhecimento sólido em LFA seja de extrema importância para a formação de um bacharel em Ciência da Computação e áreas afins, os algoritmos e técnicas abordadas na disciplina são complexos e de difícil assimilação. Diante disso, este trabalho apresenta LFApp, um aplicativo móvel para ensino de LFA. O aplicativo – desenvolvido para celulares e tablets com sistema operacional (SO) Android – provê aos alunos não somente a resolução de problemas envolvendo Linguagens Regulares e Linguagens Livres de Contexto, mas também, uma interface com caráter acadêmico que descreve e ilustra cada etapa da execução dos algoritmos de modo a apoiar os alunos no processo de aprendizagem.*

1. Introdução

Linguagens Formais e Autômatos (LFA) é uma importante área da computação que aborda modelos matemáticos que possibilitam a especificação e reconhecimento de linguagens, suas propriedades e características [8]. Considerando a importância que um conhecimento sólido de LFA tem sobre o perfil de um bacharel da área de Ciência da Computação e observando um nível de reprovação médio de 45% na disciplina de LFA ministrada na UFLA [9], foi proposta a criação de uma plataforma de estudos que possa auxiliar os alunos no processo de aprendizagem.

Diante disso, este artigo apresenta LFApp, um aplicativo móvel para ensino de LFA que provê aos alunos não somente a resolução de problemas envolvendo Linguagens Regulares e Linguagens Livres de Contexto, mas também uma interface com caráter acadêmico – sendo essa a principal contribuição deste trabalho – que descreve e ilustra cada etapa da execução dos algoritmos de modo a apoiar os alunos no processo de aprendizagem. É importante mencionar que o aplicativo foi desenvolvido para celulares e tablets com sistema operacional (SO) Android e possui o código aberto visando a ampla divulgação, uma vez que o Android é o SO mais utilizado por dispositivos móveis [6].

LFApp foca em Linguagens Livres de Contexto, tendo sido implementadas as seguintes funcionalidades: identificação do tipo de gramática, derivação, remoção de recursão no símbolo inicial, remoção de produções vazias, remoção de regras de cadeia, remoção de símbolos não-terminais, remoção de símbolos não alcançáveis, Forma Normal de Chomsky, remoção de recursão direta à esquerda, remoção de recursão indireta

à esquerda, Forma Normal de Greibach e algoritmo de reconhecimento CYK (Cocke-Younger-Kasami) [8].

Este artigo está organizado como a seguir. A Seção 2 apresenta a análise do contexto em que o aplicativo melhor se adaptaria e as motivações para a sua criação, enquanto que a Seção 3 apresenta o aplicativo desenvolvido e o que espera-se com seu uso. A Seção 4 apresenta seu projeto, implementação e ilustra os testes de unidade realizados. A Seção 5 lista os trabalhos relacionados. Por fim, a Seção 6 apresenta a conclusão e o que espera-se obter com a realização deste trabalho.

2. Contextualização

Esta seção tem o objetivo de analisar o contexto de adaptação para o aplicativo desenvolvido, destacando as motivações para seu desenvolvimento. As subseções a seguir destacam os termos importantes para a compreensão correta deste trabalho.

Hierarquia de Chomsky: Proposta por Noam Chomsky em 1956 e ilustrada na Figura 1, constitui uma classificação para linguagens, gramáticas e autômatos [2].

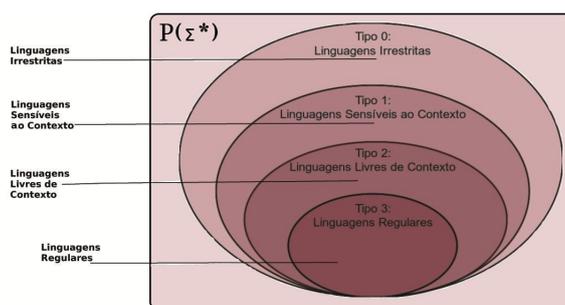


Figura 1. Hierarquia de Chomsky adaptado de [8]

A hierarquia de Chomsky é composta por quatro categorias, sendo cada uma delas um subconjunto próprio da categoria superior. As quatro categorias são: Linguagens Recursivamente Enumeráveis (0), Linguagens Sensíveis ao Contexto (1), Linguagens Livres de Contexto (2) e Linguagens Regulares (3). Os níveis 2 e 3 são amplamente empregados na descrição de linguagens de programação e na implementação de interpretadores e compiladores [1]. Este trabalho mantém como foco a categoria de nível 2, Linguagens Livres de Contexto, empregando algoritmos e técnicas que atuam sobre esse nível da hierarquia.

Gramática: É um conjunto de regras de produções no seguinte formato $\mu \rightarrow \nu$ que permitem gerar uma linguagem. A definição formal de uma gramática é uma quádrupla (V, Σ, P, S) , onde V representa o conjunto de símbolos não-terminais (variáveis), Σ o conjunto de símbolos terminais, P o conjunto de regras e S o símbolo não-terminal inicial [8]. Segundo a Hierarquia de Chomsky, a Tabela 1 descreve os quatro tipos de gramáticas, em que cada tipo representa uma categoria.

Forma Normal: GLCs são muito flexíveis na formação de suas regras, o que pode ser vantajoso no processo de criação de uma gramática. No entanto, essa flexibilidade não

Tabela 1. Hierarquia de Chomsky

Tipo	Linguagem	Gramática	Regras $\mu \rightarrow \nu$
0	Rec. Enumerável	Irrestrita (GI)	$\mu \in (V \cup \Sigma)^+, \nu \in (V \cup \Sigma)^*$
1	Recursiva	Sensível ao Contexto (GSC)	$\mu \in (V \cup \Sigma)^+, \nu \in (V \cup \Sigma)^+, \mu \leq \nu $
2	Livre de Contexto	Livre de Contexto (GLC)	$\mu \in V, \nu \in (V \cup \Sigma)^*$
3	Regular	Regular (GR)	$\mu \in V, \nu \in \lambda \mid \Sigma \mid \Sigma V$

define uma estrutura padrão para esse tipo de gramática, o que não permite estabelecer relações gerais sobre a gramática, derivações e linguagens [8]. Com o objetivo de tornar as GLCs mais concisas, formas normais impõem regras e simplificam a representação, sem alterar o poder de expressão da linguagem. A Tabela 2 descreve a Forma Normal de Chomsky (FNC) e Greibach (FNG), e o formato de suas regras.

Tabela 2. Formais Normais

Forma Normal de Chomsky (FNC)		Forma Normal de Greibach (FNG)	
$A \rightarrow BC$	onde $B, C \in V$	$A \rightarrow aA_1A_2...A_n$	onde $A_i \in V - \{S\}$
$A \rightarrow a$	onde $a \in \Sigma$	$A \rightarrow a$	onde $a \in \Sigma$
$S \rightarrow \lambda$		$S \rightarrow \lambda$	

Derivação: Consiste no processo de substituição dos símbolos lidos em uma dada palavra pelas regras de uma gramática. Dada uma gramática G e uma palavra p , o objetivo é verificar se a partir de seu símbolo inicial S deriva-se p , i.e., $S \Rightarrow^* p$. As derivações podem ser mais à esquerda ou mais à direita. A derivação mais à esquerda consiste em substituir os símbolos mais à esquerda da forma sentencial, enquanto que, de forma análoga, a mais à direita consiste em substituir os símbolos mais à direita. O processo de derivação ocorre em todas as categorias das linguagens presentes na Hierarquia de Chomsky [8].

3. Solução Proposta

Esta seção apresenta LFApp, um aplicativo que provê a implementação de diversas funcionalidades em um ambiente de caráter acadêmico. O objetivo é prover o aluno com uma descrição sobre o processo de resolução de forma a estimular seu aprendizado. Nas subseções seguintes são descritas e ilustradas as funcionalidades implementadas. Para ilustrar as funcionalidades, adota-se a seguinte GLC, a qual foi projetada de forma a ser possível observar as transformações propostas pela maioria dos algoritmos implementados:

$$\begin{array}{ll}
 S \rightarrow aABC \mid a \mid S & C \rightarrow cC \mid c \\
 A \rightarrow aA \mid \lambda & D \rightarrow abc \mid E \\
 B \rightarrow bcB \mid bc & E \rightarrow Ea \mid a
 \end{array}$$

A Figura 2 apresenta a tela inicial do LFApp com a gramática de referência.

Identificação do tipo de gramática: Essa funcionalidade consiste em identificar em qual nível da hierarquia de Chomsky a gramática se encontra. A Figura 3 ilustra o resultado dessa funcionalidade na gramática de referência. Basicamente, o algoritmo analisa o formato das regras da gramática e a classifica no nível mais restrito possível. Nesse exemplo, além de classificar a gramática como GLC, proveu-se uma tabela de classificação de regras e um exemplo de uma regra que não se enquadra na definição do nível superior (GR).

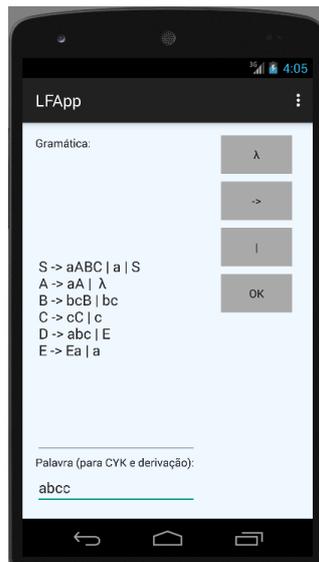


Figura 2. Tela inicial do sistema com gramática de referência

Identificação da gramática

A classificação de uma gramática é feita pelo tipo de suas regras ($u \rightarrow v$). A tabela abaixo mostra o formato de regras características de cada nível:

(3) GR	$u \in V$	$v \in \lambda \mid \Sigma \mid \Sigma V$
(2) GLC	$u \in V$	$v \in (V \cup \Sigma)^*$
(1) GSC	$u \in (V \cup \Sigma)^+$	$v \in (V \cup \Sigma)^+$
(0) GI	$u \in (V \cup \Sigma)^+$	$v \in (V \cup \Sigma)^*$

Resultado:

- Na gramática informada, a regra $S \rightarrow aABC$ não pertence ao conjunto das gramáticas regulares. Logo, a gramática inserida é uma Gramática Livre de Contexto (GLC).

Derivação

Resultado:

S => aABC
 => aλBC
 => abcC
 => abcc

Palavra: abcc

Ambiguidade: Uma Gramática Livre de Contexto é ambígua caso possua mais de uma derivação mais à esquerda. Para a gramática inserida foram realizadas até 1000 tentativas de encontrar uma ambiguidade, porém, não foi encontrada outra forma de gerar a palavra abcc.

Figura 3. Identificação do tipo da gramática

Figura 4. Derivação

Derivação: Essa funcionalidade realiza o processo de derivação mais à esquerda, i.e., tenta gerar uma dada palavra a partir de uma dada gramática. Essa funcionalidade também avalia se a gramática é ambígua. Verificar se uma gramática é ambígua é considerado um problema indecidível, logo, são realizadas até 1.000 tentativas de encontrar formas diferentes de gerar a mesma sentença. Em caso positivo, a gramática é classificada como ambígua, caso contrário, nada se conclui. A Figura 4 ilustra o processo de derivação sobre a gramática de referência e a palavra *abcc*.

Remoção de recursão no símbolo inicial: Essa funcionalidade consiste em transformar a gramática de forma que o símbolo inicial se limite a iniciar derivações. A Figura 5 ilustra o resultado dessa funcionalidade aplicada à gramática de referência. O algoritmo para remoção de recursão no símbolo inicial consiste em verificar se a gramática gera uma derivação do tipo $S \Rightarrow^* \alpha S \beta$. Caso possua, é criado um novo símbolo não-terminal S' e inserida uma nova regra do tipo $S' \rightarrow S$, removendo então a recursão no símbolo inicial. Caso contrário, a gramática não possui recursão no símbolo inicial e não é alterada. Nesse exemplo, além de transformar a gramática, proveu-se uma explicação do problema e do algoritmo descrito acima [8].

Remoção de Recursão no Símbolo Inicial

Resultado:

```

S' -> S
S -> a | aABC | S
A -> λ | aA
B -> bc | bcB
C -> c | cC
D -> abc | E
E -> Ea
    
```

Algoritmo:
 O símbolo inicial deve se limitar a iniciar derivações, não podendo ser uma variável recursiva. Logo, não deve ser possível ter derivações do tipo $S \rightarrow^* \alpha S \beta$.

A gramática $G = (V, \Sigma, P, S)$ possui o símbolo inicial S recursivo. Logo, existe uma GLC $G' = (V \cup \{S'\}, \Sigma, P \cup \{S' \rightarrow S\}, S')$, tal que $L(G') = L(G)$ e o novo símbolo inicial S não é recursivo.

Figura 5. Remoção de recursão no símbolo inicial

Remoção de produções vazias: Durante o processo de derivação de uma palavra, a forma sentencial pode conter variáveis que não contribuem para a formação de uma sentença, as quais são classificadas como produções vazias. Essa funcionalidade remove as produções vazias em uma GLC, garantindo a não contratibilidade da gramática. Para a realização dessa transformação, implementou-se o algoritmo λ -rules [8] que torna a gramática não-contrátil ou essencialmente não-contrátil. A Figura 6 ilustra o resultado da execução dessa funcionalidade na gramática de referência. Além da transformação realizada sobre a gramática, o aplicativo proveu uma explicação do processo de determinação das variáveis anuláveis (1), da substituição das variáveis (2) e da remoção dos símbolos λ (3).

Remoção de Produções Vazias

Resultado:

```

S -> aBC | a | aABC | S
A -> aA | a
B -> bcB | bc
C -> cC | c
D -> abc | E
E -> a | Ea
    
```

Algoritmo:
 O algoritmo para remoção de regras λ consiste em 3 passos:

(1) Determinar o conjunto das variáveis anuláveis.
 $NULL = \{A \mid \{A \rightarrow \lambda\} \in P\}$
repita
 PREV = NULL
para cada $A \in V$ **faça**
 se $A \rightarrow w e w \in PREV^*$ **faça**
 $NULL = NULL \cup \{A\}$
até $NULL == PREV$

	NULL	PREV
(0)	{A}	\emptyset
(1)	{A}	{A}

(2) Adicionar regras em que as ocorrências de variáveis nulas são omitidas. Por exemplo, assumamos a regra $A \rightarrow BABA$ e B é uma variável anulável. Logo, são inseridas as seguintes regras: $A \rightarrow ABa$, $A \rightarrow BAa$ e $A \rightarrow Aa$.

```

S -> aBC | a | aABC | S
A -> λ | aA | a
B -> bcB | bc
C -> cC | c
D -> abc | E
E -> a | Ea
    
```

(3) Remover as regras λ .

```

S -> aBC | a | aABC | S
A -> aA | a
B -> bcB | bc
C -> cC | c
D -> abc | E
E -> a | Ea
    
```

Figura 6. Remoção de produções vazias

Remoção de regras de cadeia: Ocasionalmente ocorrem em GLCs regras com o formato $A \rightarrow B$, as quais são denominadas regras de cadeia. Essas regras não contribuem de forma direta para o processo de formação de uma palavra e devem ser removidas por meio de um algoritmo que localize as regras de cadeia e as substitua pelas produções referentes a variável destino da cadeia [8]. Por exemplo, em uma gramática que possua a seguinte regra $A \rightarrow B$, será realizada a substituição de B em A por todas as produções existentes em B . A Figura 7 ilustra o processo de remoção de regras de cadeia na gra-

mática de referência. Além da transformação realizada sobre a gramática, o aplicativo proveu o conjunto de cadeias de cada variável (1), o destaque das regras de cadeia (2) e uma explicação do processo de substituição das mesmas (3).

Remoção de Regras de Cadeia

Resultado:

```
S -> a | aABC
A -> λ | aA
B -> bc | bcB
C -> c | cC
D -> a | Ea | abc
E -> a | Ea
```

Algoritmo:
A remoção de regras de cadeia substitui as ocorrências de uma cadeia diretamente pelas regras da variável renomeada.

(1) O primeiro passo do algoritmo é montar as cadeias de cada variável.

```
CHAIN(A) = {A}
PREV = ∅
repita
NEW = CHAIN(A) - PREV
PREV = CHAIN(A)
para cada B ∈ NEW faça
para cada B → C faça
CHAIN(A) = CHAIN(A) ∪ {C}
até CHAIN(A) == PREV
```

(2) Destacar as cadeias encontradas.

```
S -> a | aABC | S
A -> λ | aA
B -> bc | bcB
C -> c | cC
D -> abc | E
E -> a | Ea
```

(3) Substituir as cadeias encontradas.

```
S -> a | aABC
A -> λ | aA
B -> bc | bcB
C -> c | cC
D -> a | Ea | abc
E -> a | Ea
```

Variável	Cadeia
(S)	{S}
(A)	{A}
(B)	{B}
(C)	{C}
(D)	{D, E}
(E)	{E}

Figura 7. Remoção de regras de cadeia

Remoção de variáveis que não geram terminais: Essa funcionalidade consiste em remover as variáveis que não geram terminais da gramática. Para ilustrar o funcionamento dessa funcionalidade sem grandes alterações na gramática de referência, a regra $E \rightarrow a$ foi removida, fazendo com que a variável E não gere terminais. A Figura 8 ilustra o processo de remoção de variáveis que não geram terminais na gramática de referência. Além da transformação realizada sobre a gramática, o aplicativo apresenta a construção do conjunto $TERM$ que possui as variáveis que produzem terminais diretamente ou indiretamente (1), e o processo de remoção das variáveis que não estão em $TERM$ (2).

Remoção de símbolos não alcançáveis: Essa funcionalidade remove símbolos não alcançáveis a partir do símbolo inicial S , i.e., aqueles não-terminais A que não aparecem em uma derivação da seguinte forma $S \Rightarrow^* \alpha A \beta$. A Figura 9 ilustra o processo de remoção de variáveis inalcançáveis na gramática de referência [8]. Além da transformação realizada sobre a gramática, o aplicativo apresenta a construção do conjunto $REACH$ que possui as variáveis alcançáveis (1) e o processo de remoção das variáveis que não estão em $REACH$ (2).

Forma Normal de Chomsky (FNC): Essa funcionalidade padroniza as regras da gramática em um dos formatos da FNC, como descrito na Seção 2. O algoritmo requer que a GLC de entrada não possua símbolo inicial recursivo, produções vazias, regras de cadeia e símbolos inúteis (variáveis que não geram terminais e símbolos não alcançáveis), logo, a ordem da aplicação dos algoritmos é importante. Portanto, o aplicativo pode automaticamente transformar a gramática para satisfazer as pré-condições do algoritmo utilizado [8]. A Figura 10 ilustra o processo de transformação da gramática de referência para a FNC. Além da transformação realizada sobre a gramática, o aplicativo destaca as regras não válidas na FNC (1) e as novas regras criadas (2).

Remoção de variáveis que não geram terminais

Resultado:

S → a | aABC | S
 A → λ | aA
 B → bc | bcB
 C → c | cC
 D → abc

Algoritmo:

Remove as regras que não geram terminais. Consiste de dois passos:

(1) Determinar quais variáveis geram terminais direta e indiretamente.

TERM = {A | existe uma regra $A \rightarrow w \in P$, com $w \in \Sigma^*$ }

repita

PREV = TERM

para cada A ∈ V **faça**

se $A \rightarrow w \in P$ e $w \in (PREV \cup \Sigma)^*$ **então**

TERM = TERM ∪ {A}

até PREV == TERM

	TERM	PREV
(0)	{D, A, S, B, C}	{}
(1)	{D, A, S, B, C}	{D, A, S, B, C}

(2) Remover as variáveis que não estão em {S, A, B, C, D}, i.e., {E}.

S → a | aABC | S

A → λ | aA

B → bc | bcB

C → c | cC

D → abc | E

E → Ea

Figura 8. Remoção de símbolos não-terminais

Remoção de símbolos não alcançáveis

Resultado:

S → a | aABC | S
 A → λ | aA
 B → bc | bcB
 C → c | cC

Algoritmo:

Remover as variáveis não alcançáveis no processo de derivação de uma palavra.

(1) Determinar quais variáveis são alcançáveis a partir do símbolo inicial S.

REACH = {S}

PREV = ∅

repita

NEW = REACH - PREV

PREV = REACH

para cada A ∈ NEW **faça**

para cada A → w **faça**

adicione as variáveis de w em REACH

até REACH == PREV

	REACH	PREV	NEW
(0)	{S}	{}	{}
(1)	{A, S, B, C}	{S}	{S}
(2)	{A, S, B, C}	{A, S, B, C}	{A, B, C}

(2) Remover as variáveis que não estão em {S, A, B, C}, i.e., {D, E}.

S → a | aABC | S

A → λ | aA

B → bc | bcB

C → c | cC

D → abc | E

E → a | Ea

Figura 9. Remoção de símbolos não alcançáveis

Forma Normal de Chomsky

Remoção de recursão no símbolo inicial:

Remoção de produções vazias:

Remoção de regras de cadeia:

Remoção de símbolos não terminais:

Remoção de símbolos não alcançáveis:

Forma Normal de Chomsky:

Resultado:

S' → T₃T₆ | T₃T₄ | a T₂ → c
 A → a | T₃A T₃ → a
 B → T₁T₅ | T₁T₂ T₄ → BC
 C → T₂C | c T₅ → T₂B
 T₁ → b T₆ → AT₇
 T₇ → BC

Algoritmo:

Uma GLC $G = (V, \Sigma, P, S)$ está na Forma Normal de Chomsky se suas regras tem uma das seguintes formas

- $A \rightarrow BC$ onde $B, C \in V - \{S\}$

- $A \rightarrow a$ onde $a \in \Sigma$

- $S \rightarrow \lambda$

(1) Identificar as regras que não estão na Forma Normal de Chomsky.

S' → aBC | a | aABC

A → aA | a

B → bc | bcB

C → c | cC

(2) Transformar tais regras em um dos formatos válidos

S' → T₃T₆ | T₃T₄ | a

A → a | T₃A

B → T₁T₅ | T₁T₂

C → T₂C | c

T₁ → b

T₂ → c

T₃ → a

T₄ → BC

T₅ → T₂B

T₆ → AT₇

T₇ → BC

Figura 10. Forma Normal de Chomsky

Remoção de recursão direta à esquerda: Essa funcionalidade remove recursões diretas à esquerda, as quais podem produzir *loops* infinitos em analisadores sintáticos descendentes [1]. A Figura 11 ilustra o processo de remoção de recursão direta à esquerda na gramática de referência. Além da transformação realizada sobre a gramática, o aplicativo identifica as regras com recursão (1) e o processo de remoção de tais regras (2).

Remoção de Recursão Direta à Esquerda

Resultado:

$ \begin{aligned} S &\rightarrow a \mid aABC \\ A &\rightarrow \lambda \mid aA \\ B &\rightarrow bcB \mid bc \\ C &\rightarrow cC \mid c \\ D &\rightarrow abc \mid E \\ E &\rightarrow a \mid aZ_1 \\ Z_1 &\rightarrow aZ_1 \mid a \end{aligned} $	$ \begin{aligned} S &\rightarrow a \mid aABC \\ A &\rightarrow \lambda \mid aA \\ B &\rightarrow bc \mid bcB \\ C &\rightarrow c \mid cC \\ D &\rightarrow abc \mid E \\ E &\rightarrow a \mid Ea \end{aligned} $
---	--

(2) O segundo passo é resolver a recursão.

Algoritmo:

Recursividade direta à esquerda pode produzir "loops infinitos" em analisadores sintáticos descendentes (top-down).

(1) O primeiro passo é identificar a recursão.

Suponha a regra genérica diretamente recursiva à esq.:

$$A \rightarrow A\mu_1 \mid A\mu_2 \mid \dots \mid A\mu_m \mid v_1 \mid v_2 \mid \dots \mid v_n$$

Regra equivalente não-recursiva à esquerda:

$$\begin{aligned}
 A &\rightarrow v_1 \mid v_2 \mid \dots \mid v_n \mid v_1Z \mid v_2Z \mid \dots \mid v_nZ \\
 Z &\rightarrow \mu_1Z \mid \mu_2Z \mid \dots \mid \mu_mZ \mid \mu_1 \mid \mu_2 \mid \dots \mid \mu_m
 \end{aligned}$$

$$Z_1 \rightarrow aZ_1 \mid a$$

Figura 11. Remoção de recursão direta à esquerda

Remoção de recursão indireta à esquerda: Essa funcionalidade remove recursões indiretas à esquerda. A Figura 12 ilustra o processo de remoção de recursão indireta à esquerda na gramática de referência. No entanto, o algoritmo utilizado [8] requer que a gramática inserida não possua ciclos, o que impede a execução da funcionalidade na gramática de referência. Caso as pré-condições sejam satisfeitas, além da transformação realizada sobre a gramática, o aplicativo ordena as variáveis (1), localiza as recursões (2) e descreve o processo de alteração das regras (3).

Remoção de Recursão Direta e Indireta à Esquerda

Resultado:

$$\begin{aligned}
 S &\rightarrow a \mid aABC \mid S \\
 A &\rightarrow \lambda \mid aA \\
 B &\rightarrow bc \mid bcB \\
 C &\rightarrow c \mid cC \\
 D &\rightarrow abc \mid E \\
 E &\rightarrow a \mid Ea
 \end{aligned}$$

Algoritmo:

A gramática inserida possui ciclos.
A regra $S \rightarrow S$ é um ciclo.

Figura 12. Remoção de recursão indireta à esquerda

Forma Normal de Greibach (FNG): Essa funcionalidade padroniza as regras da gramática em um dos formatos da FNG, como descrito na Seção 2. O objetivo é garantir que não ocorra recursões diretas e indiretas à esquerda. A Figura 13 ilustra o processo de transformação da gramática de referência para a FNG. Além da transformação realizada sobre a gramática – que usualmente envolve a remoção de recursão à esquerda direta e indireta – o aplicativo destaca as regras não válidas na FNG (1) e as novas regras criadas (2).

Forma Normal de Greibach

Remoção de recursão à esquerda:

Resultado:

$S' \rightarrow T_3T_6 | T_3T_4 | a$ $T_4 \rightarrow bT_5C | bT_2C$
 $A \rightarrow a | T_3A$ $T_5 \rightarrow cB$
 $B \rightarrow T_1T_5 | T_1T_2$ $T_6 \rightarrow aT_7 | aAT_7$
 $C \rightarrow T_2C | c$ $T_7 \rightarrow bT_2C | bT_5C$
 $T_1 \rightarrow b$
 $T_2 \rightarrow c$
 $T_3 \rightarrow a$

Algoritmo:

A remoção de recursão à esquerda consiste em ordenar as variáveis da gramática e organizar as regras da forma que a variável do lado esquerdo sempre possua valor menor do que a variável do lado direito.

(1) Ordenar as variáveis da gramática.

Variável	Valor
S'	1
A	2
B	3
C	4
T1	5
T3	6
T2	7
T6	8
T7	9
T4	10
T5	11

(2) Localizar as recursões.

$S' \rightarrow T_3T_6 | T_3T_4 | a$ $T_4 \rightarrow BC$
 $A \rightarrow a | T_3A$ $T_5 \rightarrow T_2B$
 $B \rightarrow T_1T_5 | T_1T_2$ $T_6 \rightarrow AT_7$
 $C \rightarrow T_2C | c$ $T_7 \rightarrow BC$
 $T_1 \rightarrow b$
 $T_2 \rightarrow c$
 $T_3 \rightarrow a$

(3) Alterar as regras

$S' \rightarrow T_3T_6 | T_3T_4 | a$ $T_3 \rightarrow a$
 $A \rightarrow a | T_3A$ $T_4 \rightarrow bT_5C | bT_2C$
 $B \rightarrow T_1T_5 | T_1T_2$ $T_5 \rightarrow cB$
 $C \rightarrow T_2C | c$ $T_6 \rightarrow aT_7 | aAT_7$
 $T_1 \rightarrow b$ $T_7 \rightarrow bT_2C | bT_5C$
 $T_2 \rightarrow c$

Forma normal de Greibach:

Resultado:

$S' \rightarrow aT_4 | a | aT_6$ $T_3 \rightarrow a$
 $A \rightarrow aA | a$ $T_4 \rightarrow bT_5C | bT_2C$
 $B \rightarrow bT_2 | bT_5$ $T_5 \rightarrow cB$
 $C \rightarrow cC | c$ $T_6 \rightarrow aT_7 | aAT_7$
 $T_1 \rightarrow b$ $T_7 \rightarrow bT_2C | bT_5C$
 $T_2 \rightarrow c$

Algoritmo:

Uma GLC $G = (V, \Sigma, P, S)$ está na FN de Greibach se suas regras têm uma das seguintes formas:

$- A \rightarrow aA_1A_2A_3 \dots A_n$ onde $a \in \Sigma$ e $A_1 \dots A_n \in V - \{S\}$
 $- A \rightarrow a$ onde $a \in \Sigma$
 $- A \rightarrow \lambda$

(1) Localizar as regras que não estão na Forma Normal de Greibach.

$S' \rightarrow T_3T_6 | T_3T_4 | a$ $T_3 \rightarrow a$
 $A \rightarrow a | T_3A$ $T_4 \rightarrow bT_5C | bT_2C$
 $B \rightarrow T_1T_5 | T_1T_2$ $T_5 \rightarrow cB$
 $C \rightarrow T_2C | c$ $T_6 \rightarrow aT_7 | aAT_7$
 $T_1 \rightarrow b$ $T_7 \rightarrow bT_2C | bT_5C$
 $T_2 \rightarrow c$

(2) Transformar tais regras em um dos formatos válidos

$S' \rightarrow aT_4 | a | aT_6$ $T_6 \rightarrow aT_7 | aAT_7$
 $A \rightarrow aA | a$ $T_7 \rightarrow bT_2C | bT_5C$
 $B \rightarrow bT_2 | bT_5$
 $C \rightarrow cC | c$
 $T_1 \rightarrow b$
 $T_2 \rightarrow c$
 $T_3 \rightarrow a$
 $T_4 \rightarrow bT_5C | bT_2C$
 $T_5 \rightarrow cB$

Figura 13. Forma Normal de Greibach

Algoritmo de reconhecimento CYK: Essa funcionalidade verifica se uma palavra pode ser gerada por uma GLC. O algoritmo CYK utiliza de uma abordagem *bottom-up* com o intuito de analisar a sentença associando-a com as regras da gramática. A Figura 14 ilustra o processo de reconhecimento da palavra *abcc* na gramática de referência. Além da construção da matriz triangular inferior, o aplicativo ilustra o passo-a-passo de sua construção. Nesse exemplo, o símbolo inicial da gramática (S') encontra-se no topo da matriz, logo pode-se concluir que a palavra *abcc* pertence a gramática.

CYK

Resultado:

	{S', T ₆ }			
	-		{T ₄ , T ₇ }	
	-		{B}	{C}
	{A, S', T ₃ }	{T ₁ }	{C, T ₂ }	{C, T ₂ }
a	b	c	c	c

Algoritmo:

(1) O primeiro passo do algoritmo é adicionar à tabela as variáveis que produzem os respectivos terminais diretamente.

Há alguma regra que gere a diretamente? Sim.

$T_3 \rightarrow a, A \rightarrow a, S' \rightarrow a$.

	{A, S', T ₃ }	{T ₁ }	{C, T ₂ }
a	b	c	c

(2) O segundo passo do algoritmo é adicionar à tabela as variáveis que produzem as sentenças de tamanho dois.

Há alguma regra que gere AT_1, ST_1, T_3T_1 ? Não.

		{B}	{C}
	{A, S', T ₃ }	{T ₁ }	{C, T ₂ }
a	b	c	c

(3) O terceiro passo do algoritmo é adicionar à tabela as variáveis que produzem as sentenças de tamanho três.

Há alguma regra que gere $AB, S'B, T_3B, -C, -T_2$? Não.

		{T ₄ , T ₇ }	
		{B}	{C}
	{A, S', T ₃ }	{T ₁ }	{C, T ₂ }
a	b	c	c

E assim por diante.

Figura 14. Reconhecimento CYK

4. Projeto, Implementação e Testes

A Figura 15 apresenta as duas principais classes do aplicativo: Grammar que representa uma gramática de qualquer nível da Hierarquia de Chomsky que é composta por um conjunto de Rule que representa uma regra do tipo $\mu \rightarrow \nu$. Objetos Grammar são imutáveis, logo todas as funcionalidades sempre retornam um novo objeto Grammar como resposta.

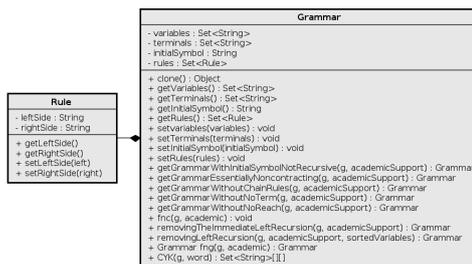


Figura 15. Diagrama de classes

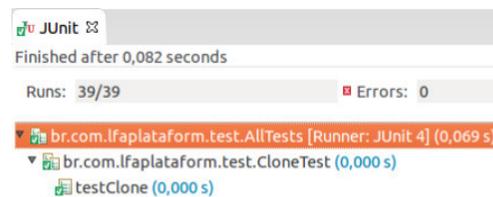


Figura 16. Testes de unidade

Com o objetivo de evitar erros de codificação, todos os algoritmos desenvolvidos foram testados através de testes de unidade, que verificam cada unidade da aplicação [7], garantindo o correto funcionamento do LFApp. Além de garantir o comportamento esperado por parte dos algoritmos, os testes também asseguram que a implementação esteja organizada de forma a permitir uma boa manutenibilidade e extensibilidade. Portanto, utilizando o *framework* JUnit¹, foram desenvolvidos 39 casos de testes, os quais são executados antes da publicação de uma nova versão do aplicativo. A Figura 16 ilustra um *screenshot* da execução de um teste de regressão no aplicativo. O teste de regressão consiste em submeter as versões mais recentes do software aos mesmos testes que as versões anteriores já foram submetidas [7]. O objetivo é garantir que as funcionalidades previamente existentes permanecem corretas, i.e., o software não regrediu.

5. Trabalhos Relacionados

Existem diversas ferramentas similares à proposta neste artigo. Yandre et al. [3] propuseram a ferramenta SCTMF (Software para a Criação e Teste de Modelos Formais) com o objetivo de auxiliar o ensino de LFA. A ferramenta consiste em um software para a criação e testes de modelos formais em máquinas aceitadoras da hierarquia de Chomsky, tais como Autômatos Finitos (AFs), Autômatos com Pilha (APs), Autômatos Linearmente Limitados (ALLs) e Máquinas de Turing (MTs). Em contraste com LFApp, seu principal foco é em máquinas aceitadoras, sem possuir algoritmos sobre gramáticas.

Menezes et al. [10] propuseram *Language Emulator*, um ambiente com foco em Linguagens Regulares (LRs) que apoia o ensino de LFA. Dentre suas funcionalidades estão algoritmos sobre Expressão Regular (ER), Gramática Regular (GR), AFs, e máquinas de Mealy e Moore. A ferramenta é completa no que se refere à classe das LRs, implementando as técnicas e algoritmos abordados na área de conhecimento. No entanto, em contraste com LFApp, não possui implementação de modelos referentes à classe das LLCs.

Jukemura et al. [5] descrevem GAM, uma ferramenta que simula diferentes tipos de máquinas abstratas para apoiar o ensino de Teoria da Computação. Dentre suas funcionalidades estão simulação de AFs, transformação de AFND para AFD, transformação de AFD para ER, transformação de ER para AFND- λ , simulação de APs e simulação de MTs. Mesmo possuindo funcionalidades no nível de LLCs (e.g., simulação de AP), o qual é o foco de LFApp, GAM não possui funcionalidades que lidam com transformações de gramáticas e formas normais.

¹<http://www.junit.org>

Muito conhecida no âmbito de Teoria da Computação e LFA, a ferramenta JFLAP disponibiliza funcionalidades com foco principal em autômatos. Para a classe das LLCs, o JFLAP disponibiliza quatro tipos de transformações, sendo elas: conversão de APs para GLCs e mais três algoritmos para conversão de GLCs em APs [4]. Mesmo possuindo foco na classe das LLCs, ao contrário da ferramenta descrita neste trabalho, o JFLAP não apresenta como foco principal as transformações de GLCs, o que é amplamente abordado neste trabalho.

6. Considerações Finais

Linguagens Formais e Autômatos (LFA) é uma importante disciplina da computação que aborda modelos matemáticos que possibilitam a especificação e reconhecimento de linguagens, suas propriedades e características. No entanto, os algoritmos e técnicas abordadas na disciplina são complexos e de difícil assimilação, o que implica em um alto índice de reprovações. Diante disso, este trabalho apresentou LFApp, um aplicativo – desenvolvido para celulares e *tablets* com SO Android – que provê aos alunos não somente a resolução de problemas envolvendo Linguagens Regulares e Linguagens Livres de Contexto, mas também, uma interface com alto caráter acadêmico que descreve e ilustra cada etapa da execução dos algoritmos de modo a apoiar os alunos no processo de aprendizagem.

Como principais contribuições do LFApp, espera-se (i) prover um ambiente de estudos interativo para os alunos de forma a auxiliar no processo de ensino e aprendizagem de LFA e (ii) promover a ideia do uso de aplicativos móveis educacionais. Como trabalhos futuros, pretende-se (i) estender LFApp para máquinas aceitadoras desenvolvendo, por exemplo, algoritmos para simulação de AF/AP/ALL/MT, conversão de AFND para AFD, minimização de AFD, etc., (ii) expandir a ferramenta para abordar todos os níveis da Hierarquia de Chomsky e (iii) desenvolver versões para outras plataformas (iOS e Windows Mobile).

A ferramenta LFApp e seu código fonte estão publicamente disponíveis em:

<http://www.dcc.ufla.br/~terra/lfapp>

Agradecimentos: Este trabalho foi apoiado pela FAPEMIG.

Referências

- [1] Alfred V Aho, Ravi Sethi, and Jeffrey D Ullman. *Compiladores: Princípios, técnicas e ferramentas*. LTC, 2 edition, 2008.
- [2] Noam Chomsky. Three models for the description of language. *IRE Transactions on Information Theory*, 2(3):113–124, 1956.
- [3] Yandre M. e G. da Costa, Rafael C. de Meneses, and Flávio R. Uber. Uma ferramenta para auxílio didático no ensino de teoria da computação. In *XVI Workshop sobre Educação em Computação (WEI)*, pages 208–217, 2008.
- [4] Eric Gramond and Susan H Rodger. Using JFLAP to interact with theorems in automata theory. *ACM SIGCSE Bulletin*, 31(1):336–340, 1999.
- [5] Anibal S. Jukemura, Hugo A. D. do Nascimento, and Joaquim Q. Uchôa. GAM: Um simulador para auxiliar o ensino de linguagens formais e de autômatos. *XIII Workshop sobre Educação em Computação (WEI)*, pages 2432–2443, 2005.

- [6] Farhad Manjoo. A murky road ahead for Android, despite market dominance. *The New York Times*, 2015.
- [7] Roger S Pressman. *Engenharia de software*. Bookman, 7 edition, 2011.
- [8] Thomas A. Sudkamp. *Languages and machines: an introduction to the theory of Computer Science*. Addison-Wesley, 3 edition, 2005.
- [9] Ricardo Terra. Índice de reprovação na disciplina de Linguagens Formais e Autômatos. Technical report, UFLA, 2016.
- [10] Luiz Filipe M. Vieira, Marcos Augusto M. Vieira, and Newton J. Vieira. Language emulator, a helpful toolkit in the learning process of computer theory. In *35th SIGCSE Technical Symposium on Computer Science Education (SIGCSE)*, pages 135–139, 2004.