

Projeto 01 - Mapa Conceitual sobre Linguagens Formais

Caio Nery, Gustavo de Oliveira, Linsmar Vital, Luca Argolo e Thiago Vieira

GLOSSÁRIO DO MAPA CONCEITUAL

- Azul - Informações chaves
- Roxo - Exemplos
- Azul claro - Pequena definição
- Laranja - Tema principal

INTRODUÇÃO À LINGUAGENS FORMAIS

Um dos primeiros relatos de Linguagens Formais se dão com Frege em sua obra *Notação Conceitual: Uma Linguagem Formal*, decalcada da aritmética e do Pensamento Puro, criando a lógica de predicados axiomática, que podia distinguir perfeitamente problemas de generalidade múltipla. Um dos propósitos declarados de Frege era isolar os princípios apenas lógicos de inferência, para que ocorra uma representação adequada da prova matemática, sem nenhum apelo à interpretação pessoal (uma vez que você estabelece o que foi dito ela pode ser interpretada de apenas uma forma).

De forma informal, temos que Linguagem é um instrumento de representação, conceitualização e comunicação através de símbolos escritos ou falados, símbolos estes que podem ser abrangentes ou específicos a depender da semântica, contexto social, etc, utilizando-se de regras predefinidas para juntar e dar significado a estes símbolos. Já as linguagens formais pode-se dizer que são mecanismos formais para representação e especificação de linguagens e problemas através de palavras. Elas podem ser representadas de maneira finita e precisa através de sistemas com sustentação matemática.

É algo fundamental à Teoria da Computação pois a própria computação pode ser entendida como uma linguagem, pois ela é utilizada como forma de conceitualizar o mundo, assim como a matemática, de forma que podemos manipular de maneira efetiva e raciocinar sobre essas representações abstratas.

Com isto, temos diversas aplicações na computação que são resultados diretos do estudo desta área, como análise sintática e léxica, fundamentais para construção de compiladores e as próprias linguagens de programação, análises de textos, etc.

CONCEITOS

Símbolo ou Caractere - Entidade básica, não definida formalmente / base para definições / Exemplo: letras ou dígitos

Alfabeto Σ - Conjunto finito de símbolos ou caracteres

- Conjunto infinito não é alfabeto

- Conjunto Vazio ϕ pode ser considerado um alfabeto

Palavra - Sequência finita de símbolos justapostos

Função de um número natural no conjunto de elementos que está incluído na sentença. ($w : n \rightarrow \Sigma$)

Palavra vazia - Cadeia sem símbolos

Chamamos de palavra vazia $\lambda : 0 \rightarrow \Sigma$

Sub Palavra - qualquer sequência de símbolos contíguos da palavra

Σ^* - Conjunto de todas as palavras possíveis sobre o alfabeto Σ . ($\Sigma^* \subseteq \mathbb{N} \times \Sigma$)

Comprimento ou tamanho de uma palavra $|w|$ - número de símbolos que compõem a palavra

Concatenação de palavras - operação binária sobre um conjunto de palavras que associa duas palavras pela justaposição que tem tamanho igual a soma das palavras concatenadas ($w_1.w_2 : |w_1| + |w_2| \rightarrow \Sigma$) tal que

$$w_1.w_2(i) = \{w_1(i) \text{ se } i < |w_1| \text{ ou } w_2(w_1 - i) \text{ caso contrário}\}$$

Linguagem L sobre um alfabeto Σ - qualquer conjunto contido em Σ^* ($L \subseteq \Sigma^*$).

Pode ser pensada como todas as possíveis palavras construtíveis nessa linguagem de acordo com as regras que a definem. Como por exemplo, o conjunto de palíndromos sobre Σ .

Considerando uma linguagem com infinitas palavras sobre um alfabeto é inviável representá-la completamente e verificar se cada uma dessas palavras pertence ao conjunto de palavras do alfabeto, para esse tipo de situação podemos utilizar de formalismos, que nada mais são que uma maneira de especificar de forma finita linguagens (eventualmente) finitas, para fazer tal verificação, sendo eles:

- Formalismos Reconhecedores - Estruturas ou representações utilizadas para reconhecer se um determinado elemento pertence a uma linguagem ou não. Ex: Máquina de Estados.
- Formalismos Geracionais - Gera exemplos que pertencem a determinada linguagem. Ex: Gramática.

Derivação - aplicação de uma regra de produção e a aplicação sucessiva de regras de derivação é chamado de fecho transitivo, que permite derivar palavras da linguagem.

Gramática - Conjunto finito de regras que quando aplicadas sucessivamente, geram palavras e o conjunto de palavras geradas por uma gramática define a Linguagem. É um formalismo axiomático de geração, uma gramática pode ser definida como uma tupla $G = \langle V, \Sigma, S, P \rangle$ com: um conjunto finito de variáveis $V \neq \emptyset$, um alfabeto Σ , um símbolo inicial $S \in V$ e um conjunto de regras de reescrita $P \subseteq (V \cup \Sigma)^* \times (V \cup \Sigma)^*$.

Linguagem gerada por uma Gramática - $L(G) = \{w \in E^* | S \Rightarrow_G^* w\}$

Máquina de Estados - Com ela podemos modularizar o comportamento de determinado sistema, a partir de um conjunto finito ou limitado de estados de processamentos. Uma máquina de estados pode ser definida como uma tupla ($M = \langle Q, \Sigma, q_0, \delta, F \rangle$) onde contém um conjunto não-vazio Q de estados da máquina, um alfabeto de entrada Σ , um

estado inicial da máquina $q_0 \in Q$, uma função de progressão da máquina $\delta \subseteq Q \times (\Sigma \cup \{\lambda\}) \times Q$ e um conjunto de estados objetivos da máquina $F \subseteq Q$.

- Na computação - usado para representar abstratamente o comportamento de determinado sistema.

- Uma linguagem pode ser dita como sendo todas as palavras que são aceitas por uma máquina de estados

Tipos de máquinas para representar programas -

Máquina de estados finitos (Q é finito) -

Quantidade finita de memória: sem alocação dinâmica de memória ou chamada de procedimentos.

Máquina de pilhas / filas ($q = \langle q_i, \alpha \rangle$) -

Manipulação limitada da memória: valores consumidos as vezes não podem ser resgatados.

Máquina não determinística (δ não é função) -

Incerteza do processamento.

Hierarquia de Chomsky - Noam Chomsky propôs uma hierarquia de gramáticas para descrever problemas, com a hierarquia podemos dar tipos a gramática dessa forma:

- De tipo 3 (regular), se toda produção é da forma $A \rightarrow w$ ou $A \rightarrow wB$ ($w \in \Sigma^*$) (adiciona uma ou nenhuma variável, cresce apenas em uma direção)
- De tipo 2 (livre de contexto), se toda produção é da forma $A \rightarrow \alpha$ ($\alpha \in (V \cup \Sigma)^*$) (não tem restrição de quantidade e posição de variável)
- De tipo 1 (sensível a contexto), quando se tem uma variável em um contexto específico, se toda produção é da forma $\alpha A \gamma \rightarrow \alpha \beta \gamma$ ($\alpha, \beta, \gamma \in (V \cup \Sigma)^*$) (caso essa variável for sucedida ou precedida, ela é reescrita como outra coisa)
- De tipo 0 (estrutura de frase), se toda produção é da forma $\alpha \rightarrow \beta$ ($\alpha, \beta \in (V \cup \Sigma)^*$) (sem restrição nenhuma)

Sentença - Sentença em linguística, também denominada “frase”, é um enunciado de sentido completo, a unidade mínima de comunicação. (Cunha e Cintra, 2008, p.133).

Semântica - Na linguística, trata-se do componente do sentido das palavras e da interpretação das sentenças e dos enunciados.

Sintaxe - Componente que determina as relações formais que interligam os constituintes da sentença, atribuindo-lhe uma estrutura

Morfologia - Estuda as formas das palavras em diferentes usos e construções. Trata das estruturas internas das palavras e dos seus constituintes significativos

DESCRITIVA

Alfabeto é o conjunto finito que todos diversos símbolos ou caracteres, estes que por sua vez são a entidade básica não definida formalmente que serve como base de nossas definições. Nota-se que um alfabeto pode ser um conjunto vazio, sem nenhum símbolo.

Já palavra é uma sequência finita de símbolos que existem no conjunto alfabeto, justapostos, ou seja, sem espaços entre si. Importante ressaltar a importância da palavra

vazia, que trata-se de uma **cadeia sem símbolos**. Outro conceito importante de palavra, é o de número de símbolos que compõem a palavra, chamado de comprimento ou tamanho de uma palavra. Há uma operação definida sobre o conjunto das palavras chamada concatenação, que é a operação binária que associa duas palavras pela justaposição. Seu **tamanho é igual a soma do tamanho das palavras concatenadas**. Outra definição fundamental é a do alfabeto estrela, definido como o conjunto de todas as palavras possíveis de se formar a partir de um alfabeto dado.

Linguagem sobre um alfabeto é qualquer conjunto contido no conjunto de todas as palavras. De maneira informal, pode ser pensada como **todas as possíveis palavras construídas nessa linguagem de acordo com as regras que a definem**.

Considerando uma linguagem com infinitas palavras sobre um alfabeto é inviável representá-la completamente e verificar se cada uma dessas palavras pertence ao conjunto de palavras do alfabeto, para esse tipo de situação podemos utilizar de formalismos, que nada mais são que uma maneira de especificar de forma finita linguagens (eventualmente) finitas, para fazer tal verificação, sendo eles: formalismos geracionais que gera exemplos que pertencem a determinada linguagem, e formalismos reconhecedores que são estruturas ou representações utilizadas para reconhecer se um determinado elemento pertence a uma linguagem ou não.

Numa **máquina de estados**, um dos formalismos reconhecedores, nós podemos descrever o comportamento de determinado sistema, a partir de um conjunto finito ou limitado de estados de processamentos. Uma linguagem pode ser dita como sendo todas as palavras que são aceitas por uma máquina de estados.

Ainda falando de máquinas, há tipos de máquina para representar tipos de programa: máquina de estados finitos, quantidade finita de memória: sem alocação dinâmica de memória ou chamada de procedimentos (Q é finito); manipulação limitada da memória: valores consumidos as vezes não podem ser resgatados (Máquina de pilhas / filas ($q = \langle q_i, \alpha \rangle$)); e máquina não determinística, quando δ não é função.

Como exemplo para um formalismo geracional, podemos fazer da **gramática**, que nada mais é que um conjunto finito de **regras que, quando aplicadas sucessivamente, geram palavras, e o conjunto de palavras geradas por uma gramática define a linguagem**.

E sobre gramáticas, temos um conceito fundamental, que se trata da **Hierarquia de Chomsky**. Que é uma categorização das gramáticas que as define de acordo com o seu poder expressivo e consequentemente o seu custo computacional, afinal, quanto maior o poder expressivo de uma linguagem, mais custoso é para gerar tais palavras. São quatro tipos, sendo eles, em ordem crescente de custo computacional: **tipo 3 ou gramáticas regulares**, que possuem regras que trocam/adicionam apenas uma ou nenhum variável, **tipo 2 ou livre de contexto**, caso possuam regras que não tenham restrições de quantidade de variável, **tipo 1 ou sensível ao contexto**, caso trabalhem com uso de variáveis mas caso elas estejam em subpalavras específicas e do **tipo 0 ou de estrutura de frase**, que são as mais custosas computacionalmente e contém regras que não tem restrição e podem substituir qualquer subpalavra por outra.

REFERÊNCIAS

XVII SEMANA DE HUMANIDADES, 17., 2009, Natal. **LINGUAGENS FORMAIS: ENFEITIÇAMENTO DA INTELIGÊNCIA HUMANA**. Natal: Departamento de Letras –

UFRN, 2009. 7 p.7 f. Disponível em: <https://cchla.ufrn.br/humanidades2009/>. Acesso em: 22 out. 2020.

FURTADO, Olinto José Varela. **LINGUAGENS FORMAIS E COMPILADORES**. 2014. Disponível em: https://www.ime.usp.br/~jef/tc_gramaticas.pdf. Acesso em: 22 out. 2020.

MENEZES, P. Blauth. **Linguagens Formais e Autômatos**. Rio de Janeiro: UFRJ, 2010. Color. Disponível em: <http://www.ic.uff.br/~ueverton/files/LF/>. Acesso em: 22 out. 2014.

CUNHA, Celso; CINTRA, Luís F. Lindley. **Complementos Verbais: Frase e a sua constituição**. In: Nova gramática do português contemporâneo. 5.ed. Rio de Janeiro: Lexikon, 2008, p. 133-135.