

فصل اول:

زبان و گرامر

سه مفهوم اساسی

در درس نظریه زبانها و ماشین ها، سه مفهوم زیر بررسی می شوند:

زبان	زبان، مجموعه ای از رشته ها روی یک الفبا می باشد.
گرامر	گرامر، ابزاری برای تولید زبان می باشد.
ماشین	ماشین (اتوماتا)، ابزاری برای پذیرش زبان می باشند. توسط ماشین می توان تشخیص داد که آیا یک رشته مربوط به زبان هست یا نه.

✓ الفباء (alphabet) : مجموعه‌ی متناهی از نمادها یا کاراکترها را یک الفباء گویند. الفباء را معمولاً با نماد Σ نشان می‌دهند. $\Sigma = \{a, b, c, \dots, x, y, z\}$ و $\Sigma = \{0, 1\}$ هر دو الفباء هستند ولی $X = \{1, 2, 3, \dots\}$ و $Y = \{\}$ هیچ‌کدام الفباء نیستند. چون X مجموعه‌ی متناهی نیست و Y هم یک مجموعه تهی است.

✓ رشته (string) بر روی Σ : یک دنباله متناهی از نمادهای Σ را یک رشته بر روی الفباء Σ گویند. توجه داشته باشید که نمادهای تشکیل دهنده یک رشته لزوماً متمایز نیستند. به عنوان مثال، "xyz" یک رشته بر روی $\Sigma = \{a, b, c, \dots, x, y, z\}$ و "10110" یک رشته بر روی $\Sigma = \{0, 1\}$ است.

✓ **طول رشته (length):** تعداد نمادهای تشکیل دهنده‌ی یک رشته را طول رشته گویند. طول رشته w را با $|w|$ نشان می‌دهند. مثلاً اگر $w = "a_1 a_2 \cdots a_n"$ آنگاه $|w| = n$. اگر $n = 0$ آنگاه رشته‌ی فاقد نماد بوده و آن را **رشته تهی (empty string)** گویند. رشته‌ی تهی را با نمادهای ε و λ نشان می‌دهند و $|\lambda| = 0$ به عنوان مثال اگر $w = abcd$ آنگاه $|w| = 4$ و اگر $w = 010$ آنگاه $|w| = 3$.

✓ **وارون یک رشته (Reverse):** وارون رشته‌ی w که آن را با نمادهای w^{-1} یا w^R نشان می‌دهند رشته‌ای است که از نوشتن نمادهای آن به ترتیب عکس بدست می‌آید. بنابراین اگر $w = a_1 a_2 \cdots a_n$ آنگاه $w^R = a_n a_{n-1} \cdots a_1$. به عنوان مثال اگر $w = abcd$ آنگاه $w^R = dcba$. توجه داشته باشید که $(w^{-1})^{-1} = w$ و $(uv)^{-1} = v^{-1}u^{-1}$.

✓. توان یک رشته. اگر W رشته‌ی مفروض باشد آنگاه توانهایی از W به صورت زیر تعریف می‌شود:

$$w^n = \underbrace{ww \cdots w}_{n \text{ times}} \quad \text{آنگاه } n \geq 1 \text{ و } w^0 = \lambda$$

✓. الحاق (concatenation): فرض کنید u و V دو رشته است. uV یعنی الحاق u و V و عبارت است از رشته‌ای که از الحاق نمادهای V به انتهای راست نمادهای u بدست می‌آید. مثلا اگر $u = a_1 a_2 \cdots a_m$ و $v = b_1 b_2 \cdots b_n$ آنگاه $uv = a_1 a_2 \cdots a_m b_1 b_2 \cdots b_n$. بدیهی است که $|uv| = |u| + |v|$.

در ضمن رشته تهی λ عضو همانی نسبت به عمل الحاق است یعنی به ازاء هر رشته‌ای مانند w بر روی Σ

$$\lambda w = w\lambda = w$$

مثال . $\Sigma = \{a, b\}$ و $w = ababaaa$ یک رشته بر روی Σ است. طول و وارون رشته را مشخص کنید.

جواب. $|w| = 7$ و $w^R = aaababa$

مثال ۱. با فرض $w = abba$ مطلوب است w^0 و w^2 .

جواب. $w^0 = \lambda$ و $w^2 = abbaabba = ab^2a^2b^2a$.

توجه: فرض کنید $a \in \Sigma$ و W یک رشته بر روی Σ است. تعداد تکرار نماد a در رشته W را معمولاً به صورت

$|w|_a$ یا $n_a(w)$ نشان می‌دهیم. مثلاً اگر $w = ababaaa$ آنگاه $|w|_a = 5$ و $|w|_b = 2$

تعریف توانهایی از الفباء مفروض Σ

• $\Sigma^0 = \{\lambda\}$

• Σ^k یعنی مجموعه کلیه رشته‌ها بطول $k (k \geq 1)$ بر روی الفباء Σ .

اگر Σ دارای n نماد باشد آنگاه تعداد رشته‌ها بطول k بر روی Σ برابر n^k خواهد بود.

مثال ۱. $\Sigma = \{0,1,2\}$ دارای ۳ نماد است. مجموعه کلیدی رشته‌ها به طول‌های 1 و 2 و 3 را با ذکر تعداد عناصر هر کدام نشان دهید.

جواب. Σ^1 یعنی مجموعه کلیدی رشته‌ها بطول 1 و دارای $3^1 = 3$ رشته: $\Sigma^1 = \{0,1,2\}$

Σ^2 یعنی مجموعه کلیدی رشته‌ها بطول 2 و دارای $3^2 = 9$ رشته:

$$\Sigma^2 = \{00,01,02,10,11,12,20,21,22\}$$

Σ^3 یعنی مجموعه کلیدی رشته‌ها بطول 3 و دارای $3^3 = 27$ رشته:

$$\Sigma^3 = \{000,001,002,010,011,012,020,021,022,100,101,102,110,111,112,120,121,122,200,201,202,210,211,212,220,221,222\}$$

- Σ^* (سیکما استار) یعنی مجموعه کلیه رشته‌های ممکن و متناهی از الفباء Σ که رشته‌ی تهی λ را نیز شامل می‌شود.

$$\Sigma^* = \Sigma^0 \cup \Sigma^1 \cup \Sigma^2 \cup \dots \cup \Sigma^n \cup \dots$$

- Σ^+ (سکما پلاس) یعنی مجموعه کلیه رشته‌های ممکن و متناهی و غیر تهی از الفباء Σ .

$$\Sigma^+ = \Sigma^* - \{\lambda\} = \Sigma^1 \cup \Sigma^2 \cup \dots \cup \Sigma^n \cup \dots$$

مثال . اگر $\Sigma = \{a, b\}$ الفباء مفروض باشد $\Sigma^0, \Sigma^1, \Sigma^2, \Sigma^3, \Sigma^*, \Sigma^+$ را به صورت مجموعه نشان دهید.

$$\Sigma^0 = \{\lambda\}, \quad \Sigma^1 = \{a, b\}, \quad \Sigma^2 = \{aa, ab, ba, bb\} \quad \text{جواب.}$$

$$\Sigma^3 = \{aaa, aab, aba, abb, baa, bab, bba, bbb\}$$

$$\Sigma^* = \{\lambda, a, b, aa, ab, ba, bb, aaa, aab, aba, baa, bab, bba, bbb, \dots\}$$

زبان‌های فرمال (زبان‌های صوری) (Formal Languages)

هر مجموعه‌ای از رشته‌های متناهی از عناصر Σ به عبارت دیگر هر زیرمجموعه‌ای از Σ^* را یک زبان گویند. منظور از زبان صوری زبانی است که کلمات تشکیل دهنده آن بدون در نظر گرفتن معنی آنها مورد توجه بوده و مجموعه‌ای است از رشته‌های متشکل از نمادهای مورد نظر که با بکاربردن مجموعه‌ای از قوانین و فرمول‌های ریاضی و قابل پردازش برای ماشین بدست می‌آیند.

مثال . اگر $\Sigma = \{a, b\}$ الفباء مفروض باشد آنگاه $L_1 = \{a, aa, aab\}$, $L_2 = \{\lambda, abba, baba, aa, ab\}$ و $L_3 = \{a^n b^n \mid n \geq 0\}$ چند نمونه از زبانهای مختلف بر روی $\Sigma = \{a, b\}$ است. توجه داشته باشید که رشته‌های λ , "ab" و "aabb" متعلق به زبان L_3 هستند ولی $abb \notin L_3$.

عملیات بر روی زبان‌ها (operations on languages)

فرض کنید U و V رشته‌هایی بر روی Σ هستند. به ازاء هر زبان L و L_1 و L_2 بر روی Σ و با استفاده از عملیات معمول بر مجموعه‌ها و سایر عملیات دیگر به شرح زیر می‌توانیم به زبان‌های جدید به صورت زیر دسترسی داشته باشیم:

$$L_1 \cup L_2 = \{u \mid u \in L_1 \text{ or } u \in L_2\} = L_1 + L_2 \quad \checkmark \text{ اجتماع:}$$

$$L_1 \cap L_2 = \{u \mid u \in L_1 \text{ and } u \in L_2\} \quad \checkmark \text{ اشتراک:}$$

$$L_1 - L_2 = \{u \mid u \in L_1 \text{ and } u \notin L_2\} \quad \checkmark \text{ تفاضل:}$$

$$\bar{L} = \Sigma^* - L \quad \checkmark \text{ متمم:}$$

$$L_1 L_2 = \{uv \mid u \in L_1 \text{ and } v \in L_2\} \quad \checkmark \text{ الحاق (عمل ضرب):}$$

$$L^R = \{u^R \mid u \in L\} \quad \checkmark \text{ وارون:}$$

$\checkmark$ توان: اگر L زبان مفروضی باشد آنگاه توان‌هایی از L به صورت زیر تعریف می‌شود:

$$L^0 = \{\lambda\}, \quad L^1 = L, \quad L^n = \underbrace{LL \cdots L}_{n \text{ times}} \quad (\text{الحاق } n \text{ بار زبان } L)$$

$$L^* = L^0 \cup L^1 \cup L^2 \cup \dots \cup L^n \cup \dots \quad \checkmark \text{ عمل } * \text{ (star-closure): (Kleen's Star)}$$

$$L^+ = L^1 \cup L^2 \cup \dots \cup L^n \cup \dots \quad \checkmark \text{ عمل } + \text{ (positive-closure):}$$

مثال ۱. اگر $L_1 = \{a, ab, aaaa\}$ و $L_2 = \{bb, ab\}$ آنگاه $L_1 \cup L_2$, $L_1 \cap L_2$ و $L_1 - L_2$ را بدست آورید.

$$L_1 \cup L_2 = \{a, ab, bb, aaaa\} \quad \text{جواب.}$$

$$L_1 \cap L_2 = \{ab\}$$

$$L_1 - L_2 = \{a, aaaa\}$$

مثال ۲. اگر $L = \{01\}$ آنگاه L^* و L^+ را بدست آورید.

$$L^* = \{01\}^* = \{\lambda, 01, 0101, 010101, \dots\} = \{\{01\}^n \mid n \geq 0\} \quad \text{جواب.}$$

$$L^+ = \{01\}^+ = \{01, 0101, 010101, \dots\} = \{\{01\}^n \mid n > 0\}$$

مثال ۱. اگر $L = \{ab, aab, baba\}$ آنگاه ورون L را تشکیل دهید.

$$L^R = \{ba, baa, abab\} \quad \text{جواب.}$$

مثال ۲. اگر $L_1 = \{01, 1001\}$ و $L_2 = \{11, 00, 1\}$ مطلوب است $L_1 L_2$

$$L_1 L_2 = \{01, 1001\} \times \{11, 00, 1\} = \{\underline{0111}, \underline{0100}, \underline{011}, \underline{100111}, \underline{100100}, \underline{10011}\}$$

مثال . اگر $L = \{a, ab\}$ آنگاه L^2 و L^3 را بدست آورید.

$$L^2 = \{a, ab\} \times \{a, ab\} = \{aa, aab, aba, abab\} \quad \text{جواب.}$$

$$\begin{aligned} L^3 &= L^2 L = \{aa, aab, aba, abab\} \times \{a, ab\} \\ &= \{aaa, aaab, aaba, aabab, abaa, abaab, ababa, ababab\} \end{aligned}$$

مثال: تعداد اعضای L_1L_2 را تعیین کنید. ($L_1 = \{10,1\}$, $L_2 = \{01,011,11\}$)

مثال: با فرض $L = \{a^n b^n \mid n \geq 0\}$ مطلوب است L^2 و L^R

تعیین و تشخیص زبان‌ها

تعیین و تشخیص زبان‌ها به طرق مختلف به شرح زیر امکان پذیر است:

الف. مشخص کردن یک زبان با لیست کردن یک به یک رشته‌های آن.

به عنوان مثال با فرض $\Sigma = \{a, b\}$, زبان‌های L_1, L_2 با لیست کردن کلمات آنها در زیر ارائه شده است

$$L_1 = \{a, aa, aaa, ab, ba, aba\}$$

$$L_2 = \{ \lambda, ab, aabb, aaabbb, aaaabbbb, \dots \}$$

ب. معرفی یک زبان با استفاده از خواص رشته‌های تشکیل دهنده آن.

اغلب می‌توان انواع زبان‌های صوری را به سادگی از طریق بیان خواص مشترک رشته‌های تشکیل دهنده آن‌ها به صورت یک مجموعه نشان داد. بدین وسیله نه تنها توضیح مناسب یک زبان بلکه تفهیم خواص آن به نحو مطلوبی امکان پذیر می‌گردد.

مثال . با فرض $\Sigma = \{a, b\}$ زبانهای L_1, L_2, L_3, L_4, L_5 را از روی خواص رشته‌های تشکیل دهنده آنها به شرح زیر مشخص کنید.

✓ L_1 شامل رشته‌هایی است که با نماد a آغاز و طول هر رشته حداکثر برابر ۳ است.

جواب.

$$L_1 = \{ax \mid x \in \{a, b\}^*, |ax| \leq 3\} = \{a, aa, ab, aaa, aab, aba, abb\}$$

✓ L_2 شامل رشته‌هایی است که وارون هر رشته با خود رشته برابر است (palindrome).

$$L_2 = \{w \in \Sigma^* \mid w = w^R\}$$

جواب.

پ. معرفی یک زبان با استفاده از گرامر آن

طبق تعریف، زبان L بر روی الفباء Σ مجموعه‌ای است از رشته‌های متناهی از عناصر Σ و زیر مجموعه‌ای است از Σ^* . ولی باید توجه داشت که در حالت کلی یک زبان L روی یک الفبا Σ کلیه‌ی رشته‌های ممکن از اعضاء Σ را شامل نبوده، بلکه تعدادی از رشته‌ها را شامل می‌شود. بنابراین به قواعد مشخصی نیاز داریم که توسط آن بتوان رشته‌های مورد نیاز را ایجاد و همچنین تعلق یا عدم تعلق یک رشته به یک زبان را تشخیص داد. یکی از روشهای کلی و مفید در تشریح زبان‌ها گرامرها هستند.

تعریف: گرامر عبارت است از مجموعه‌ای از قواعد متناهی که با دنبال کردن آن می‌توان جملات معتبر از یک زبان را ایجاد و تعلق یا عدم تعلق یک رشته موردنظر به آن زبان را تشخیص داد.

مشخصات یک گرامر

گرامر G که آن را با نماد $G = (V, T, S, P)$ نشان می‌دهیم دارای چهار مشخصه به شرح زیر است:

- V مجموعه متناهی از نمادها که هر عنصر آن را یک متغیر (Variable) یا یک غیرپایانه (Nonterminal) گویند. متغیرها را معمولا با حروف بزرگ مانند A و B و C و S و نظایر آن نشان می‌دهند.
- T مجموعه‌ی متناهی و غیرتهی از نمادها بنام الفباء. هر عنصر T را یک پایانه یا یک ترمینال (Terminal) گویند. پایانه‌ها را معمولا با حروف کوچک مانند a و b و c و بعضی نمادهای ویژه نشان می‌دهند. توجه داشته باشید که V و T دو مجموعه‌ی مجزا هستند به عبارت دیگر $V \cap T = \emptyset$.
- S عنصر مشخصی است از متغیرها ($S \in V$) که آنرا نماد شروع (Start symbol) گویند.
- P مجموعه‌ی متناهی است که هر عنصر آنرا یک تولید (Production) یا یک قانون جایگزینی گویند. قوانین تولید را اغلب به صورت $\alpha \rightarrow \beta$ نشان می‌دهند که در آن $\alpha \in (V \cup T)^+$ و $\beta \in (V \cup T)^*$. توجه داشته باشید که سمت چپ تولید یعنی α حداقل دارای یک نماد غیر ترمینال بوده و رشته‌ی β می‌تواند یک رشته تهی هم باشد که در این صورت $\alpha \rightarrow \lambda$ و آنرا یک محصول تهی گویند.

مثال: نحوه تولید رشته $aabbcc$ را با توجه به گرامر زیر نمایش دهید.

$$1. S \rightarrow aSBC \mid abc$$

$$2. cB \rightarrow Bc$$

$$3. bB \rightarrow bb$$

$$4. cC \rightarrow cc$$

حل: شماره قاعده استفاده شده در هر مرحله نمایش داده شده است:

$$S \xrightarrow{1} aSBC \xrightarrow{1} aabcBC \xrightarrow{2} aabBcC \xrightarrow{3} aabbcC \xrightarrow{4} aabbcc$$

مثال: گرامر سازنده زبان $L(G) = \{a^n b^{n+1} : n \geq 0\}$ را بنویسید.

$$S \rightarrow Ab$$

$$A \rightarrow aAb \mid \lambda$$