نظریه زبانها و اتوماتها (فصل دوم)

نظریه زبانها و اتوماتها به عنوان یک درس 3 واحدی در دوره ی کارشناسی علوم کامپیوتر تدریس می شود. منبع این درس کتاب نظریه زبانها و ماشینها نوشته ی توماس سودکمپ و ترجمه مهندس سید حجت الله جلیلی است.

در این پست خلاصه ای از فصل دوم این کتاب را می بینید.

مقدمه:
یک زبان یک مجموعه از رشته ها روی یک الفباست. این جامع ترین تعریفی است که هیچ محدودیتی برای شکل رشته های تشکیل دهنده زبان ارائه نشده است. ولی اغلب زبانهایی که ما می شناسیم مثل زبانهای طبیعی ( فارسی و انگلیسی) یا زبانهای کامپیوتری یا زبانهای ریاضی. از رشته های دلخواه روی یک الفبا تشکیل نمی شوند بلکه دارای ویژگیهای خاصی می باشند که این ویژگیها قالب نحوی زبان را مشخص می کند. ما می توانیم از تعاریف بازگشتی و عملیات مجموعه ای برای اجرای این محدودیتها استفاده کنیم.

رشته ها و زبانها:
یک رشته یک دنباله متناهی از یک الفبا میباشد. رشته ها که به صورت بازگشتی از الفبای خاصی به دست می آیند عناصر پایه ای برای تعریف زبانها هستند. در حقیقت رشته ها از عناصر الفبا که غیر قابل تجزیه هستند به دست میآیند و چون عناصر الفبا غیر قابل تجزیه هستند ما آنها را کاراکتر مینامیم. فرض کنید Σ یک الفبا باشد $\sum^*$ (مجموعه رشته های روی Σ) به صورت بازگشتی زیر تعریف میشود:

مرحله اول (پایه): $\lambda \in \Sigma$ ( علامت $\lambda$ نشانگر رشته ی تهی میباشد. یعنی فرض میکنیم که $\lambda$ یک رشته بر روی الفباست.)

مرحله دوم (گام بازگشت): اگر $W \in \sum^*$ و $a \in \sum$ آنگاه $wa \in \sum^*$

مرحله سوم (همبستگی): $W \in \sum^*$ است اگر آن بتواند با تکرار متنهای از مرحله گام بازگشت به دست آید.

برای یک الفبای غیر تهی $\sum$، $\sum^*$ شامل تعداد نامتناهی از عناصر است. اگر $\sum=\{a\}$ آنگاه $\sum^*$ شامل رشته های : $\lambda$ ، a ، aa ، aaa و ... خواهد بود. اگر $\sum$ شامل n عنصر باشد آنگاه $n^k$ رشته به طول k در $\sum^*$ وجود دارد.

مثال:
$\sum = \{a,b\}$ . تمام رشته ها به طول 3 مساوی خواهند بود با $2^3 = 8$
رشته ها: aaa aab aba abb baa bab bba bbb

اگر زبانی مثل انگلیسی را در نظر بگیریم همه رشته های موجود (جملات) که از کلمات انگلیسی (در اینجا کلمات انگلیسی الفبا هستند) میتوانند به دست آیند جزء زبان انگلیسی نیستند. مثلا جمله زیر:
Car face eat sandwich
بنابر این برای ایجاد رشته ها در زبان قوانینی وجود دارد. زبان های شامل رشته های متناهی روی یک الفبا هستند. بنابراین: یک زبان روی یک الفبا $\sum$ یک زیر مجموعه متناهی از مجموعه نامتناهی $\sum^*$ میباشد.

الحاق رشته ها:
فرض کنید $u,v \in \sum^*$ . الحاق u,v در حقیقت چسباندن عناصر این رشته ها به ترتیب از چپ به راست به هم میباشد. مثلا فرض می کنیم u=ak و v=bx الحاق uv مساوی خواهد بود با akbx .

مثال:
فرض کنید u=ab ، v=ca و w=bb :
Uv=abca　　uw=abbb　　(uv)w=abcabb

با دقت در مثال بالا به این نتیجه میرسیم که عمل الحاق دارای خاصیت شرکت پذیری میباشد. پس الحاق یک عمل دودویی شرکت پذیر است.

زیر رشته:
گوییم u یک زیر رشته از v است اگر رشته هایی مثل x,y وجود داشته باشند به طوری که v=xuy . توجه کنید اگر رشته x تهی باشد u یک پیشوند برای v و اگر رشته y تهی باشد u یک پسوند برای v است.

معکوس یک رشته رشته ایست که برعکس نشان داده میشود و آن را با نماء $^R$ نشان میدهند. مثلا اگر u=caab باشد آنگاه $u^R = baac$ .

مشخصات متناهی زبانها:
مثال:
میخواهیم زبانی ایجاد کنیم به نام L روی الفبای {a,b} که این خصوصیات را داشته باشد: 1- همه رشته ها با a شروع شوند. 2- همه رشته ها طول زوج داشته باشند.

مرحله اول (پایه): $aa, ab \in L$
مرحله دوم (گام بازگشت): اگر $u \in L$ باشد آنگاه uaa , uab , uba , ubb عضو L هستند.
مرحله سوم (همبستگی): یک رشته عضو L است اگر آن بتواند با تکرار متناهی از مرحله گام بازگشت از عنصر پایه به دست آید.

(فکر کنید):
زبان L شامل رشته هایی روی {a,b} است که قبل از هر b فورا یک a میآید. مراحل ایجاد زبان را بنویسید.

یک روش دیگر برای تولید زبان استفاده از عملیات مجموعه ای برای ایجاد مجموعه های پیچیده از رشته ها از مجموعه های ساده است.
مثلا الحاق زبانهای X , Y به صورت XY :
$$XY = \{uv \mid u \in X, v \in Y\}$$

مثال:
فرض کنید X={a,b,c} و Y={ab,cd} در اینصورت:
XY=aab,acd,bab,bcd,cab,ccd
$$X^0 = \{\lambda\}$$
$$X^1 = X = \{a,b,c\}$$

برای هر i، $X^i$ شامل رشته هایی به طول i در $\sum^*$ میباشد. این مشاهده مارا با عملگر مجموعه ای دیگری بنام kleen star آشنا میکند که مفهوم آن به این صورت است $X^*$ که شامل تمام رشته های با هر طولی از X ها روی $\sum^*$ میباشد یعنی:
$$X^* = \bigcup_{i=0}^{\infty} X^I \quad \text{و همچنین} \quad X^+ = \bigcup_{i=1}^{\infty} X^i$$

$X^*$ مجموعه تمام رشته های است که میتوانند از عناصر X ساخته شوند. $X^+$ مجموعه رشته های غیر تهی از X است.

$$X^+ = X X^*$$

مثال:
زبان $\{a,b\}^* \{bb\} \{a,b\}^*$ شامل تمامی رشته ها روی {a,b} است که دارای زیر رشته bb است. الحاق مجموعه bb مارا از وجود bb در هر رشته از L مطمئن میسازد.

مجموعه های $\{a,b\}^*$ نیز مشخص میکنند که هر تعداد از a,b با هر ترتیبی میتوانند قبل یا بعد از bb قرار گیرند.

یک مثال جالب (فکر کنید):
$\{aa,bb,ab,ba\}^*$ شامل تمامی رشته ها با طول زوج روی {a,b} است و $\{a,b\}^* - \{aa,bb,ab,ba\}^*$ شامل تمامی رشته ها با طول فرد میباشد.

عبارات و مجموعه های با قاعده:
یک مجموعه باقاعده است اگر بتواند با استفاده از عملیات اجتماع، الحاق و kleen star از مجموعه تهی، مجموعه شامل رشته تهی و مجموعه الفبا تولید شود. مجموعه های باقاعده خانواده بزرگ و مهمی از زبان ها، هم در تئوری زبانهای صوری و هم زبان های ماشین با حالات متناهی را تشکیل میدهند.

تعریف:
مرحله اول (پایه): $\phi$ ، $\lambda$ و {a} به ازای هر $a \in \Sigma$ مجموعه های باقاعده روی $\Sigma$ هستند.
مرحله دوم (گام بازگشت): فرض کنید X,Y مجموعه های باقاعده روی $\Sigma$ باشند. $X \cup Y, XY, X^*$ مجموعه های باقاعده روی $\Sigma$ هستند.
مرحله سوم (همبستگی): X یک مجموعه باقاعده روی $\Sigma$ است. اگر آن بتواند با تکرار متناهی از مرحله گام بازگشت از عناصر پایه به دست آید.

عبارات با قاعده از مجموعه های با قاعده به صورت زیر ساخته میشوند.
$$\{a,b\}^* = (a \cup b)^*, \{a\} = a, \{a\}\{b\} = ab$$

مثال:
میخواهیم عبارت باقاعده برای مجموعه رشته هایی که شامل زیررشته aa نمیباشند ایجاد کنیم. تمامی a ها حداقل بایستی به یک b ختم شوند. عبارت با قاعده به صورت زیر خواهد بود.
$$\begin{aligned} & b^*\{ab^+)^* \cup b^*(ab^+)^* a \\ &= b^*(ab^+)^*(\lambda \cup a) \\ &= b^*(abb^*)^*(\lambda \cup a) \\ &= (b \cup ab)^*(\lambda \cup a) \end{aligned}$$

پیوست:
1) برای هر رشته w و یک $i \geq 0$ ، $(w^R)^i = (w^i)^R$
2) رشته های aaa و bbababb متقارن هستند. رشته های متقارن را میتوان به صورت: $\{w \mid w = w^R)$ نشان داد.

http://cstacs.blogfa.com
1386/12/01
ساعت 5:44 بامداد
احد جعفرزاده