Theory of Computation 1
Sets and Regular Expressions

Frank Stephan

Department of Computer Science

Department of Mathematics

National University of Singapore

fstephan@comp.nus.edu.sg

Languages

Examples of Languages (Sets)

(a) Set of all natural numbers in binary notation: either $\mathbf{0}$ or $\mathbf{1}$ followed by arbitrarily many digits from $\mathbf{0}, \mathbf{1}$.

(b) Set of all possible computer programs in syntax of programming language C: Tools can translate a formal description of C syntax into a syntax checker.

(c) Set of all C programs which pass a compiler without error messages: Compilers check more than just syntactical correctness.

(d) Set of all C programs which do not have bugs: No computer program can solve this task completely.

(e) Set of all texts of books written in English and published between 1066 and 1492: Exhaustive list describes this set.

Theory of Computation

How does one describe above sets? How does one modify descriptions? Do descriptions allow membership checks?

Languages

Language = Set of Strings over an Alphabet.
Alphabet $\mathbf{\Sigma}$, for example $\mathbf{\Sigma} = \{0, 1, 2\}$. Always finite.

Finite languages
$\mathbf{L_1} = \emptyset$, no elements.
$\mathbf{L_2} = \{\varepsilon\}$, set consisting of empty string.
$\mathbf{L_3} = \{\mathbf{00}, \mathbf{01}, \mathbf{02}, \mathbf{10}, \mathbf{11}, \mathbf{12}, \mathbf{20}, \mathbf{21}, \mathbf{22}\}$, all elements of length $\mathbf{2}$.
$\mathbf{L_4} = \{\varepsilon, \mathbf{0}, \mathbf{00}, \mathbf{000}, \mathbf{0000}\}$, all strings of $\mathbf{0}$s up to length $\mathbf{4}$.
$\mathbf{L_5} = \{\mathbf{01}, \mathbf{001}, \mathbf{02}, \mathbf{002}\}$, all strings consisting of one or two $\mathbf{0}$s followed by a 1 or $\mathbf{2}$.

Operations with Languages

Union:

$\mathbf{L \cup H = \{u : u \in L \text{ or } u \in H\}}$;

$\mathbf{\{00, 01, 02\} \cup \{01, 11, 21\} = \{00, 01, 02, 11, 21\}}$;

$\mathbf{\{0, 00, 000\} \cup \{00, 000, 0000\} = \{0, 00, 000, 0000\}}$.

Intersection:

$\mathbf{L \cap H = \{u : u \in L \text{ and } u \in H\}}$;

$\mathbf{\{00, 01, 02\} \cap \{01, 11, 21\} = \{01\}}$;

$\mathbf{\{0, 00, 000\} \cap \{00, 000, 0000\} = \{00, 000\}}$.

Set Difference:

$\mathbf{L - H = \{u : u \in L \text{ and } u \notin H\}}$;

$\mathbf{\{00, 01, 02\} - \{01, 11, 21\} = \{00, 02\}}$.

Concatenation:

$\mathbf{000 \cdot 1122 = 0001122}$;

$\mathbf{L \cdot H = \{v \cdot w : v \in L \text{ and } w \in H\}}$;

$\mathbf{\{0, 00\} \cdot \{1, 2\} = \{01, 001, 02, 002\}}$.

Kleene Star and Plus

Definition

$\mathbf{L^* = \{\varepsilon\} \cup L \cup L \cdot L \cup L \cdot L \cdot L \cup \ldots}$
$\mathbf{= \{w_1 \cdot w_2 \cdot \ldots \cdot w_n : n \geq 0 \text{ and } w_1, w_2, \ldots, w_n \in L\}}$;
$\mathbf{L^+ = L \cup L \cdot L \cup L \cdot L \cdot L \cup \ldots}$
$\mathbf{= \{w_1 \cdot w_2 \cdot \ldots \cdot w_n : n > 0 \text{ and } w_1, w_2, \ldots, w_n \in L\}}$.

Examples

$\emptyset^* = \{\varepsilon\}$.
$\mathbf{\Sigma}^*$ is the set of all words over $\mathbf{\Sigma}$.
$\mathbf{\{0\}^* = \{\varepsilon, 0, 00, 000, 0000, \ldots\}}$.
$\mathbf{\{00, 01, 10, 11\}^*}$ are all binary words of even length.
$\varepsilon \in \mathbf{L}^+$ iff $\varepsilon \in \mathbf{L}$.

Notation

Often $\mathbf{w}^*$ in place of $\{\mathbf{w}\}^*$;
Often $\mathbf{w} \cdot \mathbf{L}$ in place of $\{\mathbf{w}\} \cdot \mathbf{L}$.

Regular Languages

Regular expressions are either finite sets listed by their elements or obtained from other regular expressions by forming the Kleene star, Kleene plus, union, intersection, set-difference or concatenation.

A language is regular iff it can be described by a regular expression.

Regular sets have many different regular expressions.

For example, $\{0, 00\} \cdot \{1, 2\}$ and $\{01, 001, 02, 002\}$ describe the same set. Also 0^* and $(00)^* \cup 0 \cdot (00)^*$ describe the same set.

Intersections and set difference are traditionally not used in regular expressions, as every regular set has an expression only using union, concatenation and Kleene star.

The complement of a language L is $\Sigma^* - L$.

Quiz

Which three sets are described by two of the following regular expressions?

1. $\{\mathbf{00}, \mathbf{000}\}^+$;
2. $\{\mathbf{000}, \mathbf{0000}\}^+$;
3. $\mathbf{00} \cdot \mathbf{0}^*$;
4. $\mathbf{000} \cdot \mathbf{0}^*$;
5. $\{\mathbf{000}, \mathbf{0000}\} \cup (\mathbf{000000} \cdot \mathbf{0}^*)$;
6. $\{\mathbf{00}, \mathbf{01}, \mathbf{02}, \mathbf{10}, \mathbf{11}, \mathbf{12}\}$;
7. $\mathbf{0}^*\mathbf{1}^*\mathbf{2}^*$;
8. $(\mathbf{0}^*\mathbf{1}^*\mathbf{2}^*)^*$;
9. $(\{\mathbf{0}, \mathbf{1}\} \cdot \{\mathbf{0}, \mathbf{1}, \mathbf{2}\}^*) \cap (\{\mathbf{0}, \mathbf{1}, \mathbf{2}\} \cdot \{\mathbf{0}, \mathbf{1}, \mathbf{2}\})$.

Exercises 1.6, 1.7 and 1.8

Exercise 1.6

Assume $\mathbf{A}$ has $\mathbf{3}$ and $\mathbf{B}$ has $\mathbf{2}$ elements. How many elements do the following sets have at least and at most; it depends on the actual choice which of the bounds is realised: $\mathbf{A} \cup \mathbf{B}$, $\mathbf{A} \cap \mathbf{B}$, $\mathbf{A} \cdot \mathbf{B}$, $\mathbf{A} - \mathbf{B}$, $\mathbf{A}^* \cap \mathbf{B}^*$.

Exercise 1.7

Let $\mathbf{A}, \mathbf{B}$ be finite sets and $|\mathbf{A}|$ be the number of elements of $\mathbf{A}$. Is the following formula correct:

$$|\mathbf{A} \cup \mathbf{B}| + |\mathbf{A} \cap \mathbf{B}| = |\mathbf{A}| + |\mathbf{B}|?$$

Prove your answer.

Exercise 1.8

Make a regular expression without intersection and set difference for $\mathbf{0}^*\mathbf{1}^*\mathbf{0}^*\mathbf{1}^* \cap (\mathbf{11})^*(\mathbf{00})^*(\mathbf{11})^*(\mathbf{00})^*$.

Tutorial

Register for a tutorial group
During the semester, you can make up to TEN marks in tutorials (full course = 100 marks).

(a) Write up to four exercises up in the Forum, at most one per week; these can be reserved after the lecture at the lecturer; two marks for each exercise, at most eight marks.

(b) Present in the tutorial group one exercise per week; you have in your own group priority on exercises you write up, however, you can also present other exercises not signed up by any member of your group; five presentations, one mark; eight presentations, two marks; one presentation per week is counted.

Tests

Midterm Tests are inside the class and take up a part but not all of the lectures on Tuesday 17 September 2019 and Tuesday 29 October 2019; each midterm test counts 20 marks. The midterm tests are in the second half of the lectures on those days.

The final examination counts 50 marks. It is 28 November 2019 at 17:00 hrs and the duration is 2 hours. Please consult NUS webpages for more information and doublecheck the information there.

Summary: Tutorial 10 marks, Mitderms 20 + 20 marks, Final 50 marks.
https://nusmods.com/modules/CS4232/theory-of-computation

Theorem of Lyndon and Schützenberger

Theorem

If two words $\mathbf{v}, \mathbf{w}$ satisfy $\mathbf{vw} = \mathbf{wv}$ then $\exists \mathbf{u}\, [\mathbf{v}, \mathbf{w} \in \mathbf{u}^*]$;
If all words $\mathbf{v}, \mathbf{w} \in \mathbf{L}$ satisfy $\mathbf{vw} = \mathbf{wv}$ then $\exists \mathbf{u}\, [\mathbf{L} \subseteq \mathbf{u}^*]$.

Proof

Case $\mathbf{v} = \varepsilon$ or $\mathbf{w} = \varepsilon$: $\mathbf{u} = \mathbf{vw}$.
Case $\mathbf{v} \neq \varepsilon$ and $\mathbf{w} \neq \varepsilon$ and $|\mathbf{v}| = |\mathbf{w}|$: $\mathbf{v} = \mathbf{w}$.
Case $\mathbf{v} \neq \varepsilon$ and $\mathbf{w} \neq \varepsilon$ and $|\mathbf{v}| < |\mathbf{w}|$: Let $\mathbf{k}$ be greatest common divisor of $|\mathbf{v}|, |\mathbf{w}|$ and $\mathbf{u}$ be the first $\mathbf{k}$ symbols of $\mathbf{v}$.
There are $\mathbf{i}, \mathbf{j}$ with $\mathbf{v} = \mathbf{u_1 u_2 \ldots u_i}$ and $\mathbf{w} = \mathbf{u_1 u_2 \ldots u_j}$ for words $\mathbf{u_1, u_2, \ldots, u_j}$ of length $\mathbf{k}$.
Note that $\mathbf{v^j w^i} = \mathbf{w^i v^j}$ and $|\mathbf{v^j}| = |\mathbf{w^i}|$, as both have the length $\mathbf{ijk}$. Thus $\mathbf{v^j} = \mathbf{w^i}$.
For each $\mathbf{u_h}$ there is a position of $\mathbf{u_h}$ in $\mathbf{w^i}$ where $\mathbf{u_1}$ is at the same position in $\mathbf{v^j}$. Thus $\mathbf{u_h} = \mathbf{u_1}$.
So $\mathbf{v}, \mathbf{w} \in \mathbf{u}^*$ for $\mathbf{u} = \mathbf{u_1}$.

Example, Second Part

Example: Let $\mathbf{v} = \mathbf{abcd}$ and $\mathbf{w} = \mathbf{abcdef}$ and $\mathbf{vw} = \mathbf{wv}$. Now $\mathbf{k} = \mathbf{2}$ (greatest common divisor of $\mathbf{4}, \mathbf{6}$).

$$\begin{aligned} \mathbf{v}^{\mathbf{3}} &= \mathbf{ab\,cd\,ab\,cd\,ab\,cd}; \\ \mathbf{w}^{\mathbf{2}} &= \mathbf{ab\,cd\,ef\;\;ab\,cd\,ef}. \end{aligned}$$

Now $\mathbf{ab} = \mathbf{ab}$ at $\mathbf{0}, \mathbf{1}$, $\mathbf{ab} = \mathbf{ef}$ at $\mathbf{4}, \mathbf{5}$, $\mathbf{ab} = \mathbf{cd}$ at $\mathbf{8}, \mathbf{9}$. So $\mathbf{ab} = \mathbf{cd} = \mathbf{ef}$ and $\mathbf{v}, \mathbf{w} \in (\mathbf{ab})^*$.

Second Part: Let $\mathbf{v}$ be shortest nonempty word of $\mathbf{L}$ and $\mathbf{u}$ be shortest word with $\mathbf{v} \in \mathbf{u}^*$.
Let $\mathbf{w} \in \mathbf{L}$ be arbitrary.
There is $\tilde{\mathbf{u}}$ with $\mathbf{v}, \mathbf{w} \in \tilde{\mathbf{u}}^*$.
Now $\tilde{\mathbf{u}}^{\mathbf{i}} = \mathbf{u}^{\mathbf{j}} = \mathbf{v}$ for some $\mathbf{i}, \mathbf{j}$.
Thus $\tilde{\mathbf{u}}, \mathbf{u} \in \hat{\mathbf{u}}^*$ as in Part 1 for some $\hat{\mathbf{u}}$ where $|\hat{\mathbf{u}}|$ is greatest common divisor of $|\mathbf{u}|$ and $|\tilde{\mathbf{u}}|$.
By choice of $\mathbf{u}$, $\hat{\mathbf{u}} = \mathbf{u}$ and $\mathbf{w} \in \mathbf{u}^*$.
So $\mathbf{L} \subseteq \mathbf{u}^*$.

Structural Induction

Theorem

Let $\mathbf{P}$ be a property of sets such that the following holds:

- Every finite set (including $\emptyset$) satisfies $\mathbf{P}$;
- If $\mathbf{L}, \mathbf{H}$ satisfy $\mathbf{P}$ so does $\mathbf{L} \cup \mathbf{H}$;
- If $\mathbf{L}, \mathbf{H}$ satisfy $\mathbf{P}$ so does $\mathbf{L} \cdot \mathbf{H}$;
- If $\mathbf{L}$ satisfies $\mathbf{P}$ so does $\mathbf{L}^*$.

Then all regular sets satisfy $\mathbf{P}$.

Proof.

Let $\mathbf{L}(\sigma)$ be the set generated by the regular expression σ. Here finite sets and the operations union, concatenation and Kleene star are permitted for regular expressions.

Now it is shown that there is no shortest regular expression σ such that $\mathbf{L}(\sigma)$ does not satisfy $\mathbf{P}$.

Shortest Expression

Assume that σ is a shortest expression not satisfying $\mathbf{P}$; if there are several shortest ones, σ is just any of these.

- If σ is a list of a finite set then $\mathbf{P}$ satisfies $\mathbf{L}(\sigma)$;
- If $\sigma = (\rho \cup \tau)$ then ρ, τ are shorter than σ and $\mathbf{L}(\rho), \mathbf{L}(\tau)$ satisfy $\mathbf{P}$ and so does $\mathbf{L}(\sigma) = \mathbf{L}(\rho) \cup \mathbf{L}(\tau)$;
- If $\sigma = (\rho \cdot \tau)$ then ρ, τ are shorter than σ and $\mathbf{L}(\rho), \mathbf{L}(\tau)$ satisfy $\mathbf{P}$ and so does $\mathbf{L}(\sigma) = \mathbf{L}(\rho) \cdot \mathbf{L}(\tau)$;
- If $\sigma = \tau^*$ then τ is shorter than σ and $\mathbf{L}(\tau)$ satisfies $\mathbf{P}$ and so does $\mathbf{L}(\sigma) = (\mathbf{L}(\tau))^*$.

So there is no case in which $\mathbf{L}(\sigma)$ would not satisfy $\mathbf{P}$, thus this σ does not exist and there is no regular expression σ for which $\mathbf{L}(\sigma)$ does not satisfy $\mathbf{P}$. All regular languages satisfy $\mathbf{P}$.

Strengthening the Theorem

Theorem

Let $\mathbf{P}$ be a property of sets such that the following holds:

- The empty set and the set $\{\varepsilon\}$ satisfy $\mathbf{P}$;
- For every letter $\mathbf{a}$, the set $\{\mathbf{a}\}$ satisfies $\mathbf{P}$;
- If $\mathbf{L}, \mathbf{H}$ satisfy $\mathbf{P}$ so does $\mathbf{L} \cup \mathbf{H}$;
- If $\mathbf{L}, \mathbf{H}$ satisfy $\mathbf{P}$ so does $\mathbf{L} \cdot \mathbf{H}$;
- If $\mathbf{L}$ satisfies $\mathbf{P}$ so does $\mathbf{L}^*$.

Then all regular sets satisfy $\mathbf{P}$.

This strengthening is just based on the fact that every finite set of strings can be formed using concatenation and union from the sets containing a single letter word, the set containing the empty word and the empty set.

Polynomial and Exponential Growth

Definition

A language $\mathbf{L}$ has polynomial growth iff there is a polynomial $\mathbf{p}$ such that for all $\mathbf{n}$ there are in $\mathbf{L}$ at most $\mathbf{p(n)}$ many words shorter than $\mathbf{n}$.

A language $\mathbf{L}$ has exponential growth iff there are constants $\mathbf{h, k}$ such that $\mathbf{L}$ contains for all $\mathbf{n}$ at least $\mathbf{2^n}$ words which are shorter than $\mathbf{h \cdot n + k}$.

Theorem

Every regular set has either polynomial or exponential growth.

This will be proven by structural induction.

Examples

Every finite set has polynomial growth, as one plus the number of elements is a polynomial which is an upper bound as required.
The set $\mathbf{0^*1^*}$ has polynomial growth as there are $\mathbf{n(n+1)/2}$ many words shorter than $\mathbf{n}$ in this set.
The set $\mathbf{0^* \cup 1^*}$ has polynomial growth as there are at most $\mathbf{2n}$ many words shorter than $\mathbf{n}$ in this set.

The set $\mathbf{\{00, 11\}^* \cdot \{222222\}}$ has exponential growth as it has for all $\mathbf{n}$ at least $\mathbf{2^n}$ words shorter than $\mathbf{7 + 2n}$.
The set $\mathbf{\{0000, 1111\}^*}$ has exponential growth as it has for all $\mathbf{n}$ at least $\mathbf{2^n}$ words shorter than $\mathbf{1 + 4n}$.

The set $\mathbf{\{00, 11\}^* \cdot \emptyset}$ is empty and has polynomial growth.

Rules for Growth

Finite sets have polynomial growth.

If $\mathbf{L}$ and $\mathbf{H}$ have polynomial growth then so do $\mathbf{L} \cup \mathbf{H}$ and $\mathbf{L} \cdot \mathbf{H}$.

If $\mathbf{L}$ or $\mathbf{H}$ have exponential growth then so does $\mathbf{L} \cup \mathbf{H}$.

The sets $\mathbf{L} \cdot \emptyset$ and $\emptyset \cdot \mathbf{L}$ have polynomial growth.

If $\mathbf{L}$ and $\mathbf{H}$ are not empty and at least one of them has exponential growth so does $\mathbf{L} \cdot \mathbf{H}$.

If $\mathbf{L}$ contains $\mathbf{v}, \mathbf{w}$ with $\mathbf{vw} \neq \mathbf{wv}$ then $\mathbf{L}^*$ has exponential growth else $\mathbf{L}^*$ has polynomial growth.

Let $\mathbf{P}(\mathbf{L})$ say that the language $\mathbf{L}$ has either polynomial or exponential growth. Then the rules imply that all finite sets satisfy $\mathbf{P}$ and that, whenever $\mathbf{L}, \mathbf{H}$ satisfy $\mathbf{P}$ so do $\mathbf{L} \cup \mathbf{H}$, $\mathbf{L} \cdot \mathbf{H}$ and $\mathbf{L}^*$. Thus all regular sets satisfy $\mathbf{P}$ by structural induction.

Quiz

Does $\mathbf{L \cap H}$ have exponential growth whenever $\mathbf{L}$ and $\mathbf{H}$ have exponential growth?

Does $\mathbf{\{0101, 010101\}^*}$ have exponential growth?

Does $\mathbf{\{000, 001, 011, 111\}^* \cdot \{0000, 1111\}}$ have exponential growth?

Does the (non-regular) set $\mathbf{\{w \in \{0,1\}^* : w}$ has at most $\mathbf{\log(|w|)}$ many $\mathbf{1s\}}$ have polynomial growth?

Does the set $\mathbf{\{w \in \{0,1\}^* : w}$ has at most $\mathbf{\log(|w|)}$ many $\mathbf{1s\}}$ have exponential growth?

Is there a polynomial $\mathbf{p}$ such that every set of polynomial growth has at most $\mathbf{p(n)}$ elements shorter than $\mathbf{n}$ for every $\mathbf{n}$?

Rules for Regular Expressions

(a) $\mathbf{L} \cup \mathbf{L} = \mathbf{L}$, $\mathbf{L} \cap \mathbf{L} = \mathbf{L}$, $(\mathbf{L}^*)^* = \mathbf{L}^*$, $(\mathbf{L}^+)^+ = \mathbf{L}^+$;

(b) $(\mathbf{L} \cup \mathbf{H})^* = (\mathbf{L}^* \cdot \mathbf{H}^*)^*$ and if $\varepsilon \in \mathbf{L} \cap \mathbf{H}$ then $(\mathbf{L} \cup \mathbf{H})^* = (\mathbf{L} \cdot \mathbf{H})^*$;

(c) $(\mathbf{L} \cup \{\varepsilon\})^* = \mathbf{L}^*$, $\emptyset^* = \{\varepsilon\}$ and $\{\varepsilon\}^* = \{\varepsilon\}$;

(d) $\mathbf{L}^+ = \mathbf{L} \cdot \mathbf{L}^* = \mathbf{L}^* \cdot \mathbf{L}$ and $\mathbf{L}^* = \mathbf{L}^+ \cup \{\varepsilon\}$;

(e) $(\mathbf{L} \cup \mathbf{H}) \cdot \mathbf{K} = (\mathbf{L} \cdot \mathbf{K}) \cup (\mathbf{H} \cdot \mathbf{K})$ and $\mathbf{K} \cdot (\mathbf{L} \cup \mathbf{H}) = (\mathbf{K} \cdot \mathbf{L}) \cup (\mathbf{K} \cdot \mathbf{H})$;

(f) $(\mathbf{L} \cup \mathbf{H}) \cap \mathbf{K} = (\mathbf{L} \cap \mathbf{K}) \cup (\mathbf{H} \cap \mathbf{K})$ and $(\mathbf{L} \cap \mathbf{H}) \cup \mathbf{K} = (\mathbf{L} \cup \mathbf{K}) \cap (\mathbf{H} \cup \mathbf{K})$;

(g) $(\mathbf{L} \cup \mathbf{H}) - \mathbf{K} = (\mathbf{L} - \mathbf{K}) \cup (\mathbf{H} - \mathbf{K})$ and $(\mathbf{L} \cap \mathbf{H}) - \mathbf{K} = (\mathbf{L} - \mathbf{K}) \cap (\mathbf{H} - \mathbf{K})$.

Inequality Rules

(a) $\mathbf{L} \cdot \mathbf{L}$ can be different from $\mathbf{L}$: $\{0\} \cdot \{0\} = \{00\}$;

(b) $(\mathbf{L} \cap \mathbf{H})^* \subseteq \mathbf{L}^* \cap \mathbf{H}^*$;
Properness: $\mathbf{L} = \{00\}$, $\mathbf{H} = \{000\}$, $(\mathbf{L} \cap \mathbf{H})^* = \{\varepsilon\}$,
$\mathbf{L}^* \cap \mathbf{H}^* = \{000000\}^*$;

(c) If $\{\varepsilon\} \cup (\mathbf{L} \cdot \mathbf{H}) = \mathbf{H}$ then $\mathbf{L}^* \subseteq \mathbf{H}$;
Properness: $\mathbf{L} = \{\varepsilon\}$, $\mathbf{H} = \{0\}^*$;

(d) If $\mathbf{L} \cup (\mathbf{L} \cdot \mathbf{H}) = \mathbf{H}$ then $\mathbf{L}^+ \subseteq \mathbf{H}$;
Properness: $\mathbf{L} = \{\varepsilon\}$, $\mathbf{H} = \{0\}^*$;

(e) $(\mathbf{L} \cap \mathbf{H}) \cdot \mathbf{K} \subseteq (\mathbf{L} \cdot \mathbf{K}) \cap (\mathbf{H} \cdot \mathbf{K})$;
Properness: $(\{0\} \cap \{00\}) \cdot \{0, 00\} = \emptyset \subset \{000\} =$
$(\{0\} \cdot \{0, 00\}) \cap (\{00\} \cdot \{0, 00\})$;

(f) $\mathbf{K} \cdot (\mathbf{L} \cap \mathbf{H}) \subseteq (\mathbf{K} \cdot \mathbf{L}) \cap (\mathbf{K} \cdot \mathbf{H})$;
Properness: $\{0, 00\} \cdot (\{0\} \cap \{00\}) = \emptyset \subset \{000\} =$
$(\{0, 00\} \cdot \{0\}) \cap (\{0, 00\} \cdot \{00\})$.

Characterising Kleene Star

Corollary 1.17. For any set $\mathbf{L}$, the following statements characterise $\mathbf{L}^*$ and $\mathbf{L}^+$:

(a) $\mathbf{L}^*$ is the smallest set $\mathbf{H}$ such that $\{\varepsilon\} \cup (\mathbf{L} \cdot \mathbf{H}) = \mathbf{H}$;

(b) $\mathbf{L}^*$ is the smallest set $\mathbf{H}$ such that $\{\varepsilon\} \cup (\mathbf{L} \cdot \mathbf{H}) \subseteq \mathbf{H}$;

(c) $\mathbf{L}^+$ is the smallest set $\mathbf{H}$ such that $\mathbf{L} \cup (\mathbf{L} \cdot \mathbf{H}) = \mathbf{H}$;

(d) $\mathbf{L}^+$ is the smallest set $\mathbf{H}$ such that $\mathbf{L} \cup (\mathbf{L} \cdot \mathbf{H}) \subseteq \mathbf{H}$.

In the above, one could also use $\mathbf{H} \cdot \mathbf{L}$ in place of $\mathbf{L} \cdot \mathbf{H}$.

Exercise 1.18

Which three of the following sets are not equal to any of the other sets:

(a) $\{01, 10, 11\}^*$;

(b) $((\{0, 1\} \cdot \{0, 1\}) - \{00\})^*$;

(c) $(\{01, 10\} \cdot \{01, 10, 11\} \cup \{01, 10, 11\} \cdot \{01, 10\})^*$;

(d) $(\{01, 10, 11\} \cdot \{01, 10, 11\})^* \cup \{01, 10, 11\} \cdot (\{01, 10, 11\} \cdot \{01, 10, 11\})^*$;

(e) $\{0, 1\}^* - \{0, 1\} \cdot \{00, 11\}^*$;

(f) $((\{01\}^* \cup \{10\})^* \cup \{11\})^*$;

(g) $(\{\varepsilon\} \cup (\{0\} \cdot \{0, 1\}^* \cap \{1\} \cdot \{0, 1\}^*))^*$.

Explain your answer.

Exercise 1.19

Make a regular expression which contains all those decimal natural numbers which start with $\mathbf{3}$ or $\mathbf{8}$ and have an even number of digits and end with $\mathbf{5}$ or $\mathbf{7}$.

Make a further regular expression which contains all odd ternary numbers without leading $\mathbf{0}$s; here a ternary number is a number using the digits $\mathbf{0, 1, 2}$ with $\mathbf{10}$ being three, $\mathbf{11}$ being four and $\mathbf{1212}$ being fifty. The set described should contain the ternary numbers $\mathbf{1, 10, 12, 21, 100, 102, 111, 120, 122, 201, \ldots}$ which are the numbers $\mathbf{1, 3, 5, 7, 9, 11, 13, 15, 17, 19, \ldots}$ in decimal.

Exercise 1.20

Let S be the smallest class of languages such that

- every language of the form u^* for a nonempty word u is in S;
- the union of two languages in S is again in S;
- the concatenation of two languages in S is again in S.

Prove by structural induction the following properties of S:

(a) Every language in S is infinite;

(b) Every language in S has polynomial growth.

Lay out all inductive steps explicitly without only citing results in this lecture.

Exercise 1.21

Let $\mathbf{L}$ satisfy the following statement: For all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbf{L}$, either $\mathbf{uv} = \mathbf{vu}$ or $\mathbf{uw} = \mathbf{wu}$ or $\mathbf{vw} = \mathbf{wv}$. Which of the following statements are true for all such $\mathbf{L}$:

- All $\mathbf{x}, \mathbf{y} \in \mathbf{L}$ satisfy $\mathbf{xy} = \mathbf{yx}$;
- All sufficiently long $\mathbf{x}, \mathbf{y} \in \mathbf{L}$ satisfy $\mathbf{xy} = \mathbf{yx}$;
- The language $\mathbf{L}$ has polynomial growth.

Give an answer to these questions and prove them.

Exercises 1.22-1.23

In the following, each digit, the symbol ε, the symbol $\cdot$, the symbol $\cup$, the comma and each set bracket and each normal bracket have length 1 and the length of the expression is the number of all the symbols in it (counting repetitions).

Exercise 1.22

Let $\mathbf{L}$ consist of all words which contain each of the letters $\mathbf{0}, \mathbf{1}, \mathbf{2}, \mathbf{3}$ exactly once. Make a regular expression generating $\mathbf{L}$ which has at most length $\mathbf{100}$.

Exercise 1.23

Make a regular expression for the set $\{\mathbf{w} \in \{\mathbf{0}\}^* : |\mathbf{w}| \leq \mathbf{9}\}$ which has at most length 26.

Exercises 1.24-1.26

Let V be the set of vowels, W be the set of consonants and S be the set of punctuation marks and T be the set of spacings (blancs and new lines and so on).

Exercise 1.24. Make a regular expression (using above sets) of all words which contain at least two vowels and before, after and between vowels is exactly one consonant.

Exercise 1.25. Make a regular expression of all sentences where each sentence consists of words containing one vowel and arbitrarily many consonants and between two words are spacings and after the last word is a punctuation mark.

Exercise 1.26. Make a regular expressions generating texts of sentences separted by spacings where sentences are as above with the only difference that words can have one or two vowels and up to four consonants.