1 Leon's (WIP) ITCS Exam Notes

Basically adapted from Chris Dalziel's notes :)

Finite Automata

Definition: Finite Automata

A finite automaton takes a string as input and replies "yes" or "no". If an automaton A replies "yes" on a string S we say that A "accepts" S.

Definition: Deterministic Finite Automata

A deterministic finite automaton (DFA) is a quintuple $(Q, \Sigma, q_0, \delta, F)$ where

- Q is a finite set of states
- Σ is an alphabet
- $q_0 \in Q$ is the initial state
- $\delta : Q \times \Sigma \to Q$ is the transition function
- $F \subseteq Q$ is the set of final states

A DFA accepts a string $w \in \Sigma^*$ iff $\delta^*(q_0, w) \in F$, where δ^* is δ applied successively for each symbol in w.
The language of a DFA A is the set of all strings accepted by a, $\mathcal{L} \subseteq \Sigma^*$ is the set of all strings accepted by A.
The transition function is a total function which gives exactly one next state for each input symbol, i.e. it is deterministic

Definition: Nondeterministic Finite Automata

Non-determinism would mean that δ can return more than one successor state, it instead returns a set of possible states - no states is an empty set. A NFA is a quintuple $(Q, \Sigma, q_0, \delta, F)$ where:

- Q is a finite set of states
- Σ is an alphabet
- $q_0 \in Q$ is the initial state
- $\delta : Q \times \Sigma \to \mathcal{P}(Q)$ is the transition function
- $F \subseteq Q$ is the set of final states

The only difference between the definition of a DFA and that of an NFA is that in an NFA δ returns an element from the powerset of Q, $\mathcal{P}(Q)$

Adding non-determinism doesn't change "expressivity". Given an NFA A there is an equivalent DFA D such that $\mathcal{L}(D) = \mathcal{L}(A)$ and vice versa.

Definition: ϵ-NFA

If we allow non-deterministic state chnages that don't consume any input symbols, we can label silent moves using ϵ - meaning the empty string We define the ϵ closure $E(q)$ of a state q as the set of all states reachable from q by silent moves. That is, $E(q)$ is the least set satisfying:

- $q \in E(q)$
- For any $s \in E(q)$ we also have $\delta(s, \epsilon) \subseteq E(q)$

DFA, NFA, ϵ-NFA are all equal in expressive power

Regular Languages

Definition: Regular Languages

Any language which can be accepted by a finite automaton is called a regular language.
Regular languages are also those recognised by Regular Expressions

Definition: Regular Language Closure Properties

For two languages L_1 and L_2, the following operations satisfy the closure property, i.e. for a member $x \in X$, and an operation ϕ we have that $\phi(x) \in \mathbb{R}$ for all x.

- **Union**: $L_1 \cup L_2$ is the language that includes all strings of L_1 and all strings of L_2.
- **Intersection**: $L_1 \cap L_2$ is the language that includes all strings of L_1 that are not in L_2, and vice versa
- **Sequential Composition**: $L_1 L_2$ is the language of strings that consist of strings in L_1 followed by a string in L_2.
- **Kleene closure**: L^* is the language of strings that consist wholly of zero or more strings in L.

$$L^* = \bigcup_{i \in \mathbb{N}} L^i$$

- **Complement**: $\bar{L}$ is the language of every string not in L.

Definition: Regular Expressions

Regular characterise the regular languages, just like finite automata do. The following table shows the syntax and semantics of a regex.

Syntax	Semantics	
a	$[\![a]\!] = \{a\}$	$(a \in \Sigma)$
$\emptyset$	$[\![\emptyset]\!] = \emptyset$	
ϵ	$[\![\epsilon]\!] = \{\epsilon\}$	
$R_1 \cup R_2$	$[\![R_1 \cup R_2]\!] = [\![R_1]\!] \cup [\![R_2]\!]$	
$R_1 \circ R_2$	$[\![R_1 \circ R_2]\!] = [\![R_1]\!][\![R_2]\!]$	
R^*	$[\![R^*]\!] = [\![R]\!]^*$	

Definition: Generalised NFAs

A **generalised NFA**, or GNFA is an NFA where:

- Transitions have **regular expressions** on them instead of symbols
- There is only one unique final state
- The transition relation if **full**, except that the initial state has no incoming transitions, and the final state has no outgoing transitions

Theorem 1: Pumping Lemma

If $L \subseteq \Sigma^*$ is regular, then there is a **pumping length** $p \in \mathbb{N}$ such that for any $w \in L$ where $|w| \geq p$, we may split w into three piexes $w = xyz$ satisfying three conditions:

- $xy^iz, \quad \forall i \in \mathbb{N}$
- $|y| > 0$
- $|xy| \leq p$

Note that if the pumping lemma fails then the language is not regular, but the inverse is not necessarily true.

Theorem 2: Myhill-Nerode Theorem

Let $L \subseteq \Sigma^*$ and $x, y \in \Sigma^*$. If there exists a suffix string z such that $xz \in L$, but $yz \notin L$ or vice versa, then x and y are **distinguishable** by L. If x and y are not distinguishable by L, then we say that $x \equiv_L y$ - this is an equivalence relation. A regular language satisfies the following

- The number of equivalence classes $\equiv_L$ is finite.
- The number of equivalence classes is equal to the number of states in the minimal DFA accepting L (not as important)

Therefore, to show a language is non-regular, show that it has infinite equivalence classes - that is, we find an infinite sequence $u_0 u_1 \dots$ of strings such that for any i, j where $i \neq j$, there is a string w_{ij} such that $u_i w_{ij} \in L$ but $u_j w_{ij} \notin L$ or vice-versa

Context-Free Languages

Definition: Context-free Languages

By adding recursion to regexs we can begin to recognise some non-regular languages. All regular languages are also context free.

Definition: Context-free Grammars

A language is context-free iff it is recognised by a Context-free Grammar (CFG), which is a 4-tuple (N, Σ, P, S) where:

- N is a finite set of variables or non-terminals
- Σ is a finite set of terminals
- $P \subseteq N \times (N \cup \Sigma)^*$ is a finite set of rules or productions
  - Typically productions are written $A \to aBc$
  - Productions with common heads can be combined, $A \to a$ and $A \to Aa$ can be combined into $A \to a \mid Aa$
- $S \in N$ is the starting variable

We use α, β, γ to refer to sequences of terminals
We make a derivation step $\alpha A \beta \Rightarrow_G \alpha\gamma\beta$ whenever $(A \to \gamma) \in P$;
The language of a CFG G is:

$$\mathcal{L}(G) = \{w \in \Sigma^* \mid S \Rightarrow_G^* w^*\}$$

Where $\Rightarrow_G^*$ is the reflexive, transitive, closure of $\Rightarrow_G$.
Context-free grammars are ambiguous. They are closed under union, concatenation, and kleene star, but not under intersection or complementation

Definition: Eliminating Ambiguity

We want to eliminate ambiguity in CFGs while still accepting all the same strings. This can be done for our language of regular expressions:

- First defining atomic expressions: $A \to (S)|\emptyset|\epsilon|a|b$
- Then ones which use Kleene Star: $K \to A|A^*$
- Then ones which may use left-associative composition: $C \to K|C \circ K$
- Finally expressions which use unions: $S \to C|S \cup C$

The order of operations here is therefore bottom to top; unions come before compositions, which come before Kleene etc

Definition: Push-down Automata

Push-down automata are to CFGs what finite automatas are to regular expressions. They are implementationally identical to ϵ-NFAs with the addition of a stack. The recursive element of CFGs is implemented using a standard last-in-first-out stack.

Transitions in a push-down automata take the form $x, y \to z$ which is read as "consume the input x, popping y off the stack, and push z onto the stack". We can allow actions that don't consume, pop, or push by setting variables to ϵ.

Definition: Formal Def. of PDAs

A **push-down automaton** is a 6-tuple $(Q, \Sigma, \Gamma, \delta, q_0, F)$ where Q, Σ, Γ are all finite sets. Γ is the stack alphabet, and δ now may take a stack symbol as input or return one as output:

$$\delta : Q \times \Sigma_\epsilon \times \Gamma_\epsilon \to \mathcal{P}(Q \times \Gamma_\epsilon)$$

All other compoments are as with ϵ-NFAs

A string w is accepted by a PDA if it ends in a final state, i.e. $\delta^*(q_0, w, \epsilon)$ gives a state q and a stack γ wuch that $q \in F$.

Theorem 3: CFG to PDA

A language is context-free if and only if it is recognised by a push-down automaton. The proof is left as an exercise to the reader.

Theorem 4: Pumping CFLs intro

Suppose a CFG has n non-terminals, and we have a parse tree of height $k > n$. Then the same non-terminal V must have appeared as its own descendant in the tree

- **Pumping down**: Cut the tree at the higher occurance of V and replace it with the subtree at the lower occurance of V
- **Pumping up**: Cut at the lower occurance and replace it with a fresh copy of the higher occurance

Theorem 5: Pumping Lemma for CFLs

If L is context-free then there exists a pumping length $p \in \mathbb{N}$ such that if $w \in L$ with $|w| \geq p$ then w may be split into **five** pieces $w = uvxyz$ such that

- $uv^i xy^i z \in L$ for all $i \in \mathbb{N}$
- $|vy| > 0$
- $|vxy| \leq p$

Definition: Chomsky Grammars

Context-free grammars are a special case of Chomsky Grammars. Chomsky grammars are similar to CFGs, except that the left-hand side of a production may be any string that includes at least one non-terminal. An example is shown below

$$\begin{aligned} S &\to abc \mid aAbc \\ Ab &\to bA \\ Ac &\to Bbcc \\ bB &\to Bb \\ aB &\to aaA \mid aa \end{aligned}$$

Such a grammar is called **context-sensitive**

Definition: The Chomsky Heirarchy

A grammar $G = (N, \Sigma, P, S)$ is of type:

0. (or **computably enumerable**) in the general case
1. (or **context sensitive**) if $|\alpha| \leq |\beta|$ for all productions $\alpha \to \beta$, except we also allow $S \to \epsilon$ if S foes not occur on the RHS of any rule
2. (or **context free**) if all productions are of the form $A \to \alpha$ (i.e. a CFG)
3. (or **right-linear/regular**) if all productions are of the form $A \to w$ or $A \to wB$, where $w \in \Sigma$ and $B \in N$

Algorithms for Languages

Theorem 6: Emptiness for Regular Languages

Can we write a program to determine if a given regular language is empty?
Given a finite-automaton this is an instance of graph reach-ability, so we can use a depth-first search.

Theorem 7: Emptiness for Context-free languages

Can we write a program to determine if a given context-free language is empty?
Given a CFG for our language, we can perform the following process:

1. Mark the terminals and ϵ as generating
2. Mark all non-terminals which have a production with only generating symbols in their right hand side as generating
3. Repeat until nothing new is marked
4. Check if S is marked as generating or not

Theorem 8: Equivalence of DFA

Is it possible to write a program to determine if two discrete finite automata are equivalent?

Given two DFA for L_1 and L_2, we can use our standard constructions to produce a DFA of the symmetric set difference:

$$(L_1 \cap \overline{L}_2) \cup (L_2 \cap \overline{L}_1)$$

Register Machines

Definition: Register Machines

A **register machine**, or RM, consists of:

- A fixed number m of registers $R_0 \ldots R_{m-1}$, which each holds a natural number
- A fixed program P which is a sequence of n instructions $I_0 \ldots I_{n-1}$

Each instruction is one of the following:

- `INC(i)`: which increments the register R_i by one
- `DECJZ(i, j)`: which decrements register R_i unless $R_i = 0$ in which case it jumps to instruction I_j

RMs can compute anything any other computer can

Definition: Pairing Functions for RMs

A **pairing function** is an injective function $\mathbb{N} \times \mathbb{N} \to \mathbb{N}$. An example is $f(x, y) = 2^x 3^y$.
We write $\langle x, y \rangle_2$ for $f(x, y)$. If $z = \langle x, y \rangle_2$, let $z_0 = x$ and $z_1 = y$.
This lets us encode multiple values into a single value, and a 2-tuple pairing function is enough to cram an arbitrary sequence of natural numbers into one $\mathbb{N}^* \to \mathbb{N}$

Theorem 9: Church-Turing Thesis

The Church-Turing thesis states that any problem is computable by any model of computation iff it is computable by a **Turing machine**.
For our purposes this matters for RMs, TMs, and λ-calculus.
Other examples are combinator calculus, general recursive functions, pointer machines, counter machines, cellular automata, queue automata, enzyme-based DNA computers, Minecraft, Magic the Gathering, and others.

blank filler text: Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Donec odio elit, dictum in, hendrerit sit amet, egestas sed, leo. Praesent feugiat sapien aliquet odio. Integer vitae justo. Aliquam vestibulum fringilla lorem. Sed neque lectus, consectetuer at, consectetuer sed, eleifend ac, lectus. Nulla facilisi. Pellentesque eget lectus. Proin eu metus. Sed porttitor. In hac habitasse platea dictumst. Suspendisse eu lectus. Ut mi mi, lacinia sit amet, placerat et, mollis vitae, dui. Sed ante tellus, tristique ut, iaculis eu, malesuada ac, dui. Mauris nibh leo, facilisis non, adipiscing quis, ultrices a, dui.

Morbi luctus, wisi viverra faucibus pretium, nibh est placerat odio, nec commodo wisi enim eget quam. Quisque libero justo, consectetuer a, feugiat vitae, porttitor eu, libero. Suspendisse sed mauris vitae elit sollicitudin malesuada. Maecenas ultricies eros sit amet ante. Ut venenatis velit. Maecenas sed mi eget dui varius euismod. Phasellus aliquet volutpat odio. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Pellentesque sit amet pede ac sem eleifend consectetuer. Nullam elementum, urna vel imperdiet sodales, elit ipsum pharetra ligula, ac pretium ante justo a nulla. Curabitur tristique arcu eu metus. Vestibulum lectus. Proin mauris. Proin eu nunc eu urna hendrerit faucibus. Aliquam auctor, pede consequat laoreet varius, eros tellus scelerisque quam, pellentesque hendrerit ipsum dolor sed augue. Nulla nec lacus.

Suspendisse vitae elit. Aliquam arcu neque, ornare in, ullamcorper quis, commodo eu, libero. Fusce sagittis erat at erat tristique mollis. Maecenas sapien libero, molestie et, lobortis in, sodales eget, dui. Morbi ultrices rutrum lorem. Nam elementum ullamcorper leo. Morbi dui. Aliquam sagittis. Nunc placerat. Pellentesque tristique sodales est. Maecenas imperdiet lacinia velit. Cras non urna. Morbi eros pede, suscipit ac, varius vel, egestas non, eros. Praesent malesuada, diam id pretium elementum, eros sem dictum tortor, vel consectetuer odio sem sed wisi.

Sed feugiat. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Ut pellentesque augue sed urna. Vestibulum diam eros, fringilla et, consectetuer eu, nonummy id, sapien. Nullam at lectus. In sagittis ultrices mauris. Curabitur malesuada erat sit amet massa. Fusce blandit. Aliquam erat volutpat. Aliquam euismod. Aenean vel lectus. Nunc imperdiet justo nec dolor.

Etiam euismod. Fusce facilisis lacinia dui. Suspendisse potenti. In mi erat, cursus id, nonummy sed, ullamcorper eget, sapien. Praesent pretium, magna in eleifend egestas, pede pede pretium lorem, quis consectetuer tortor sapien facilisis magna. Mauris quis magna varius nulla scelerisque imperdiet. Aliquam non quam. Aliquam porttitor quam a lacus. Praesent vel arcu ut tortor cursus volutpat. In vitae pede quis diam bibendum placerat. Fusce elementum convallis neque. Sed dolor orci, scelerisque ac, dapibus nec, ultricies ut, mi. Duis nec dui quis leo sagittis commodo.

Definition: Name

Example : Regular Expressions

At least one 0:

$$(0 \cup 1)^* 0 (0 \cup 1) *$$

At least one 1 and at least one 0:

$$((0 \cup 1)^* 01 (0 \cup 1)^*) \cup ((0 \cup 1)^* 10 (0 \cup 1)^*)$$