Automata & Formal Languages

Monday 9 October 2023

Lecture II

Today

Summer Research Festival

2pm – 6pm

MR 2, 3, 4, 5, 9, & 15

MR3

5:20pm

A motivated theorem prover

5:40pm

How to avoid arithmetic progressions

Brief reminder of Lecture I:

Pages 9 & 10:

NOTATION & REMINDER

(1) $\mathbb{N} = \{0, 1, 2, \dots\}$ 0 is included in $\mathbb{N}$

$n = \{0, \dots, n-1\}$

(2) ~~For~~ any set X, we write

X^n for the sequence/tuples/strings of length n of elts of X

(3) In particular, X^0 consists only of the empty string.
We call ~~this~~ ε (EPSILON).

(4) ~~Note that~~ if $\alpha \in X^n$, α is a function with $\mathrm{dom}(\alpha) = n$
$\mathrm{ran}(\alpha) \subseteq X$.

(5) If $\alpha \in X^n$ and $k \leq n$, then
$\alpha \upharpoonright k \in X^k$
is the unique initial segment of α of length k.

(6) $|\alpha| := \mathrm{dom}(\alpha)$. length of α.

(7) $X^* := \bigcup_{n \in \mathbb{N}} X^n$ The set of finite X-sequences

(8) Concatenation:
$\alpha \in X^n$, $\beta \in X^m$, we can define $\alpha\beta \in X^{n+m}$ by

$$\alpha\beta(k) := \begin{cases} \alpha(k) & k < n \\ \beta(\ell) & k = n+\ell \text{ with } \ell < m \end{cases}$$

(9) Slightly incorrectly, I'll write x for the length one seq. with value x.

This means that αx is just

$\underbrace{a_0 \dots a_n}_{=\alpha} x$.

Notation

X^n : X-strings of length n

ε : empty string

$|\alpha|$: length of α

X^* : finite X-strings

$\alpha\beta$: concatenation of α and β.

NOTATION & REMINDERS (continued).

(10) Define recursively

$$\alpha^0 := \varepsilon$$

$$\alpha^{n+1} := \alpha^n \alpha$$

(11) If $\alpha = x$, then x^n is the length n sequence consisting of x's.

(12) $X^+ := X^* \setminus \{\varepsilon\}$

set of non-empty finite X-strings

(13) X is infinite if there is an inj. $\mathbb{N} \longrightarrow X$

X is countable if there is a surj. $\mathbb{N} \longrightarrow X$ or $X = \emptyset$

X is finite if not infinite

X is uncountable if not countable

(14) X, Y ctble $\Longrightarrow$ $X \times Y$ is countable

[Remember Cantor zigzag function

$$z : \mathbb{N} \times \mathbb{N} \longrightarrow \mathbb{N} \quad \text{bij.}$$

If $\pi_X : \mathbb{N} \longrightarrow X$, $\pi_Y : \mathbb{N} \longrightarrow Y$ surjections. Fix n, find i, j s.t. $z(i,j) = n$

$$n \longmapsto (\pi_X(i), \pi_Y(j)).\,]$$

(15) X is countable

$\Longrightarrow X^*$ is ctble

[Since X^{n+1} is in bij. with $X^n \times X$, it follows by ind. from (14) that all sets X^n are ctble.

But $X^* = \bigcup_{n \in \mathbb{N}} X^n$, so ctble as a ctble union of ctble sets. [N&S].]

(16) If $X \neq \emptyset$, then X^* is infinite.

[If $x \in X$, then $n \longmapsto x^n$ is an inj. from $\mathbb{N}$ to X^*.]

Sequence of length n with constant value x.

(17) Cantor's Thm.

If X is infinite, then $\mathcal{P}(X)$ is uncountable.

$\mathcal{P}(X)$ = power set of X = set of all subsets of X

[Very famous proof: DIAGONALISATION

Start with any function $f: \mathbb{N} \longrightarrow \mathcal{P}(X)$ and define $D \subseteq X$ s.t. $D \notin \operatorname{ran}(f)$.]

check typed lecture notes for the proof.

Georg Cantor

Cantor, c.1910

Born	Georg Ferdinand Ludwig Philipp Cantor 3 March 1845 Saint Petersburg, Russian Empire
Died	6 January 1918 (aged 72) Halle, Province of Saxony, German Empire
Nationality	German

(18) If X is countable, then so is the set of finite subsets of X, say $\mathrm{Fin}(X)$.

[Know by (15) that X^* is countable, so only need a surj. from X^* onto $\mathrm{Fin}(X)$.

If $\alpha \in X^*$, let $\pi(\alpha) := \operatorname{ran}(\alpha)$.

That is a surjection.]

Some subtlety: to show "surjection", need for arbitrary $F \in \mathrm{Fin}(X)$ some α s.t. $\pi(\alpha) = F$. How do you define α?

[See typed lecture notes.]

§ 1.2 Formal Languages & Rewrite Systems

Ω finite set of symbols

We call any $L \subseteq \Omega^*$ a (formal) Ω-language

Compare to natural languages:

$$\Omega = \{a, b, c, d, \dots, z\}$$

then words become elements of Ω^* and a dictionary is a subset of Ω^*, thus a language.

Clearly, that's finite.

[There is little mathematical theory for finite languages. So ?]

FORMAL LANGUAGES & REWRITE SYSTEMS

Chomsky:

A fundamental principle of language is

LINGUISTIC RECURSION

GENERATION

Languages are best thought of as infinitely generated, even though only a finite fragment of these is practically relevant.

Noam Chomsky

Chomsky in 2017

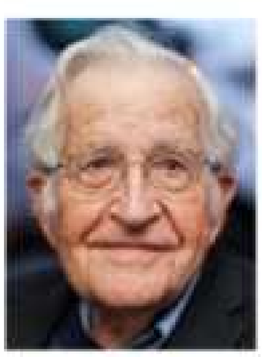
Born	Avram Noam Chomsky December 7, 1928 (age 94) Philadelphia, Pennsylvania, U.S.
Spouses	Carol Schatz (m. 1949; died 2008) Valeria Wasserman (m. 2014)
Children	3, including Aviva
Parent	William Chomsky (father)

E.g., the process of taking subordinate clauses is a *productive* feature of language. In principle, no matter how complex a sentence is, one can increase its complexity by prefixing it with "X observes that". So, we form an infinite sequence of grammatical sentences

B likes A.
C believes that B likes A.
D reports that C believes that B likes A.
E observes that D reports that C believes that B likes A.
etc.

Chomsky: The most salient feature of language is its recursive/productive nature which necessarily implies that languages should be conceived of as recursively generated & thus infinite.

GRAMMATICAL



COLOURLESS GREEN IDEAS SLEEP FURIOUSLY

FURIOUSLY SLEEP IDEAS GREEN COLOURLESS

UNGRAMMATICAL

Both sentences are nonsensical and have no meaning, but we can still clearly say that the first one is grammatical and the second one not.

If Ω is our set of symbols, we call elements of

$$\Omega^+ \times \Omega^*$$

REWRITE RULES
PRODUCTION RULES

We write
$\alpha \longrightarrow \beta$
for the rule
(α, β).

INFORMAL INTERPRETATION:

"Whenever a string contains α as a substring, we can replace it with β."

<u>Definition</u> A pair $R = (\Omega, P)$ is called a REWRITE SYSTEM if P is a finite set of rewrite rules.

<u>Prop.</u> For any fixed Ω, there are countably many rewrite systems (Ω, P).

<u>Proof.</u> By (15), Ω^* and Ω^+ are ctbl.
So by (14), $\Omega^+ \times \Omega^*$ is ctbl.
There is 1-1 correspondence between rewrite systems and $\mathrm{Fin}(\Omega^+ \times \Omega^*)$ which is ctble by (18). q.e.d.

<u>NOTATION</u> If $R=(\Omega,P)$ is a rewrite system, $\sigma,\tau\in\Omega^*$, then we write

R rewrites σ in one step as τ

$$\sigma \xrightarrow{R}_1 \tau$$

$$:\Longleftrightarrow$$

$\exists \alpha,\beta,\gamma,\delta\in\Omega^*$ s.t.

$$\sigma = \alpha\beta\gamma,$$
$$\tau = \alpha\delta\gamma,$$
$$\beta\longrightarrow\delta\in P.$$

We let the relation $\xrightarrow{R}$ be the transitive and reflexive closure of $\xrightarrow{R}_1$.

This means $\sigma\xrightarrow{R}\tau$ iff

an R-derivation of τ from σ

(1) $\sigma=\tau$

(2) there is a sequence of strings $\sigma_0,\dots,\sigma_n$ s.t. $\sigma=\sigma_0$, $\tau=\sigma_n$ and $\sigma_k\xrightarrow{R}_1\sigma_{k+1}$.

Note that a derivation of length n is a sequence of length $n+1$.

This is called derivation of length n.

R derives τ from σ.
R rewrites σ as τ
R produces τ from σ

$\mathcal{D}(R,\alpha) := \{\beta;\alpha\xrightarrow{R}\beta\}$

GRAMMARS

Def. $G = (\Sigma, V, P, S)$ is called a (formal) grammar over Σ

if

1. $\Sigma \cap V = \emptyset$
2. $\Omega := \Sigma \cup V$ is finite nonempty set of symbols

terminal symbols or letters

nonterminal symbols or variables

Σ is called the alphabet

3. (Ω, P) is a rewrite system
4. $S \in V$ called the start symbol

We call $\mathbb{W} := \Sigma^*$ the set of WORDS.

The definitions for rewrite systems still make sense:

$$\alpha \xrightarrow{G}_1 \beta, \quad \mathcal{D}(G, \alpha)$$

$$\alpha \xrightarrow{G} \beta$$

$$\mathcal{L}(G) := \mathbb{W} \cap \mathcal{D}(G, S)$$

LANGUAGE generated by G.

Example $G_0 = (\Sigma, V, P_0, S)$

$$\Sigma := \{a\}$$
$$V := \{S\}$$
$$P_0 := \{S \longrightarrow aaS, S \longrightarrow a\}$$

Claim $\mathcal{L}(G_0) = \{a^{2n+1} ; n \in \mathbb{N}\}$

Easy to see that $\supseteq$;
to see $\subseteq$, observe that every derivable string has odd length.

[Details will be repeated at the beginning of Lecture II.]