Phases of a compiler

Figure 1.6, page 5 of text
Process of building lexical analyzer

5) The minimal DFA is our lexical analyzer.

Diagram:

- Language
- Regex
- NFA
- DFA
- DFA (DFA)
  - Character stream
  - Token stream
  - Lexical analyzer
focus last time

NFA to DFA conversion
focus today
DFA minimization
NFA for \((a|b)^*abb\)
DFA for \((a|b)^*abb\)
Minimization Algorithm
**NFA -> DFA algorithm**

(set of states construction - page 153 of text)

**INPUT:** An DFA $D = (S, \Sigma, \delta, s_0, F)$

**OUTPUT:** A DFA $D' = (S', \Sigma, \delta', s'_0, F')$ such that

- $S'$ is as small as possible, and
- $L(D) = L(D')$

**ALGORITHM:**

1. Let $\pi = \{ F, S-F \}$
2. Let $\pi' = \pi$. For every group $G$ of $\pi$:
   - partition $G$ into subgroups such that two states $s$ and $t$ are in the same subgroup iff for all input symbols $a$, states $s$ and $t$ have transitions on $a$ to states in the same group of $\pi$
   - Replace $G$ in $\pi'$ by the set of all subgroups formed
3. if $\pi' = \pi$ let $\pi'' = \pi$, otherwise set $\pi = \pi'$ and repeat 2.
4. Choose one state in each group of $\pi''$ as a representative for that group.
   - a) The start state of $D'$ is the representative of the group containing the start state of $D$
   - b) The accepting states of $D'$ are the representatives of those groups that contain an accepting state of $D$
   - c) Adjust transitions from representatives to representatives.
DFA

D = (S, Σ, s₀, δ, F)

S = {A, B, C, D, E}
Σ = {a, b}
s₀ = A
δ = {(A,a)→B, (A,b)→C,
(B,a)→B, (B,b)→D,
(C,a)→B, (C,b)→C,
(D,a)→B, (D,b)→E,
(E,a)→B, (E,b)→C
}
F = {E}
Finding the minimal set of distinct sets of states

\[ \pi_0 = \{ F, S-F \} = \{ \{ E \}, \{ A, B, C, D \} \} \]

Pick a non-singleton set \( X = \{ A, B, C, D \} \) from \( \pi_0 \) and check behavior of states on all transitions on symbols in \( \Sigma \) (are they to states in \( X \) or to other groups in the partition?)

\[
\begin{align*}
(A,a) & \rightarrow B, \ (B,a) \rightarrow B, \ (C,a) \rightarrow B, \ (D,a) \rightarrow B \\
(A,b) & \rightarrow C, \ (B,b) \rightarrow D, \ (C,b) \rightarrow C, \ (D,b) \rightarrow E
\end{align*}
\]

\( D \) behaves differently, so put it in its own partition.
Finding the minimal set of distinct sets of states

$$\pi_1 = \{ \{E\}, \{A, B, C\}, \{D\} \}$$

Pick a non-singleton set $$X = \{A, B, C\}$$ from $$\pi_1$$ and check behavior of states on all transitions on symbols in $$\Sigma$$ (are they to states in $$X$$ or to other groups in the partition?)

$$(A, a) \rightarrow B, \ (B, a) \rightarrow B, \ (C, a) \rightarrow B$$

$$(A, b) \rightarrow C, \ (B, b) \rightarrow D, \ (C, b) \rightarrow C$$

B behaves differently, so put it in its own partition.
Finding the minimal set of distinct sets of states

\( \pi_2 = \{ \{E\}, \{A, C\}, \{B\}, \{D\} \} \)

Pick a non-singleton set \( X = \{A, C\} \) from \( \pi_2 \) and check behavior of states on all transitions on symbols in \( \Sigma \) (are they to states in \( X \) or to other groups in the partition?)

\( (A,a)\to B, (C,a)\to B \)
\( (A,b)\to C, (C,b)\to C \)

A and C both transition outside the group on symbol a, to the same group (the one containing B). Therefore A and C are indistinguishable in their behaviors, so do not split this group.
Finding the minimal set of distinct sets of states

$$\pi_3 = \{ \{E\}, \{A, C\}, \{B\}, \{D\} \} = \pi_2$$

We have reached a fixed point! STOP
Pick a representative from each group

\[ \pi_{\text{final}} = \{ \{E\}, \{A, C\}, \{B\}, \{D\} \} \]
MINIMAL DFA

\[ D' = ( S', \Sigma, s'_0, \delta', F') \]

\[ S' = \{B, C, D, E\} \rightarrow \text{the representatives} \]
\[ \Sigma = \{a, b\} \rightarrow \text{no change} \]
\[ s'_0 = C \rightarrow \text{the representative of the group that contained D's starting state, A} \]
\[ \delta = (\text{on next slide}) \]
\[ F = \{E\} \rightarrow \text{the representatives of all the groups that contained any of D's final states (which, in this case, was just \{E\})} \]
The new transition function $\delta'$

- For each state $s \in S'$, consider its transitions in $D$, on each $a \in \Sigma$.

- If $\delta(s,a) = t$, then $\delta'(s,a) = r$, where $r$ is the representative of the group containing $t$. 
\[ \delta = \{(B,a) \rightarrow B, (B,b) \rightarrow D, (C,a) \rightarrow B, (C,b) \rightarrow C, (D,a) \rightarrow B, (D,b) \rightarrow E, (E,a) \rightarrow B, (E,b) \rightarrow C\} \]