Tutorial Week 5

Definition 1. *The Dictionary Matching Automaton, $\mathcal{D}(X)$, it accepts the language A^*X, where X is a set of patterns, and is defined by:*

- *set of states is $\mathrm{Pref}(X)$; initial state is the empty string;*

- *set of terminal states is $\mathrm{Pref}(X) \cap A^*X$*

- *arcs are of the form $(u, a, h(ua))$, where $h(ua)$ is the longest suffix of ua that belongs to $\mathrm{Pref}(X)$*

The trie, $\mathcal{T}(X)$, associated to the set X is the digital tree whose branches are labelled by strings of X (it contains only forward arcs). $\mathcal{T}(X)$ is the basis of the Dictionary Matching Automaton, $\mathcal{D}(X)$.

Exercise 1. *Consider now the following sets of patterns:*
$X_1 = \{aa, abab, abaab, abba\}$;
$X_2 = \{aab, abb, aaba, abab\}$.

For each of the sets do the following:
a) *Draw their corresponding tries.*
b) *Give the failure table of each trie; what is that and how do you obtain it?*
c) *Draw their corresponding representation of the Dictionary Matching Automaton, including its failure links and final states.*

Exercise 2. *Consider now the set of patterns $X = \{aa, abab, abaab, abba\}$ (this is X_1 from the previous exercise) and the text $y = ababababaababbaaabbaba$. Fill up the following table, corresponding to the searching phase (use of the optimised failure function - note that this might generate several states for one value of j). Please underline all final states.*

j		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
$y[j]$		a	b	a	b	a	b	a	a	b	a	b	b	a	a	a	b	b	a	b	a
$state$	0																				