Minimalizations of NFA Using the Universal Automaton

Libor Polák [1]

Department of Mathematics, Masaryk University,
Janáčkovo nám. 2a, 662 95 Brno, Czech Republic

1 Introduction

In this contribution we shall see how our (algebraic) approach to the so-called universal automaton of a given regular language helps to understand the process of the minimalization of a NFA. The minimality is meant with respect to the number of states. Although it is known that the problem is PSPACE-complete (see [5]) one can propose algorithms of not too high complexities to compute relatively good approximations.

The universal automaton was considered implicitly already by Conway in [3]. Its significance for the problem of the minimalization of a NFA was stated by Arnold, Dicky and Nivat in [1] where the authors proved that an arbitrary minimal NFA for a given regular language L is isomorphic to a subautomaton of the universal automaton $\mathcal{U}$ for L. They credit that result to Carrez [2]. Note that also the calculations of Kameda and Weiner [6] were done implicitly in $\mathcal{U}$.

Our view on the universal automaton is based on author's paper [8]. The methods of several previous works on minimalizations of NFA can be modified so that they fit in our approach. We formulate various conditions on sets of states of the universal automaton $\mathcal{U}$ and we investigate the relationships between them. Any such set of states P induces a subatomaton $\mathcal{U}_P$ of $\mathcal{U}$. The conditions (L) and (B) determine the sets of states P_l and P_b in a unique way, so they form bases for concrete implementations. A checking of the conditions with the exception of "$\mathcal{U}_P$ accepts L" leads to polynomial time algorithms with respect to the dimension of the so-called basic matrix for L.

Let $L \subseteq A^*$ be a regular language. Let $\mathsf{D} = \{u^{-1}L \mid u \in A^*\} = \{u_1^{-1}L, \dots, u_n^{-1}L\}$, $\widehat{\mathsf{D}} = \{Lv^{-1} \mid v \in A^*\} = \{Lv_1^{-1}, \dots, Lv_m^{-1}\}$, $\mathsf{U} = \{\ w_1^{-1}L \cap \dots \cap w_k^{-1}L \mid k \geq 0,\ w_1, \dots, w_k \in A^*\ \}$. Let $B = (\beta_{ij})$ be a matrix of type m/n with entries from $\{0,1\}$ where $\beta_{ij} = 1$ if and only if $u_j v_i \in L$. This matrix is called the *basic matrix* of the language L. Adding to the columns of B new ones which are componentwise meets of sets of columns of B ($0 \wedge 0 = 0 \wedge 1 = 1 \wedge 0 = 0,\ 1 \wedge 1 = 1$) we get the matrix U which is called the *universal matrix* of L.

The *universal automaton* of a language L is a (non-deterministic) automaton $\mathcal{U} = (\mathsf{U}, A, E, I, T)$ where $(p, a, q) \in E$ if and only if $q \subseteq a^{-1}p$ and $q \in \mathsf{U}$ is an element of I if and only if $q \subseteq L$ and $q \in T$ if and only if $1 \in q$.

Note that the states of the minimal complete deterministic automaton of L correspond to the columns of B and the states of $\mathcal{U}$ correspond to the columns of U. Moreover, we can easily compute unions and intersections of states of $\mathcal{U}$ using the matrix U. Also the following is well-known.

Proposition 1 *Let $\mathcal{U} = (\mathsf{U}, A, E, I, T)$ be the universal automaton of a regular language L over an alphabet A. Then*
(i) $\mathcal{U}$ accepts L,
(ii) for each non-deterministic automaton $\mathcal{V} = (V, A, G, J, W)$ accepting a subset of L, the mapping

$$\phi : p \to \bigcap \{u^{-1}L \mid p \text{ is reachable from a } j \in J \text{ by a path labeled by } u\ \}$$

is an automaton homomorphism of $\mathcal{V}$ into $\mathcal{U}$,
(iii) for each $q \in \mathsf{U}$, the automaton $(\mathsf{U}, A, E, \{q\}, T)$ accepts exactly the language q.

2 Results

Let $\mathcal{U} = (\mathsf{U}, A, E, I, T)$ be the universal automaton of a regular language $L \subseteq A^*$. Each $P \subseteq \mathsf{U}$ induces a subautomaton $\mathcal{U}_P = (P, A, E_P, I \cap P, T \cap P)$ of $\mathcal{U}$ where $E_P = \{(p, a, q) \in E \mid p, q \in P\}$. Clearly, the language accepted by $\mathcal{U}_P$ is a subset of L. We formulate several conditions on a subset P of U:
(A) the automaton $\mathcal{U}_P$ accepts the language L,
(C) $\mathcal{U}_P$ is complete, i.e. $(\ \forall\, p \in P\ )(\ \forall\, a \in A\ )(\ \exists\, q \in P\ )\ q \subseteq a^{-1}p$,
(I) the initial state is covered, i.e. $L = \bigcup \{p \in P \mid p \subseteq L\}$,
(D) closeness with resp. to derivatives, i.e. $(\ \forall\, p \in P\ )(\ \forall\, a \in A\ )\ a^{-1}p \in P$,
(K) (Kiel) $(\ \forall\, p \in P\ )(\ \forall\, a \in A\ )\ a^{-1}p = \bigcup \{q \in P \mid q \subseteq a^{-1}p\}$,
(W) (Waterloo) $(\ \forall\, q \in \mathsf{D}\ )\ q = \bigcup \{p \in P \mid p \subseteq q\}$,
(LM) the local minimality, i.e. $P \models (A)$ but for each $p \in P$ we have $P \setminus \{p\} \models \neg(A)$,
(L) (Lille) $P_l = \{\ p \in \mathsf{D} \mid p \text{ is union-irreducible in } (\mathsf{D}, \subseteq),\ p \neq \emptyset\ \}$,
(B) (Brno) $P_b = \{\ \bigcap \{u_j^{-1}L \mid \beta_{ij} = 1\} \mid Lv_i^{-1} \text{ union-irred. in } (\widehat{\mathsf{D}}, \subseteq)\ \}$.
We also put $P_0 = \{\ q \in \mathsf{D} \mid q \neq \bigcup \{r \in \mathsf{U} \mid r \subseteq q,\ r \neq q\}\ \}$.

The following result relates the above conditions.

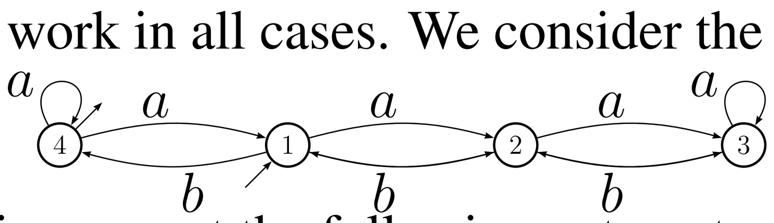
Theorem 1 *The following implications between our conditions hold.*
(i) (D) $\implies$ (K) & (C).
(ii) (L) $\implies$ (I) & (K) $\implies$ (A) $\implies$ (W).
Moreover, (iii) P_b satisfies (A).
(iv) Both P_l and P_b satisfy the condition (LM).
(v) For each $P \subseteq \mathsf{U}$ satisfying (A) we have $P_0 \subseteq P$.
(vi) None of the implications in (i) and (ii) can be reversed.

Compact proofs, algorithms for constructing the matrices B and U, and exact lines to the related works [4, 6, 7, 9] can be find in an extended version of this note.

3 Examples

In Examples 1–3 we have $A = \{a, b\}$ and $A = \{a\}$ in Ex. 4. In our diagrams of universal automata the covering relations of the order reducts are depicted by black lines, the transitions labeled by the letter a by red lines and transitions labeled by b by blue ones. We draw only the transitions $p \xrightarrow{c} c^{-1}p$, $p \in \mathsf{U}$, $c \in A$. Elements of D's are colored in yellow.

1. Kiel. This valuable example appears in [7] where the authors show that their methods do not work in all cases. We consider the following NFA.



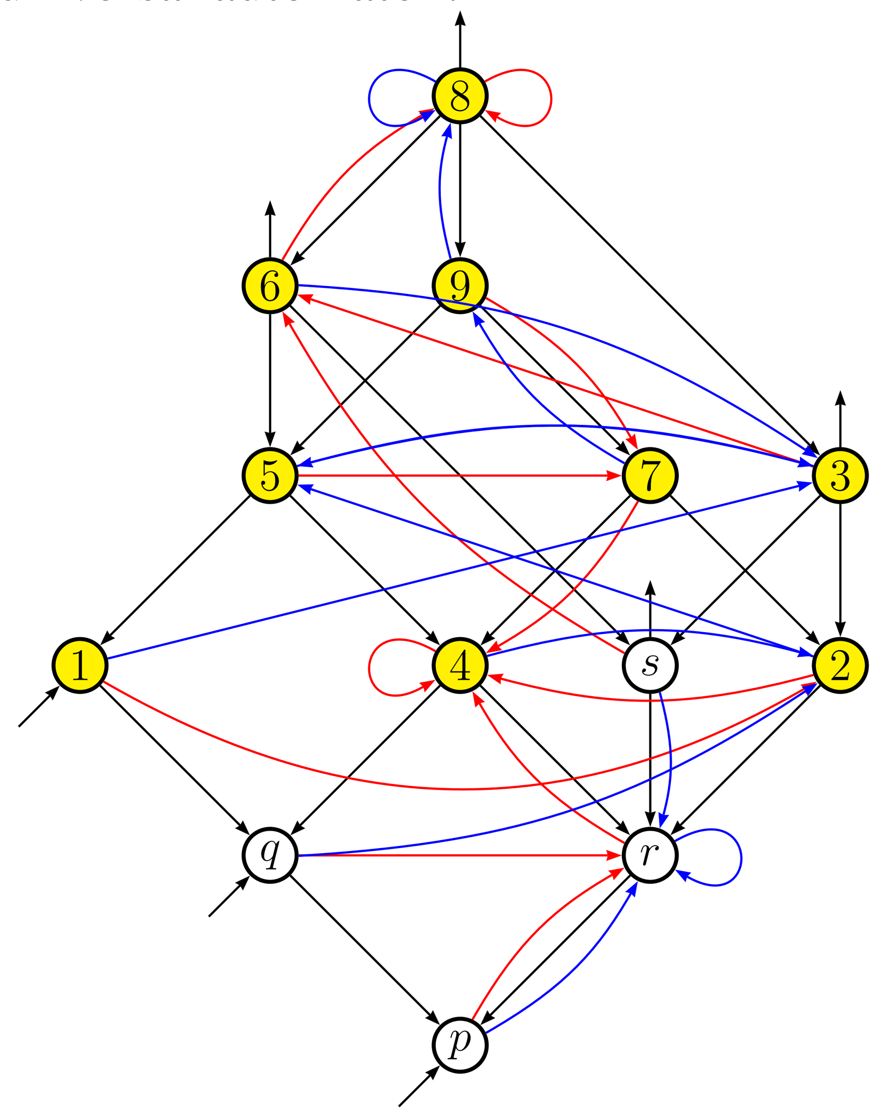
After determinization we get the following automaton.

	↓		↑			↑		↑	
	1	2	3	4	5	6	7	8	9
a	2	4	6	4	7	8	4	8	7
b	3	5	5	2	3	3	9	8	8

One can see that it is a minimal complete DFA and the universal matrix follows.

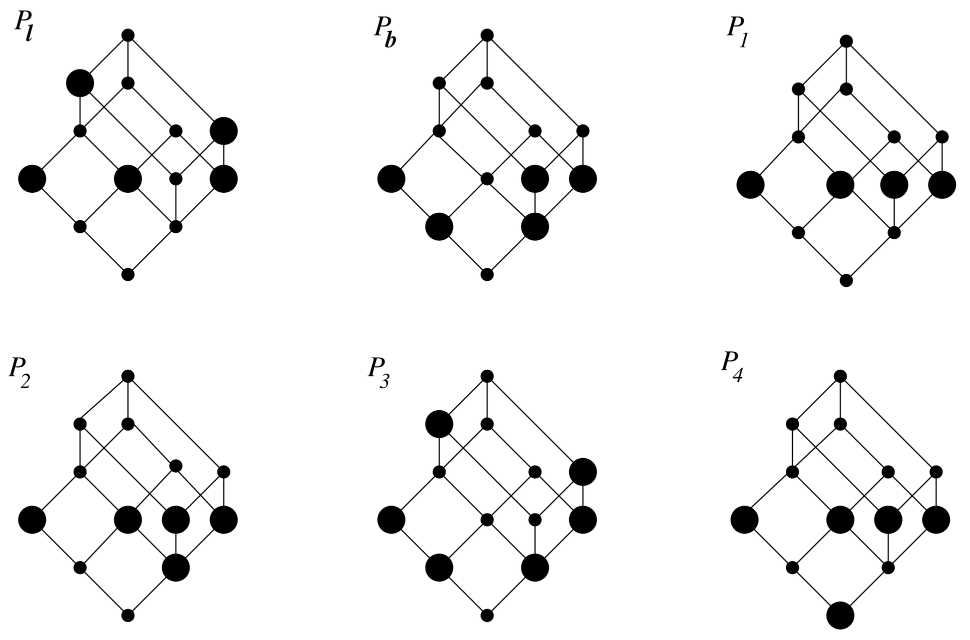
	1	2	3	4	5	6	7	8	9	p	q	r	s
1	0	0	1	0	0	1	0	1	0	0	0	0	1
b	1	0	0	0	1	1	0	1	1	0	0	0	0
b^2	0	1	1	0	0	0	1	1	1	0	0	0	0
b^3	1	0	0	1	1	1	1	1	1	0	1	0	0
ab^3	0	1	1	1	1	1	1	1	1	0	0	1	1
a^2b^3	1	1	1	1	1	1	1	1	1	1	1	1	1

Below is the universal automaton.



[1] Supported by the Ministry of Education of the Czech Republic under the project MSM 143100009

Consider the following choices for the set P.



Here $P_0 = \{1,2\}$ and the locally minimal sets of states with respect to (A) are exactly P_l, P_b, P_1 and P_3. Note also that P_1 is the ϕ-image of the automata we started with and P_2 is the image of its completion. Further, P_b, $P_2 \models (C), \neg(D), (K)$, $P_1 \models \neg(C), \neg(D), \neg(K)$, $P_4 \models (C), \neg(D), \neg(K)$.

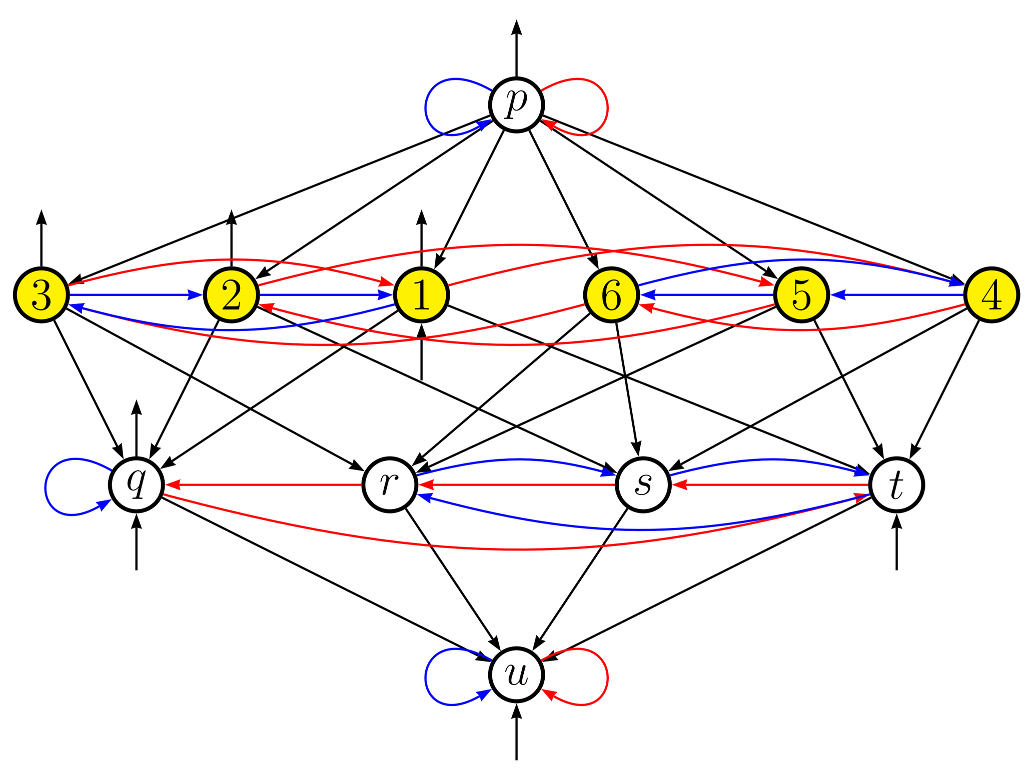
2. Lille. This example is a member of a sequence of automata used by the authors of [4] to show that their methods could lead to results which are exponentially worse than the optimum. The following table determines a complete quasideterministic (several initial states are allowed) automaton.

	↓↑	↓		
	1	2	3	4
a	2	3	4	1
b	1	4	2	3

Its universal matrix follows.

	1	2	3	4	5	6	p	q	r	s	t	u
1	1	1	1	0	0	0	1	1	0	0	0	0
a	0	0	1	0	1	1	1	0	1	0	0	0
a^2	0	1	0	1	0	1	1	0	0	1	0	0
ba	1	0	0	1	1	0	1	0	0	0	1	0

The universal automaton is depicted below.



We have $P_0 = \emptyset$ and $P_l = \{1,\ldots,6\}$, $P_b = \{q,r,s,t\}$ are the only sets of states locally minimal with respect to (A).

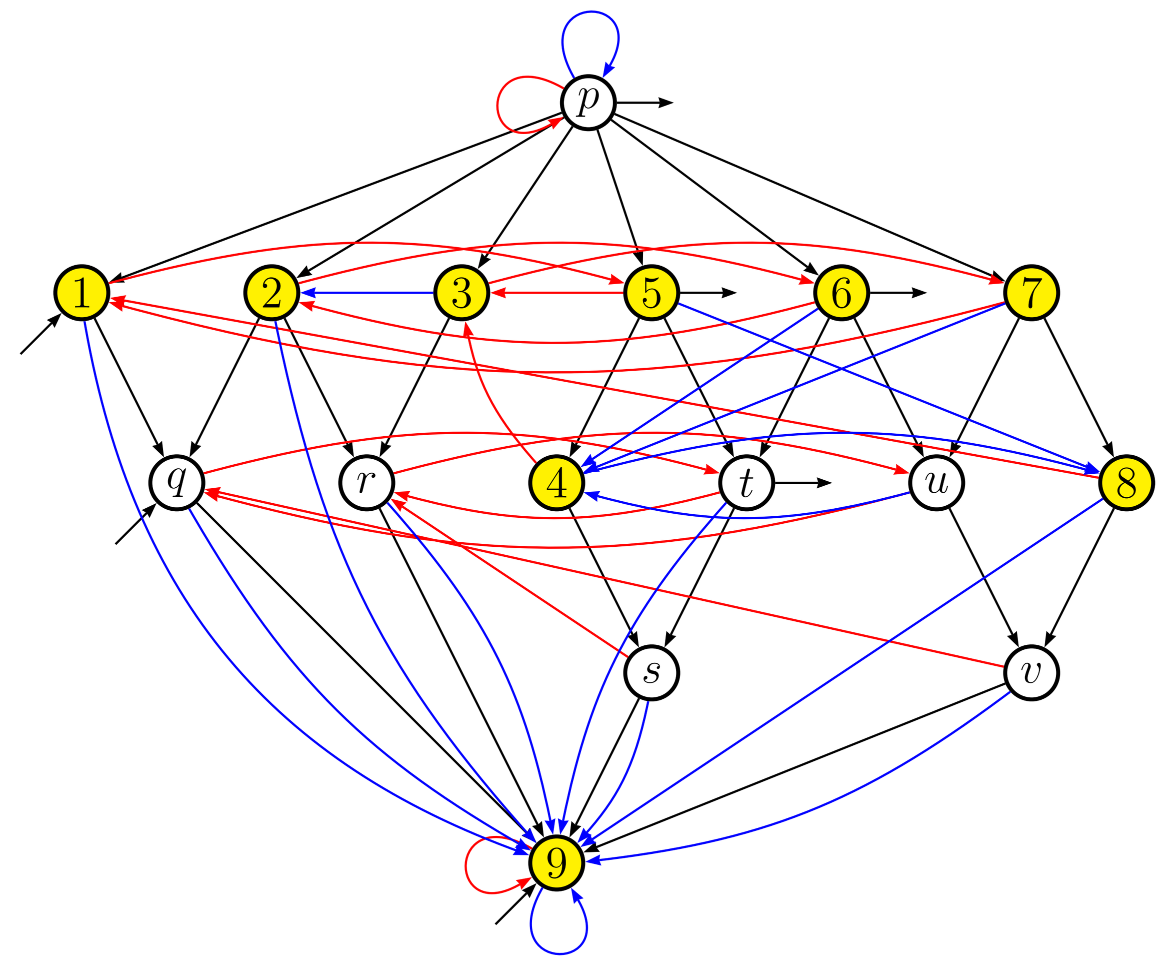
3. Waterloo. This example is taken from [6]. The following table presents an (incomplete) DFA.

	↓				↑	↑		
	1	2	3	4	5	6	7	8
a	5	6	7	3	3	2	1	1
b	–	–	2	8	8	4	4	–

Its universal matrix follows.

	1	2	3	4	5	6	7	8	9	p	q	r	s	t	u	v
1	0	0	0	0	1	1	0	0	0	1	0	0	0	1	0	0
a	1	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0
b	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
a^2	0	0	0	0	0	1	1	1	0	1	0	0	0	0	1	1
ba	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0
a^3	0	1	1	0	0	0	0	0	0	1	0	1	0	0	0	0
aba	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0
a^4	0	0	0	1	1	1	0	0	0	1	0	0	1	1	0	0
a^2ba	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
a^3ba	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0
ba^4	0	0	0	0	0	1	1	0	0	1	0	0	0	0	1	0

The universal automaton is depicted below.

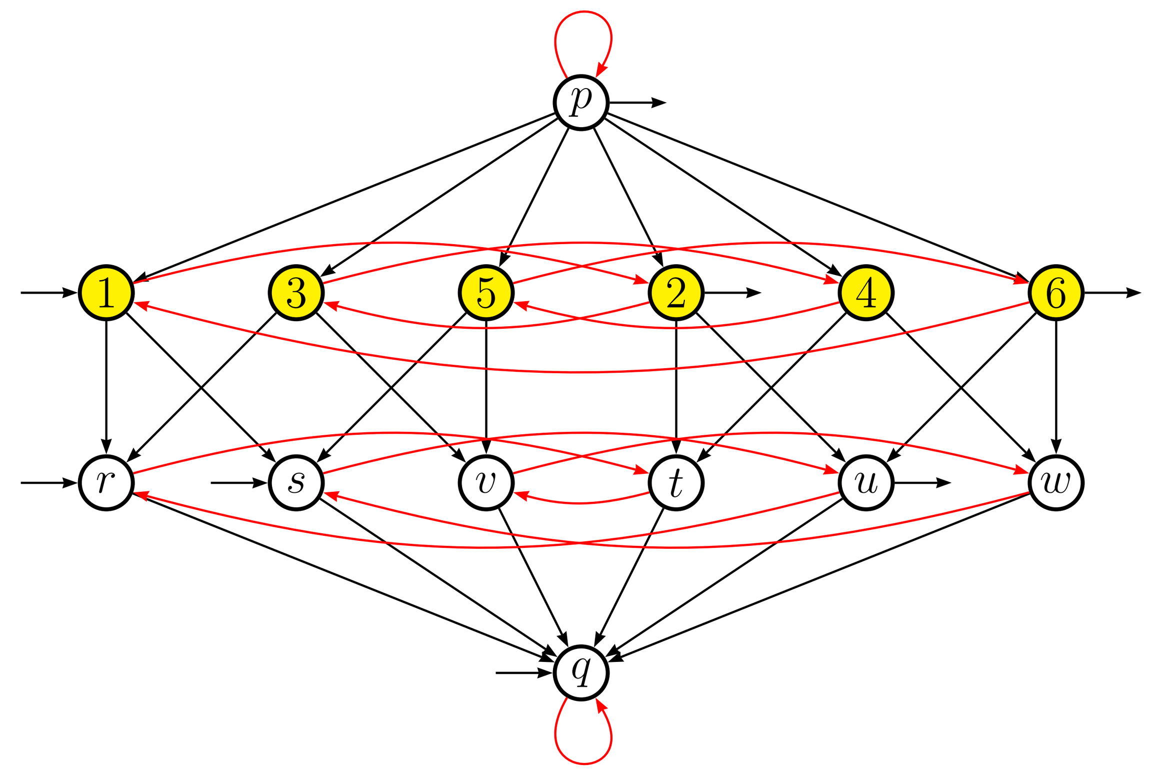


We have $P_0 = \{1,3,4,8\}$ and the locally minimal sets of states with respect to (A) are exactly $P_l = P_0 \cup \{2,5,6,7\}$, $P_b = P_0 \cup \{q,r,t,u\}$, $P_1 = P_0 \cup \{2,5,6,u\}$, $P_2 = P_0 \cup \{2,6,t,u\}$ and $P_3 = P_0 \cup \{2,6,7,t\}$. We have that $P_4 = P_0 \cup \{2,t,u\} \models (W), \neg(A)$ as already noted in [6].

4. Brno. This example is due to Ivana Vařeková, a PhD student in Brno. Consider the language $(a^6)^*(a + a^5)$ over the alphabet $A = \{a\}$. Its universal matrix follows.

	1	2	3	4	5	6	p	q	r	s	t	u	v	w
1	0	1	0	0	0	1	1	0	0	0	0	1	0	0
a	1	0	0	0	1	0	1	0	0	1	0	0	0	0
a^2	0	0	0	1	0	1	1	0	0	0	0	0	0	1
a^3	0	0	1	0	1	0	1	0	0	0	0	0	1	0
a^4	0	1	0	1	0	0	1	0	0	0	1	0	0	0
a^5	1	0	1	0	0	0	1	0	1	0	0	0	0	0

Below is the universal automaton.



Here $P_0 = \emptyset$ and the locally minimal sets of states with respect to (A) are exactly $P_l = \{1,\ldots,6\}$ and $P_b = \{r,s,t,u,v,w\}$. Note that $P_1 = \{1,2,3,4,5,q,u,v,w\} \models (W), (C), \neg(A)$.

References

[1] Arnold, A., Dicky, A. and Nivat, M.; A note about minimal non-deterministic automata, Bull. EATCS 47, 166–169 (1992)

[2] Carrez, C.; On the minimalization of non-deterministic automaton, Laboratoire de Calcul de la Faculté des Sciences de l'Université de Lille, 1970

[3] Conway, J.H.; *Regular Algebra and Finite Machines*, Chapman and Hall, Ltd., 1971

[4] Denis, F., Lemay, A. and Terlutte, A.; Residual finite state automata, Fundam. Inform. 51, No.4, 339–368 (2002)

[5] Jiang, Tao and Ravikumar, B.; A note on the space complexity of some decision problems for finite automata. Inf. Process. Lett. 40, No.1, 25–31 (1991)

[6] Kameda, T. and Weiner, P.; On the state minimalization of nondeterministic finite automata, IEEE Transactions on Computers, Vol. C-19, No. 7, 617–627 (1970)

[7] Matz, O. and Potthoff, A.; Computing small nondeterministic finite automata, *Proc. Workshop on Tools and Algorithms for the Construction and Analysis of Systems 1995*, BRICS Lecture Notes, pp. 74–88

[8] Polák, L.; Syntactic semiring and universal automaton, *Proc. Developments in Language Theory, Szeged 2003*, Springer LNCS, Vol. 2710, pp. 411–422 (2003)

[9] Tamm, H. and Ukkonen, E.; Bideterministic automata and minimal representations of regular languages, *Proc. CIAA, Santa Barbara (USA), 2003*, Springer LNCS, Vol. 2759, pp. 61–71 (2003)