Zadanie 5

W całym rozwiązaniu przyjmuję alfabet zgodny z treścią zadania. Słowa wczytywane przez automat A_n utożsamiamy ze ścieżkami przechodzonymi przez automat (jest to możliwe, gdyż duży rozmiar alfabetu pozwala łatwo z kolejnych stanów otrzymać słowo, które automat wczytał). Zauważmy, że wszystkie podwyrażenia szukanego wyrażenie E muszą generować jedynie słowa, które są legalne dla automatu A_n zaczynającego pracę w pewnym stanie (tzn. podwyrażenie E nie może być np. postaci $a_{12}a_{32}$, gdyż wtedy w $L(E)$ znalazłoby się słowo, którego nie może tam być – trudno mi jaśniej sformułować tę własność).

Niech E będzie wyrażeniem regularnym. Niech S będzie zbiorem ścieżek w automacie A_n dla pewnego n. Powiemy, że wyrażenie E opisuje zbiór ścieżek S jeśli słowa z języka $L(E)$ są utożsamione ze ścieżkami z S (nieformalnie możemy napisać $L(E)=S$). Niech p będzie dowolną, niepustą ścieżką w grafie automatu. Powiemy, że S pokrywa ścieżkę p jeśli istnieje taka ścieżka q należąca do S, że p jest podścieżką q (tzn. $p=uqw$ dla pewnych u i w). Wyrażenie E pokrywa p jeśli zbiór ścieżek opisywany przez E pokrywa p.

Lemat 1:
Jeśli E pokrywa p^k dla pewnego $k>2|E|$, to E pokrywa też każde słowo postaci p^*.
Dowód:
Rozpatrzmy automat skończony równoważny wyrażeniu E o minimalnej liczbie stanów (oznaczmy ten automat A). Na mocy twierdzenia z wykładu, ten automat ma mniej niż $2|E|$ stanów. Automat A akceptuje słowo up^kw dla pewnych u i w. Ponieważ k jest większe od liczby stanów automatu, to podczas wczytywania up^kw, z zasady szufladowej Dirchleta, dla pewnych i i j, po wczytaniu prefiksów up^i i up^ip^j , automat znajduje się w tym samym stanie. Stąd A akceptuje wszystkie słowa postaci $up^i(p^j)^*p^{k-i-j}w$. Zatem E (przez równoważność z A pokrywa każde słowo postaci p^*. Co kończy dowód lematu.

Największe k takie, że E pokrywa p^k oznaczymy przez $k(E,p)$ ($k(E,p)$ nie musi więc istnieć).

Lemat 2:
Jeśli $k(E,p)$ nie istnieje, to istnieje takie X^*, które jest podwyrażeniem E, że $k(X,p)$ istnieje, ale $k(X^*,p)$ nie istnieje.
Dowód:
Weźmy Y jako minimalne podwyrażenie E, dla którego $k(Y,p)$ nie istnieje. Zauważmy najpierw, że Y nie może być symbolem z alfabetu (gdyż wtedy $k(Y,p)$ zawsze istnieje). Gdy $Y=A+B$ lub $Y=AB$ i istnieją $k(A,p)$ oraz $k(B,p)$, to istnieje również $k(Y,p)$ – zatem Y nie może być także sumą ani złożeniem podwyrażeń. Pozostaje przypadek, gdy $Y=X^*$. Ten X spełnia tezę lematu, ze względu na sposób wyboru Y. Co kończy dowód lematu.

Pokażemy przez indukcję, że dla każdej liczby naturalnej n istnieje ścieżka p, której początek i koniec są równe 1, taka, że jeśli E pokrywa p, to $|E| \geq 2^{n-1}$. Oczywiście wtedy teza głównego twierdzenia będzie natychmiastowa: słowo utożsamione z p należy do $L(A_n)$, więc każde wyrażenie E takie, że $L(E)=L(A_n)$ musi pokrywać p. Zatem jego długość musi być nie mniejsza niż np. $(1,5)^n$ (dla odpowiednio dużych n).

Dla $n=1$ teza indukcyjna jest trywialna (p jest ścieżką jednoelementową).
Przyjmijmy, że mamy ścieżkę p spełniającą tezę dla $n-1$. Niech p_k będzie ścieżką powstałą z p przez permutację wierzchołków automatu $i \mapsto (i+k) \bmod n$. Zauważmy, że p_k ma wtedy początek w stanie k oraz nie przechodzi przez stan $(k-1) \bmod n$. (dalsza arytmetyka na indeksach p_k odbywa się implicite modulo n). Rozpatrzmy ścieżkę:

$q = p_1^m a_{12} p_2^m a_{23} \ldots p_n^m a_{n1}$, gdzie $m=2^n$.

Niech wyrażenie E pokrywa q. Oczywiście wtedy dla każdego i, gdy $k(E,p_i)$ istnieje, to $k(E,p_i) \geq 2^n$. Na mocy lematu 1, gdy $k(E,p_i)$ istnieje, to $2|E|>k(E,p_i)$, więc albo $|E| \geq 2^{n-1}$ (co kończy dowód kroku indukcyjnego), albo dla wszystkich i wartość $k(E,p_i)$ nie istnieje.

W takim razie, na mocy poczynionych wcześniej obserwacji, możemy w E wybrać dla każdego i minimalne podwyrażenie X_i^*, że $k(X_i^*,p_i)$ nie istnieje. Spośród tych podwyrażeń wybierzmy wyrażenie minimalne w sensie długości i oznaczmy X^*. Ponieważ X jest podwyrażeniem E, to na mocy poczynionej na początku uwagi, pierwszym symbolem każdego ze słów z $L(X)$ musi być a_{jk} dla ustalonego j i dowolnego k (inaczej część generowanych przez $L(E)$ słów nie mogłaby być opisem przejść A_n). Ten ustalony wierzchołek nazywać będziemy j (wierzchołek początkowy) również w dalszym wywodzie. Spośród X_i^* pokrywających p_{j+1} wybierzmy jedno, oznaczmy je Y^*. Jeśli X^* i Y^* są rozłącznymi podwyrażeniami E, to ponieważ każde z nich

jest długości przynajmniej 2^{n-2} (na mocy założenia indukcyjnego), więc $|E| \geq |X| + |Y| \geq 2^{n-1}$, co kończy dowód. W przeciwnym razie, Y^* musi być podwyrażeniem (być może niewłaściwym) X^* (ponieważ zarówno X^* jak Y^* wybraliśmy spośród X_i^*, zaś X^* miało być najkrótsze). Weźmy wyrażenie Z^* powstałe przez podstawienie w Y^* pod X^* symbolu ε. Wtedy Z^* pokrywa p_{j+1}, ponieważ Y^* pokrywało p_{j+1}, a p_{j+1} nie przechodzi przez j – wierzchołek początkowy wyrażenia X^*. Zatem z założenia indukcyjnego mamy $|Z^*| \geq 2^{n-2}$. Wyrażenie Y^* jest dłuższe od Z^* o długość podstawionego wyrażenia X^*, zatem jest długości przynajmniej 2^{n-1}. Z kolei Y^* jest podwyrażeniem E, więc również $|E| \geq 2^{n-1}$. cnd.