

Dlaczego niektóre łamigłówki są tak trudne?

*Nokia

Łukasz GRZĄDKO*

Inspiracją do napisania tego artykułu jest znana, popularna i – do czego chcę Czytelnika przekonać – całkiem niełatwa łamigłówka zwana Sudoku. Problem polega na uzupełnieniu częściowo wypełnionej planszy 9×9 w taki sposób, żeby każdy wiersz i każda kolumna oraz każdy z 9 tzw. regionów 3×3 zawierał wszystkie cyfry od 1 do 9. Czy nie przypomina to pewnego innego równie znanego problemu natury kombinatorycznej? Tak, to problem Uzupełniania Kwadratów Łacińskich, których wynalazcą był Leonhard Euler. Być może zainspirował innych, by w przyszłości stworzyli Sudoku. Zasady Uzupełniania (częściowo wypełnionych) Kwadratów Łacińskich są niemal identyczne jak w Sudoku, z tą drobną różnicą, że pomijamy wymaganie dotyczące regionów 3×3 . Gwoli uzupełnienia rysu historycznego dodajmy, że Sudoku zostało wymyślone przez Howarda Garnsa. Po raz pierwszy zostało opublikowane w 1979 r. przez magazyn *Dell Magazines*, a siedem lat później dotarło w końcu do Japonii, gdzie zdobyło dużą popularność, jak i swoje prawdziwe imię. Nazwa bowiem wzięła się z japońskiego wyrażenia „Suji wa dokushin ni kagiru” (数字は独身に限る), co dosłownie znaczy „Liczby muszą być pojedyncze”.

W wersji popularnej, w kontekście Sudoku, zawsze myślimy o planszy 9×9 . W wersji uogólnionej rozważamy siatkę złożoną z $n^2 \times n^2$ komórek, podzieloną na $n \times n$ regionów, każdy rozmiaru $n \times n$. Niektóre komórki na starcie wypełnione są liczbami naturalnymi z przedziału od 1 do n^2 . Celem jest wypełnienie pozostałych komórek tak, żeby każdy wiersz, każda kolumna oraz każdy region zawierał każdą z liczb z przedziału od 1 do n^2 dokładnie jeden raz. Podobnie możemy, oczywiście, rozważać problem Uzupełniania Kwadratów Łacińskich dla dowolnego n .

Okazuje się, że zarówno Uzupełnianie Kwadratów Łacińskich, jak i gra Sudoku są problemami NP-zupełnymi, a więc wierzymy, że bardzo trudnymi obliczeniowo. Pozostała część artykułu to właśnie próba naszkicowania, jak w ogóle można takie fakty dowodzić.

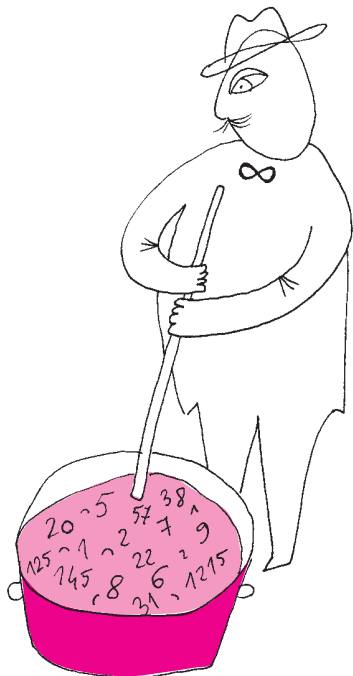
Otóż, aby udowodnić, że jakiś problem jest NP-pełny, wystarczy wziąć dowolny problem, o którym już wiemy, że jest NP-pełny i pokazać, że daje się go zredukować do naszego problemu. (Gwoli ścisłości w ten sposób tak naprawdę pokazujemy, że ten problem jest co najmniej tak trudny jak inne problemy NP-pełne. Należałoby jeszcze uzasadnić, że nie jest trudniejszy, co jest zwykle bardzo łatwe. Tutaj również, ale to pomijamy.) Zredukować, tzn. pokazać efektywną (wielomianową) konstrukcję, która przerabia dowolną instancję I_1 pierwszego problemu na instancję I_2 drugiego problemu, oczywiście z zachowaniem poprawności odpowiedzi (tzn: problem I_1 ma rozwiązanie wtedy i tylko wtedy, gdy I_2 ma rozwiązanie). Dokładnie tę metodę zaprezentujemy za chwilę. Na start weźmiemy problem 3SAT, o którym wiadomo, że jest NP-pełny, a następnie naszkicujemy ciąg redukcji:

3SAT \rightarrow TTP \rightarrow Uzupełnianie Kwadratów Łacińskich \rightarrow Sudoku.

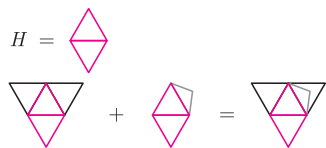
Do dzieła!

Redukcja: 3SAT \rightarrow TTP. Problem 3SAT to problem, który jest opisany przez długą koniunkcję klauzul, z których każda jest alternatywą trzech literalów (np. $(x_2 \vee \neg x_4 \vee \neg x_5) \wedge (x_1 \vee x_2 \vee x_3) \wedge (\neg x_1 \vee \neg x_3 \vee x_5)$). Pytamy, oczywiście, o spełnialność wejściowej formuły. Natomiast problem TTP (Triangulate Tripartite Problem) to problem triangulacji grafów trójdzielnych. Zanim przejdziemy do właściwej redukcji, potrzebujemy szeregu definicji.

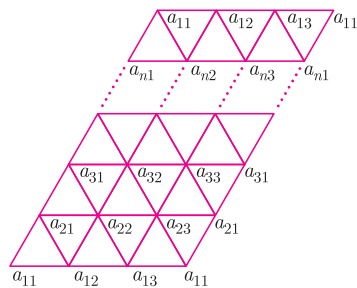
Graf *trójdzielny* to taki, którego wierzchołki można podzielić na trzy grupy w taki sposób, aby żadna krawędź nie łączyła wierzchołków z tej samej grupy. *Triangulacja* grafu to jego podział na rozłączne krawędziowo trójkąty. Rozważmy teraz dwa grafy G_1 i G_2 i przypuśćmy, że oba mają wspólny podgraf H . Możemy *skleić* G_1 z G_2 wzdłuż H . Zamiast definicji przedstawimy rysunek 1. Potrzebujemy jeszcze specjalnego grafu $H_{3,n}$, który pokazany jest na rysunku 2,



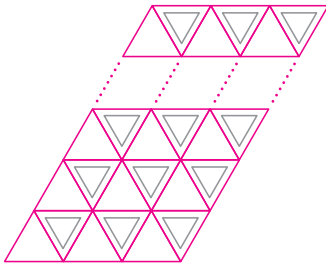
Jak łatwo zauważyć, sposób dowodzenia NP-zupełności przez redukcję ma sens, o ile już znamy jakiś problem NP-pełny. Pierwszym problemem, o którym bezpośrednio pokazano, że jest NP-pełny, był problem SAT (spełnialności formuł logicznych). Jest to treść słynnego twierdzenia Cooka z roku 1971, uważanego przez wielu za najważniejsze twierdzenie informatyki teoretycznej.



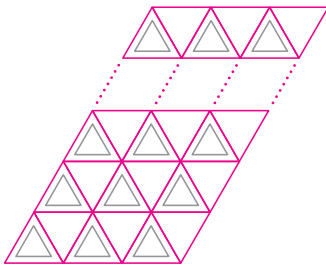
Rys. 1. Przykład sklejanego dwóch grafów wzdłuż grafu H



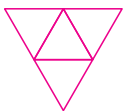
Rys. 2. Przykład grafu $H_{3,n}$. Wierzchołki na brzegach są tożsame.



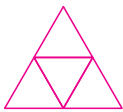
Rys. 3. Przykład F-triangulacji



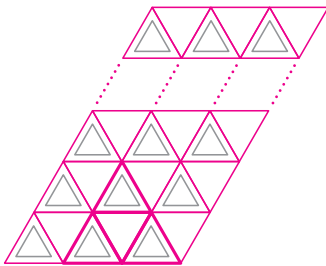
Rys. 4. Przykład T-triangulacji



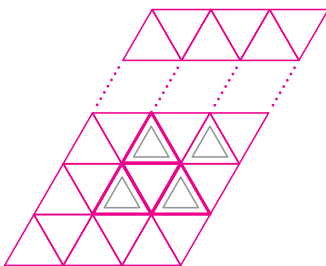
Rys. 5. Przykład T-płatka



Rys. 6. Przykład F-płatka



Rys. 7. T-triangulacja grafu H_3 oraz wyróżniony F-płatek



Rys. 8. Striangulowany graf z rysunku 7 sklejamy z powyższym grafem wzdłuż wyróżnionego F-płatka. Wymusza to T-triangulację całego grafu.

a który nie jest niczym innym jak tylko trójwymiarowym torusem zbudowanym z trójkątnej siatki.

Ten graf ma kilka interesujących własności. Przede wszystkim istnieją tylko dwie różne (dlaczego?) triangulacje grafu $H_{3,n}$, obie zilustrowane są na rysunkach 3 i 4. Tę triangulację, w której używamy trójkątów z czubkiem do góry, nazywamy T-triangulacją, a tę drugą – F-triangulacją. Pierwsza będzie odpowiadała wartości logicznej „prawda”, druga – „fałsz”. Graf $H_{3,n}$ jest grafem trójdzielnym, podobnie jak wszystkie rozważane sklejenia wielu jego kopii. Użytecznymi dla nas podgrafami grafu $H_{3,n}$ będą jeszcze F-płatek oraz T-płatek, zilustrowane na rysunkach 5 i 6.

Potrzebujemy jeszcze trzech lematów (ściśle dowody pomijamy, ale pomocne będą rysunki 7–12).

Lemat o płatkach zgodnych. Przypuśćmy, że mamy dwa grafy $H_{3,n}$ sklejone F-płatkami. Załóżmy, że znaleźliśmy triangulację wyniku tej operacji. Jest to możliwe wyłącznie wtedy, gdy albo oba grafy mają T-triangulację, albo jeden z nich ma T-triangulację, a drugi F-triangulację.

Lemat o płatkach niezgodnych. Tym razem przypuśćmy, że mamy dwa grafy $H_{3,n}$ połączone przez sklejenie T-płatka pierwszego grafu z F-płatkami drugiego grafu i znów całość została striangulowana. Jest to możliwe wyłącznie wtedy, gdy oba grafy mają każdą z trzech możliwych par konfiguracji triangulacji oprócz takiej, że pierwszy ma F-triangulację, a drugi T-triangulację.

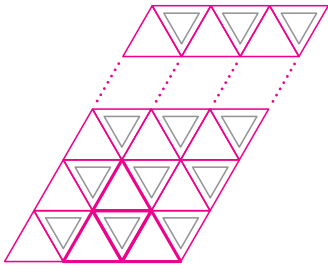
Lemat o wycięciu trójkąta. Przypuśćmy, że k grafów $H_{3,n}$ jest sklejonych wzdłuż T-płatka oraz usuwamy środkowy trójkąt typu F z części wspólnej tych grafów. Jest to możliwe wyłącznie wtedy, gdy dokładnie jeden graf ma F-triangulację, a pozostałe mają T-triangulację.

Jesteśmy wreszcie gotowi na właściwą redukcję.

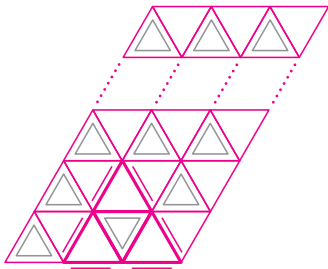
Rozważmy instancję problemu 3SAT, tj. zbiór klauzul $C = \{C_1, C_2, \dots, C_r\}$ oraz s zmiennych u_1, u_2, \dots, u_s , gdzie C_i składa się z literałów $l_{i,1}, l_{i,2}, l_{i,3}$. Wybieramy n tak duże, aby co najmniej $3r$ płatków nie kolidowało ze sobą w grafie $H_{3,n}$. Niech teraz kopia U_i grafu $H_{3,n}$ reprezentuje zmienną u_i , a kopie $C_{i,1}, C_{i,2}, C_{i,3}$ tego samego grafu reprezentują klauzulę C_i . Wszystkie te kopie sklejamy w pewien szczególny sposób. Jeśli $l_{i,j} = u_k$, to sklejamy znaleziony wolny F-płatek w $C_{i,j}$ z F-płatkami w U_k . W przeciwnym przypadku jeśli $l_{i,j} = \neg u_k$, wtedy sklejamy znaleziony wolny F-płatek w $C_{i,j}$ z T-płatkami w U_k . Ponadto łączymy grafy $C_{i,1}, C_{i,2}, C_{i,3}$ wzdłuż dowolnego wolnego F-płatka w każdym z grafów i usuwamy krawędzie centralnego trójkąta typu F. Efekt naszych działań oznaczmy jako G_3 . Pokażemy, że istnieje triangulacja grafu G_3 wtedy i tylko wtedy, gdy wejściowa formuła C jest spełnialna. To, oczywiście, zakończy dowód redukcji.

Przypuśćmy wprawdzie, że istnieje triangulacja grafu G_3 . Z poprzednich obserwacji jasno wynika, że każda ze składowych grafów ma triangulację T lub F.

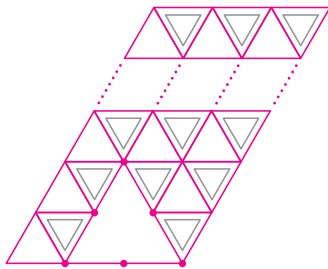
Przypuśćmy, że $l_{i,j} = u_k$ i rozważmy złączenie grafów $C_{i,j}$ i U_k . Twierdzimy, że krawędzie w sąsiedztwie złączenia mogą być triangulowane wtedy i tylko wtedy, gdy co najmniej jeden z grafów jest T-triangulowalny. Wynika to jednak wprost z lematu o płatku zgodnym. Podobnie jeśli $l_{i,j} = \neg u_k$, wtedy z lematu o płatku niezgodnym nie może jednocześnie $C_{i,j}$ być F-triangulowalny oraz U_k T-triangulowalny. Teraz rozważmy złączenie pomiędzy grafami $C_{i,1}, C_{i,2}, C_{i,3}$. Na podstawie lematu o wycięciu trójkąta wnioskujemy, że zbiór krawędzi w sąsiedztwie złączenia jest triangulowalny wtedy i tylko wtedy, gdy dokładnie jeden z grafów $C_{i,1}, C_{i,2}, C_{i,3}$ jest F-triangulowalny. Stąd wynika, że jeśli G_3 jest triangulowalny, to istnieje wartościowanie u_1, u_2, \dots, u_s , które spełnia C , mianowicie u_k ma wartość prawdziwą wtedy i tylko wtedy, gdy U_k jest T-triangulowalny. W drugą stronę: jeśli C jest spełnialna, wtedy dzielimy G_3 przez podzielenie U_k względem wartości zmiennej u_k , wybierając dla każdej



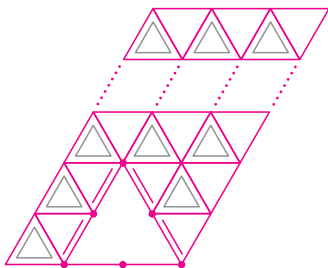
Rys. 9. F-triangulacja grafu $H_{3,n}$ oraz wyróżniony F-płatek



Rys. 10. Striangulowany graf z rysunku 9 sklejamy z powyższym grafem wzdłuż wyróżnionego F-płatka. Wymusza to T-triangulację pozostałej części powyższego grafu.



Rys. 11. F-triangulacja grafu $H_{3,n}$ oraz wyróżniony płatek z wyciętym trójkątem



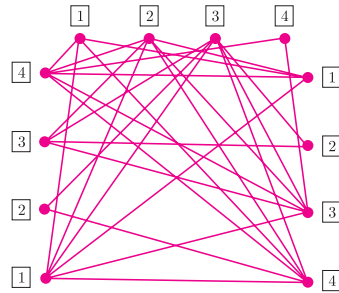
Rys. 12. Striangulowany graf z rysunku 11 sklejamy z powyższym grafem wzdłuż wyróżnionego płatka. Wymusza to T-triangulację pozostałej części powyższego grafu.

klauzuli takie $l_{i,j}$, które jest prawdziwe. Wtedy odpowiedni graf $C_{i,j}$ dzielimy według F-triangulacji, a pozostałe dwa są T-triangulowalne.

Konkretny przykład tej (przynajmy, dość zawilej) redukcji znajduje się na tylnej okładce. Pozostałe redukcje tylko zarysujemy.

Redukcja: TTP \rightarrow Uzupełnianie Kwadratów Łacińskich

Formalny opis redukcji i dowód (niestety, dość skomplikowany technicznie – potrzeba twierdzenia Halla o małżeństwach i algorytmu Hopcrofta–Karpa szukania maksymalnych skojarzeń w grafach dwudzielnych) Czytelnik Zainteresowany Szczegółami znajdzie w [1]. My pokażemy tylko poniższy przykład.



			6	2	5	7	8
8	5		7	1	2	3	6
5			8	4	1	6	7
		8		6	7	2	5
2	6	1	4	7	8	5	3
3	2	7	1	5	6	8	4
6	7	5	2	8	3	4	1
7	8	6	5	3	4	1	2

Zagadka 1. Podziel graf na rozłączne trójkąty.

Zagadka 2. Uzupełnij planszę do pełnego Kwadratu Łacińskiego.

Mam nadzieję, że *widać*, jak rozwiązanie zagadki z prawej strony przekłada się na rozwiązanie zagadki z lewej strony: jeśli liczbę S wpisemy do pustego miejsca w wierszu R w kolumnie C w zagadce prawej, to interpretujemy to jako zaznaczenie trójkąta o wierzchołkach: R z lewej, C na górze i S z prawej, w zagadce lewej.

Redukcja: Uzupełnianie Kwadratów Łacińskich \rightarrow Sudoku

Ponownie ograniczymy się do przykładu. Tym razem nie z powodu trudności technicznych, tylko dokładnie odwrotnie – wierzymy, że przykład wyjaśni wszystko. Wszystkie liczby są podane w systemie trójkowym, gdyż powinno to ułatwić zrozumienie redukcji. Problem Kwadratu jest tak „przerobiony”, aby wymusić własność, że w rozwiązaniu Sudoku wszystkie nowo wpisane liczby mają zero na pozycji jedności. Bardziej znacząca cyfra bezpośrednio koduje liczbę z Kwadratu.

0		
	1	
		2

Zagadka 3. Uzupełnij planszę do pełnego Kwadratu Łacińskiego.

00	01	02		11	12		21	22
	11	12	10	21	22		01	02
	21	22		01	02	20	11	12
01	02	10	11	12	20	21	22	00
11	12	20	21	22	00	01	02	10
21	22	00	01	02	10	11	12	20
02	10	11	12	20	21	22	00	01
12	20	21	22	00	01	02	10	11
22	00	01	02	10	11	12	20	21

Zagadka 4. Rozwiąż podane Sudoku 9×9 .

[1] Charles Colbourn, *The Complexity of Completing Partial Latin Squares*, Discrete Applied Mathematics 8 (1984).

Posłowie. Z natury rzeczy w tekście powyżej musiały się znaleźć pewne uproszczenia. Przede wszystkim, warto podkreślić, że samo pojęcie NP-zupełności nie może dotyczyć ani pojedynczego problemu, ani skończonej klasy problemów. Problemy NP-zupełne dotyczą zawsze pewnego nieskończonego ciągu problemów, które zwykle są coraz większe (stąd zresztą w tekście mowa o uogólnionym Sudoku czy o uogólnionym problemie Uzupełniania Kwadratów Łacińskich). Oznacza to, że powyższy szkic wcale nie tłumaczy wprost, dlaczego akurat łamigłówka Sudoku rozmiaru 9×9 jest trudna. Tak naprawdę możemy tylko powiedzieć, że łamigłówki Sudoku $n^2 \times n^2$ są *asymptotycznie* bardzo trudne (o ile, oczywiście, $P \neq NP$). Niemniej pozwala to przypuszczać, że również początkowe łamigłówki całego ciągu są trudne.

T. K.