

Matematyczne podstawy informatyki

Część 4: Twierdzenie Herbranda

Antoni Kościelski

9 lutego 2015

Spis treści

1 Sylwetka Herbranda	1
2 Twierdzenie Herbranda	2
2.1 Rodzaje formuł	2
2.2 Skolemizacja	2
2.3 Własność skolemizacji	3
2.4 Model Herbranda	4
2.5 Model Herbranda wyznaczony przez daną strukturę	4
2.6 Kolejny krok	5
2.7 Implementacje algorytmu Herbranda	7
2.7.1 Próba Paula Gilmora	7
2.7.2 Metoda Martina Davisa i Hilarego Putnama	7
2.7.3 Badanie sprzeczności według Davisa i Putnama	8
2.7.4 Metoda Johna Alana Robinsona	9
2.8 Przykład	13
2.9 Twierdzenie o pełności wg Herbranda	14
2.10 Dodatek: Teorie Henkina i szczegółowy dowód twierdzenia o pełności	20
3 Funkcje obliczalne według Herbranda i Gödla	23
3.1 Terminy	23
3.2 Równości i układy	24
3.3 Wyprowadzenia	24
3.4 Analiza pojęcia wyprowadzenia	24
3.5 Funkcje obliczalne	25

1 Sylwetka Herbranda

Jacques Herbrand urodził się 12 lutego 1908 roku. Był niesłychanie zdolnym człowiekiem. W latach 1928-29 przygotował rozprawę doktorską opublikowaną w Polsce w 1930 roku. Zawiera ona m.in. dowód twierdzenia o dedukcji i konstruktywny dowód twierdzenia Gödla o pełności, otrzymany w oparciu o tzw. twierdzenie Herbranda, którego dowód zawiera algorytm semirozstrzygający, czy dane zdanie jest tautologią rachunku kwantyfikatorów. Wiosną 1931 roku Herbrand miał zasugerować Gödlowi rozszerzenie definicji klasy funkcji rekurencyjnych (pochodzącej z pracy Gödla) po to, by opisać pojęcie obliczalności. Latem 1931 roku, 27 sierpnia, zginął w Alpach podczas wspinaczki.

2 Twierdzenie Herbranda

2.1 Rodzaje formuł

Formuła jest tautologią, jeżeli jest spełniona w każdej strukturze przy każdym wartościowaniu. Oczywiście, zdanie jest tautologią wtedy i tylko wtedy, gdy jest spełnione w każdej strukturze.

Formuła jest sprzeczna, jeżeli nie jest spełniona przy żadnym wartościowaniu w żadnej strukturze. Dowodzi się korzystając z lematu 5.4 z części 2, że zdanie jest sprzeczne wtedy i tylko wtedy, gdy nie jest spełnione w żadnej strukturze.

Formuła jest spełnialna, jeżeli jest struktura, w której jest spełniona przy pewnym wartościowaniu. Podobnie jak wyżej dowodzi się, że zdanie jest spełnialne wtedy i tylko wtedy, gdy jest spełnione w pewnej strukturze.

Lemat 2.1 *Dla dowolnej formuły φ zachodzą następujące równoważności:*

- 1) φ jest tautologią wtedy i tylko wtedy, gdy $\neg\varphi$ jest sprzeczna,
- 2) φ jest spełnialna wtedy i tylko wtedy, gdy φ nie jest sprzeczna. $\square$

2.2 Skolemizacja

Będziemy teraz rozważać ustalony język pierwszego rzędu, pozwalający na zapisywanie dowolnych formuł z kwantyfikatorami. Każdej formule φ tego języka przypiszemy pewną formułę φ^* .

Formułę φ^* wyliczamy w następujący sposób:

- 1) sprowadzamy φ do preneksowej postaci normalnej, a więc znajdujemy równoważną jej formułę, która zaczyna się blokiem kwantyfikatorów, po którym znajduje się formuła bezkwantyfikatorowa (można dodatkowo sprowadzić tę formułę na przykład do koniunkcyjnej postaci normalnej),
- 2) domykamy postać normalną formuły φ (dopisując do niej kwantyfikatory ogólne po wszystkich zmiennych wolnych),
- 3) dopóki są w otrzymanej formule kwantyfikatory egzystencjalne i można jej nadać postać

$$\forall x_1 \dots \forall x_n \exists y \psi$$

zastępujemy ją formułą

$$\forall x_1 \dots \forall x_n \exists y \psi[y \leftarrow f(x_1, \dots, x_n)],$$

gdzie f jest nowym symbolem funkcyjnym. Analogiczną definicję przyjmujemy w przypadku, gdy $n = 0$, wtedy symbol f powinien być nową stałą.

- 4) usuwamy z otrzymanej w ten sposób formuły kwantyfikatory dopisane w punkcie 2). Otrzymaną w ten sposób formułę oznaczamy symbolem φ^* .

2.3 Własność skolemizacji

W logice klasę formuły definiujemy nieco inaczej niż w arytmetyce. Formuła jest klasy Π_1 (Σ_1), jeżeli zaczyna się blokiem kwantyfikatorów uniwersalnych (odpowiednio: egzystencjalnych), po której jest już formuła bez kwantyfikatorów.

Najpierw powinniśmy zauważyć, że zachodzi oczywisty

Lemat 2.2 *Formuła φ^* jest klasy Π_1 . $\square$*

Lemat 2.3 *Zdanie $\forall \vec{x} \exists y \varphi$ jest spełnione w strukturze $\mathcal{A}$ wtedy i tylko wtedy, gdy w pewnym rozszerzeniu $\mathcal{A}'$ struktury $\mathcal{A}$ o interpretację symbolu funkcyjnego f jest spełnione zdanie $\forall \vec{x} \varphi[y \leftarrow f(\vec{x})]$.*

Dowód. Zdanie $\forall \vec{x} \exists y \varphi$ jest logiczną konsekwencją zdania $\forall \vec{x} \varphi[y \leftarrow f(\vec{x})]$, jest więc spełnione w strukturze $\mathcal{A}'$, a także w strukturze powstającej przez usunięcie interpretacji zbędnych symboli.

Aby dowieść implikację w drugą stronę zauważmy, że strukturze można zdefiniować całkowitą funkcję F taką, że

$$\mathcal{A} \models \varphi[\vec{x} \leftarrow \vec{a}][y \leftarrow F(\vec{a})].$$

Łatwo sprawdzić posługując się lematem ?? (lemat 5.3 z części 2), że jeżeli F uznamy za interpretację symbolu f i dodamy do $\mathcal{A}$, to w otrzymanej strukturze $\mathcal{A}'$ będzie spełnione zdanie $\forall \vec{x} \varphi[y \leftarrow f(\vec{x})]$. $\square$

Możemy teraz wyprowadzić wersję lematu 2.3 dotyczącą operacji $*$:

Lemat 2.4 *Formuła φ jest spełniona w strukturze $\mathcal{A}$ wtedy i tylko wtedy, gdy w pewnym rozszerzeniu $\mathcal{A}'$ struktury $\mathcal{A}$ o interpretację dodanych symboli funkcyjnych jest spełniona formuła φ^* .*

Dowód. Konstrukcja φ^* polega utworzeniu pewnego ciągu formuł zaczynającego się φ i kończącego się φ^* . Dla każdej pary kolejnych formuł w tym ciągu, spełnianie w strukturze $\mathcal{A}$ pierwszej jest równoważne spełnianiu w pewnym rozszerzeniu $\mathcal{A}$ drugiej. W większości przypadków tak jest na podstawie lematu 2.3.

Pierwsze przekształcenie polega na zastąpieniu formuły φ jej postacią normalną ψ . Równoważność $\varphi \Leftrightarrow \psi$ jest prawem rachunku kwantyfikatorów, na mocy twierdzenia 5.6 o poprawności (część 2) jest spełniona w każdej strukturze. Stąd spełnianie w strukturze φ jest równoważne spełnianiu w niej ψ . W pozostałych przypadkach potrzebny fakt wynika z lematu 5.5 z części 2. $\square$

Z powyższego lematów wynika seria wniosków wyjaśniających sens logiczny operacji $*$. W szczególności

Wniosek 2.5 *Zdanie φ jest spełnialne wtedy i tylko wtedy, gdy spełnialne jest φ^* .*

Dowód. Jest to właściwie oczywista konsekwencja poprzedniego lematu. $\square$

Wniosek 2.6 *Zdanie φ jest sprzeczne wtedy i tylko wtedy, gdy φ^* jest sprzeczne. $\square$*

Wniosek 2.7 *Zdanie φ jest tautologią wtedy i tylko wtedy, gdy $(\neg\varphi)^*$ jest sprzeczna. $\square$*

Wniosek 2.8 *Zdanie φ jest tautologią wtedy i tylko wtedy, gdy $\neg(\neg\varphi)^*$ jest tautologią. $\square$*

Na koniec zauważmy, że formuła $\neg(\neg\varphi)^*$ jest z oczywistych powodów równoważna formule klasy Σ_1 . Tak więc badanie, czy φ jest tautologią, sprowadza się do badania, czy jest tautologią pewna, łatwa do zalgorytmizowanego wyliczenia formuła klasy Σ_1 .

2.4 Model Herbranda

Teraz musimy rozważyć język z przynajmniej jedną stałą. Jeżeli z jakiś powodów rozważamy język bez stałych, to dodajemy do niego pewną stałą, która ma jedynie spełnić podany wymóg formalny. Niech $\mathcal{T}_0$ oznacza zbiór termów stałych w tym języku. Będziemy rozważać modele dla rozważanego języka z uniwersum równym $\mathcal{T}_0$.

W takich modelach symbole funkcyjne możemy interpretować w specyficzny sposób. Przyjmujemy, że k -argumentowy symbol f oznacza funkcję $F : \mathcal{T}_0^k \rightarrow \mathcal{T}_0$ taką, że

$$F(t_1, \dots, t_k) = f(t_1, \dots, t_k),$$

czyli operację tworzenia termu z głównym symbolem f . Analogicznie, stałą c w takim uniwersum interpretujemy jako term c .

Modele o uniwersum $\mathcal{T}_0$, w których stałe i symbole funkcyjne interpretujemy w podany sposób, nazywamy modelami Herbranda.

Lemat 2.9 *W dowolnym modelu Herbranda $\mathcal{H}$ wartość termu t przy wartościowaniu h jest równa*

$$t[x \leftarrow h(x)][y \leftarrow h(y)] \dots$$

W szczególności, wartością termu stałego t przy dowolnym wartościowaniu w modelu Herbranda jest term t . Ponadto, dla dowolnej formuły φ mamy

$$\mathcal{H} \models \varphi[h] \Leftrightarrow \mathcal{H} \models \varphi[x \leftarrow h(x)][y \leftarrow h(y)] \dots \quad \square$$

Tak więc w modelach Herbranda liczenie wartości termu przy danym wartościowaniu można sprowadzić do liczenia wartości pewnego termu stałego. Podobnie spełnianie formuły przy danym wartościowaniu jest równoważne spełnianiu pewnej formuły stałej, czyli bez zmiennych wolnych.

2.5 Model Herbranda wyznaczony przez daną strukturę

Przypuśćmy, że $\mathcal{A}$ jest dowolną, ustaloną strukturą dla rozważanego języka. Symbolem val będziemy w tym rozdziale oznaczać funkcję przyporządkowującą termowi $t \in \mathcal{T}_0$ jego wartość w strukturze $\mathcal{A}$.

Lemat 2.10 *Jeżeli h jest wartościowaniem zmiennych w modelu Herbranda, to*

$$val(t[h]) = t[val \circ h]. \quad \square$$

Struktura $\mathcal{A}$ wyznacza pewien model Herbranda $\mathcal{H}_{\mathcal{A}}$. W modelu $\mathcal{H}_{\mathcal{A}}$ symbol relacyjny r jest interpretowany jako relacja R taka, że

$$\langle t_1, \dots, t_k \rangle \in R \Leftrightarrow \mathcal{A} \models r(t_1 \dots, t_k).$$

Z powyższej definicji otrzymuje, że dla specyficznych wartościowań zmiennych, mających postać $val \circ h$ zachodzi równoważność

$$\mathcal{A} \models r(t_1, \dots, t_n)[val \circ h] \Leftrightarrow \mathcal{H}_{\mathcal{A}} \models r(t_1, \dots, t_n)[h].$$

Mamy bowiem

$$\begin{aligned} \mathcal{A} \models r(t_1, \dots, t_n)[val \circ h] &\Leftrightarrow \langle t_1[val \circ h], \dots, t_n[val \circ h] \rangle \in R_{\mathcal{A}} \Leftrightarrow \\ &\Leftrightarrow \langle val(t_1[h]), \dots, val(t_n[h]) \rangle \in R_{\mathcal{A}} \Leftrightarrow \mathcal{A} \models r(t_1[h], \dots, t_n[h]) \Leftrightarrow \\ &\Leftrightarrow \langle t_1[h], \dots, t_n[h] \rangle \in R \Leftrightarrow \mathcal{H}_{\mathcal{A}} \models r(t_1[h], \dots, t_n[h]) \Leftrightarrow \mathcal{H}_{\mathcal{A}} \models r(t_1, \dots, t_n)[h]. \end{aligned}$$

Lemat 2.11 *Jeżeli φ jest formułą bez kwantyfikatorów, a h jest wartościowaniem zmiennych w strukturze $\mathcal{H}_A$ (w zbiorze $\mathcal{T}_0$), to następujące warunki są równoważne;*

- 1) $\mathcal{A} \models \varphi[val \circ h]$
- 2) $\mathcal{H}_A \models \varphi[h]$. $\square$

Wniosek 2.12 *Jeżeli φ jest formułą klasy Σ_1 oraz $\mathcal{H}_A \models \varphi$, to $\mathcal{A} \models \varphi$.*

Wniosek 2.13 *Zdanie φ klasy Σ_1 jest tautologią wtedy i tylko wtedy, gdy jest spełnione w każdym modelu Herbranda.*

Wniosek 2.14 *Zdanie $\neg\varphi^*$ jest tautologią wtedy i tylko wtedy, gdy zdanie $\neg\varphi^*$ jest spełnione w każdym modelu Herbranda.*

Twierdzenie 2.15 *Zdanie φ jest tautologią wtedy i tylko wtedy, gdy zdanie $\neg(\neg\varphi)^*$ jest spełnione w każdym modelu Herbranda.*

W tzw. teorii modeli pokazuje się, że dla zwykłych niesprzecznych teorii T każdy nieskończony zbiór można przekształcić w model teorii T . Jeżeli nawet utożsamimy izomorficzne modele teorii T , to jest ich więcej niż każda z góry zadana moc. Modeli Herbranda jest jednak albo skończenie wiele (w przypadku skończonego uniwersum), albo continuum. Twierdzenie 2.15 mówi więc, że badając, czy φ jest tautologią, można sprawdzić spełnianie prostszej formuły $\neg(\neg\varphi)^*$ w stosunkowo małej „liczbie” modeli, dla najwyżej continuum modeli.

2.6 Kolejny krok

Twierdzenie 2.16 *Niech ψ będzie formułą bez kwantyfikatorów. Zdanie $\varphi = \exists x_1 \dots \exists x_n \psi = \exists \vec{x} \psi$ jest spełnione w każdym modelu Herbranda wtedy i tylko wtedy, gdy istnieje skończenie wiele układów termów $\vec{t}_1, \dots, \vec{t}_m \in \mathcal{T}_0^n$ takich, że formuła*

$$\psi[\vec{x} \leftarrow \vec{t}_1] \vee \dots \vee \psi[\vec{x} \leftarrow \vec{t}_m]$$

jest tautologią w sensie rachunku zdań (różne formuły atomowe uważamy za różne zmienne zdaniowe).

Dowód. Zauważmy, że formuła φ wynika z każdego członu podanej alternatywy i z całej alternatywy. Jeżeli ta alternatywa jest tautologią, to jest spełniona w każdym modelu Herbranda, w każdym takim modelu są też spełnione wszystkie wnioski z tej alternatywy i ostatecznie, w każdym modelu Herbranda jest spełniona formuła φ . Dowód w drugą stronę jest trudniejszy.

Ustawmy w ciąg wszystkie zdania atomowe (czyli formuły atomowe bez zmiennych wolnych). Na ogół jest ich nieskończenie wiele, taki jest zresztą istotny przypadek. Niech σ_n oznacza n -te zdanie atomowe w utworzonym ciągu. Będziemy teraz rozważać skończone i nieskończone ciągi zero-jedynkowe.

Aby zdefiniować model Herbranda należy podać, które zdania atomowe są w nim spełnione. Mając więc nieskończony ciąg zero-jedynkowy $\alpha \in 2^{\mathbb{N}}$ możemy zdefiniować model Herbranda $\mathcal{H}_\alpha$ przyjmując, że

$$\mathcal{H}_\alpha \models \sigma_n \iff \alpha(n) = 1.$$

Zauważmy od razu, że w ten sposób można opisać wszystkie modele Herbranda.

Przy okazji zwróćmy też uwagę na to, że taki ciąg α może być dla nas wartościowaniem w sensie znanym z rachunku zdań. Faktycznie jest to wartościowanie zdań atomowych, ale także pewnych obiektów, które możemy uważać za zmienne zdaniowe i z których możemy budować formuły rachunku zdań. Mając takie wartościowanie możemy wyliczyć wartość logiczną dowolnego zdania zapisanego w rozważanym języku bez użycia kwantyfikatorów.

Każdemu $\alpha \in 2^N$ przypiszemy teraz pewną liczbę

$$m(\alpha) = \min\{k \in N : \exists \vec{t} \in \mathcal{T}_0^n (\mathcal{H}_\alpha \models \psi[\vec{x} \leftarrow \vec{t}] \wedge \bigwedge \forall l > k \text{ formuła } \sigma_l \text{ nie jest podformułą } \psi[\vec{x} \leftarrow \vec{t}])\}.$$

Z założenia, że w modelu $\mathcal{H}_\alpha$ jest spełnione zdanie φ wynika, że liczba $m(\alpha)$ jest dobrze zdefiniowana. Tak jest, ponieważ w modelach Herbranda każdy element jest wartością termu stałego.

Liczby $m(\alpha)$ mają interesujące własności. Na przykład, jeżeli $\alpha, \beta \in 2^N$ oraz $\alpha(i) = \beta(i)$ dla wszystkich $i \leq m(\alpha)$, to $m(\alpha) = m(\beta)$.

Teraz będziemy zajmować się wartością

$$\max\{m(\alpha) : \alpha \in 2^N\}$$

i – aby zbadać tę wartość – rozważać zbiór

$$X = \{\langle \alpha(0), \dots, \alpha(m(\alpha)) \rangle : \alpha \in 2^N\}$$

skończonych ciągów zero-jedynkowych.

Przypuśćmy, że wspomniane maksimum nie jest określone. Wtedy zbiór X jest nieskończony i zawiera dowolnie długie ciągi. W takiej sytuacji istnieje nieskończony ciąg γ , którego każdy skończony ciąg początkowy jest prefiksem pewnego elementu zbioru X . Takie γ konstruujemy przez indukcję. Przypuśćmy, że zdefiniowaliśmy już $\gamma(0), \gamma(1), \dots, \gamma(n-1)$ i ciąg ten jest taki, że

$$X_n = \{\langle x_0, x_1, \dots, x_{m-1} \rangle \in X : n < m \wedge \forall i < n \gamma(i) = x_i\}$$

jest nieskończony. Oczywiście,

$$X_n = \{\langle x_0, x_1, \dots, x_{m-1} \rangle \in X : n < m \wedge x_n = 0 \wedge \forall i < n \gamma(i) = x_i\} \cup \{\langle x_0, x_1, \dots, x_{m-1} \rangle \in X : n < m \wedge x_n = 1 \wedge \forall i < n \gamma(i) = x_i\}.$$

Jeden ze składników tej sumy jest nieskończony. Jeżeli jest to pierwszy składnik, to przyjmujemy, że $\gamma(n) = 0$. Zbiór X_{n+1} jest nieskończony, gdyż różni się od pierwszego składnika najwyżej jednym elementem $\langle \gamma(0), \gamma(1), \dots, \gamma(n-1), 0 \rangle$. Analogicznie definiujemy $\gamma(n)$ w drugim z możliwych przypadków.

Każdy prefiks tak zdefiniowanego ciągu γ daje się wydłużyć do elementu X (nawet na nieskończenie wiele sposobów).

Także ciąg $\langle \gamma(0), \dots, \gamma(m(\gamma)) \rangle$ można więc wydłużyć do $\langle \beta(0), \dots, \beta(m(\beta)) \rangle$ dla pewnego β takiego, że $m(\gamma) < m(\beta)$. Ze wspomnianej wyżej własności wynika jednak, że $m(\gamma) = m(\beta)$. Uzyskana sprzeczność dowodzi, że wspomniane maksimum jest liczbą naturalną, a zbiór X jest skończony.

Teraz, dla każdego elementu $s \in X$ bierzemy ciąg α_s taki, że $s = \langle \alpha_s(0), \dots, \alpha_s(m(\alpha_s)) \rangle$ oraz termy $\vec{t}_s$, dla których

$$\mathcal{H}_{\alpha_s} \models \psi[\vec{x} \leftarrow \vec{t}_s].$$

Nietrudno zauważyć, że alternatywa

$$\bigvee_{s \in X} \psi[\vec{x} \leftarrow \vec{t}_s]$$

jest szukaną tautologią rachunku zdań. $\square$

Twierdzenie 2.17 (Herbrand) *Przypuśćmy, że mamy zdanie φ i formuła $\neg(\neg\varphi)^*$ jest równoważna $\exists x_1 \dots x_n \psi$, gdzie ψ jest formułą bez kwantyfikatorów. Zdanie φ jest tautologią wtedy i tylko wtedy, gdy istnieje skończony ciąg układów termów $\vec{t}_1, \dots, \vec{t}_m \in \mathcal{T}_0^n$ taki, że formuła*

$$\psi[\vec{x} \leftarrow \vec{t}_1] \vee \dots \vee \psi[\vec{x} \leftarrow \vec{t}_m]$$

jest tautologią. Dokładniej, jest ona prawdziwa bez względu na to, jakie są wartości logiczne formuł atomowych z tej formuły.

Dowód. Jest to wniosek z twierdzeń 2.15 i 2.16. $\square$

2.7 Implementacje algorytmu Herbranda

2.7.1 Próba Paula Gilmora

Twierdzenie Herbranda podaje zalgorytmizowany sposób badania, czy dana formuła jest prawem rachunku kwantyfikatorów, określaną jako algorytm Herbranda. Wobec twierdzenia o dedukcji może być też podstawą algorytmu badającego, czy daną formułę można dowieść w skończenie aksjomatyzowalnej teorii, np. w teorii Q Robinsona. Oczywiście, w obu wspomnianych przypadkach otrzymujemy algorytmy semirozstrzegające.

Pierwszą próbę implementacji algorytmu Herbranda podjął w 1960 roku Paul Gilmore z IBM. Próba ta została uznana za nieudaną. Gilmore wykorzystywał procedurę badania tautologii opartą o następujący

Lemat 2.18 *Formuła rachunku zdań φ w koniunkcyjnej postaci normalnej jest tautologią wtedy i tylko wtedy, gdy każdy występujący w niej człon koniunkcji jest alternatywą postaci*

$$\dots \vee p \vee \dots \vee \neg p \vee \dots$$

dla pewnej zmiennej zdaniowej p (ewentualnie dla negacji p). $\square$

2.7.2 Metoda Martina Davisa i Hilarego Putnama

Twierdzeniu 2.17 można też nadać postać

Twierdzenie 2.19 (Herbrand) *Przypuśćmy, że mamy zdanie φ i formuła $\neg(\neg\varphi)^*$ jest równoważna $\exists x_1 \dots x_n \neg\delta$, gdzie δ jest formułą bez kwantyfikatorów, najlepiej w koniunkcyjnej postaci normalnej. Zdanie φ jest tautologią wtedy i tylko wtedy, gdy istnieje skończony ciąg układów termów $\vec{t}_1, \dots, \vec{t}_m \in \mathcal{T}_0^n$ taki, że formuła*

$$\delta[\vec{x} \leftarrow \vec{t}_1] \wedge \dots \wedge \delta[\vec{x} \leftarrow \vec{t}_m]$$

jest sprzeczną formułą rachunku zdań. Dokładniej, okaże się ona spreczna, jeżeli różne jej formuły atomowe będziemy uważać za różne zmienne zdaniowe.

Dowód. Jest to oczywiste przeformułowanie twierdzenia 2.17. $\square$

Metodę badania tautologii lepiej dostosowaną do algorytmu Herbranda i opartego na wyżej przytoczonej wersji twierdzenia Herbranda zaproponowali w 1960 roku Martin Davis i Hilary Putnam. Główną jej częścią jest algorytm

$\Phi := \varepsilon$;

for all układów termów $\vec{t}$

$\Phi := \Phi \wedge \delta[\vec{x} \leftarrow \vec{t}]$;

if Φ jest spreczna, **then return** φ jest tautologią.

2.7.3 Badanie sprzeczności według Davisa i Putnama

Głównym osiągnięciem Davisa i Putnama było zwrócenie uwagi na regułę rezolucji i opracowanie opartego o tę regułę algorytmu badania sprzeczności formuł rachunku zdań w koniunkcyjnej postaci normalnej.

Koniunkcje i alternatywy w tym algorytmie są implementowane jako zbiory członów. Tak więc sposób implementacji formuł gwarantuje przemienność i łączność obu spójników. Przy takiej implementacji dobrze jest założyć, że koniunkcje i alternatywy rozumiane jako zbiory są niepuste. Taka implementacja zwykłych formuł prowadzi do zbiorów niepustych. Dalej zakładamy, że zbiór pusty nie jest ani koniunkcją, ani alternatywą.

Danymi dla algorytmu jest zbiór alternatyw literalów, czyli zbiorów zmiennych zdaniowych i ich negacji. Interesuje nas, czy jest to zbiór sprzeczny lub czy dana koniunkcja jest sprzeczna.

Dany zbiór alternatyw podczas badania sprzeczności jest upraszczany na kilka sposobów, w tym oparte o następujące lematy i podobne. W skrócie, stosujemy trzy metody przekształcania danych: usuwanie tautologii, usuwanie pojedynczych literalów i rezolucję.

Lemat 2.20 *Jeżeli formuła τ jest tautologią, to formuła $\varphi \wedge \tau$ jest sprzeczna wtedy i tylko wtedy, gdy sprzeczna jest formuła φ . $\square$*

Zgodnie z tym lematem, dany zbiór alternatyw, którego sprzeczność jest badana można uprościć usuwając z niego alternatywy, w których pewna zmienna występuje z negacją i bez. Sprzeczność obydwu takich zbiorów jest równoważna.

Lemat 2.21 *Jeżeli zmienna zdaniowa p nie występuje w ψ , to formuła $(p \vee \varphi) \wedge \psi$ jest sprzeczna wtedy i tylko wtedy, gdy ψ jest sprzeczna. $\square$*

Podany algorytm pozwala w pewnych przypadkach zastąpić badanie sprzeczności zbioru alternatyw badaniem innego zbioru, złożonego z alternatyw, w których nie występuje określony literal p . Można tak postąpić, jeżeli w danym zbiorze nie występuje $\neg p$.

Lemat 2.22 *Jeżeli zmienna zdaniowa p nie występuje w ψ , to formuła $p \wedge (p \vee \varphi) \wedge (\neg p \vee \psi)$ jest sprzeczna wtedy i tylko wtedy, gdy ψ jest sprzeczna. $\square$*

Jeżeli w danym zbiorze występuje alternatywa, której jedynym członem jest literal p , to można taki zbiór uprościć usuwając z niego wszystkie alternatywy, w których występuje literal i usuwając z pozostałych alternatyw negację tego literalu.

Główny sposób przekształcania danego zbioru odpowiada wnioskowaniu za pomocą reguły rezolucji. Badanie metodą Davisa - Putnama, czy formuła jest tautologią, może być rozumiane jako dowodzenie tego faktu metodą nie wprost stosując tylko regułę rezolucji

$$\frac{\delta \vee \varphi, \quad \neg \delta \vee \psi}{\varphi \vee \psi},$$

choć w specyficznej sytuacji i ze wspomaganie potrzebnym po to, by bez problemów zatrzymać algorytm. Ten główny sposób przekształcania jest oparty o

Lemat 2.23 *Założmy, że zmienna zdaniowa p nie występuje w formułach φ , ψ oraz δ . Wtedy formuła $(p \vee \varphi) \wedge (\neg p \vee \psi) \wedge \delta$ jest sprzeczna wtedy i tylko wtedy, gdy sprzeczna jest formuła $(\varphi \vee \psi) \wedge \delta$. $\square$*

2.7.4 Metoda Johna Alana Robinsona

Algorytm Herbranda korzystający z metody Davisa i Putnama polega szukaniu dwóch rzeczy: odpowiednich podstawień i dowodu sprzeczności. W 1965 roku ukała się praca Johna Robinsona, w której zauważył, że za jednym zamachem możemy szukać podstawień i dowodu sprzeczności. Metoda Robinsona oparta jest o taką oto wersję wniosku 2.6:

Wniosek 2.24 *Zdanie φ jest tautologią wtedy i tylko wtedy, gdy zdanie $(\neg\varphi)^*$ jest sprzeczne, lub równoważnie, gdy daje się z niego wyprowadzić sprzeczność. $\square$*

Natomiast badanie sprzeczności w tej metodzie jest oparte o regułę rezolucji bardziej ogólną niż stosowana przez Davisa.

Podstawianie i unifikatory. Najpierw musimy uogólnić operację podstawiania $\varphi[x \leftarrow t]$. Zwykle podstawiamy za jedną zmienną, mogliśmy też iterować podstawianie i podstawiać po kolei za kilka zmiennych. Teraz przez podstawianie będziemy rozumieć zbiór par złożonych ze zmiennej i termu, i będziemy stosować następującą notację

$$\theta = [x_1 \leftarrow t_1, \dots, x_n \leftarrow t_n],$$

gdzie x_i są zmiennymi, a t_i – termami. Wykonując podstawienie θ jednocześnie za zmienne x_i podstawiamy termu t_i . Definiując takie podstawianie zmieniamy sposób podstawiania za zmienne przyjmując

$$x_i\theta = x_i[x_1 \leftarrow t_1, \dots, x_n \leftarrow t_n] = t_i$$

oraz

$$y\theta = y[x_1 \leftarrow t_1, \dots, x_n \leftarrow t_n] = y$$

dla zmiennych y różnych od $x_1, \dots, x_n$.

Podstawienie $\theta = [x_1 \leftarrow y_1, \dots, x_n \leftarrow y_n]$ nazywamy przemianowaniem zmiennych, jeżeli $y_1, \dots, y_n$ są zmiennymi i zależność między x_i oraz y_i jest różnowartościowa.

Podstawienie θ jest unifikatorem termów t_1 i t_2 , jeżeli $t_1\theta = t_2\theta$. Analogicznie definiujemy unifikator formuł bez kwantyfikatorów (takie formuły można uważać za termu, w których występują także funkcje o wartościach boolowskich).

Podstawienie θ_0 jest najogólniejszym unifikatorem termów t_1 i t_2 , jeżeli jest unifikatorem tych termów i dla każdego innego unifikatora θ istnieje podstawienie θ' takie, że $t_1\theta = t_1\theta_0\theta'$ oraz $t_2\theta = t_2\theta_0\theta'$.

Reguła rezolucji wg Robinsona. Robinson zauważył, że do wyprowadzania sprzeczności w tym kontekście przydatna jest następująca reguła rezolucji

$$\frac{\delta_1 \vee \varphi, \neg\delta_2 \vee \psi}{\varphi\theta \vee \psi\theta},$$

gdzie θ jest najogólniejszym unifikatorem formuł (atomowych) δ_1 i δ_2 , rozumiana podobnie, jak w Prologu. W szczególności dopuszczamy stosowanie tej reguły w sytuacjach, gdy podane przesłanki (jedna lub obie) redukują się do formuł δ_1 i $\neg\delta_2$.

Przytoczona reguła rezolucji rozumiana literalnie nie wystarcza do wyprowadzania sprzeczności ze sprzecznych zbiorów formuł. Świadczy o tym następujący przykład: za pomocą tej reguły nie wyprowadzimy sprzeczności z formuł $A(x)$,

$\neg B(x)$ oraz $\neg A(x) \vee B(f(x))$. Możemy z nich wyprowadzić tylko $B(f(x))$. Z formuł $\neg B(x)$ i $B(f(x))$ nic nie wyprowadzimy, ponieważ napisy $B(x)$ i $B(f(x))$ nie dają się zunifikować. Mimo to formuła $\forall x (A(x) \wedge \neg B(x) \wedge (\neg A(x) \vee B(f(x))))$ jest sprzeczna.

Wobec tego dopuszczamy jeszcze możliwość stosowania następującej reguły wnioskowania:

$$\frac{\varphi}{\varphi\theta}$$

dla dowolnego przemianowania zmiennych θ . Zwykle obie podane tutaj reguły stosujemy w powiązany sposób: najpierw w formułach, które mają stać się przesłankami reguły rezolucji przemianowujemy zmienne, a następnie stosujemy rezolucję.

W przytoczonym przykładzie łatwo uzyskać sprzeczność po przemianowaniu w formule $\neg B(x)$ zmiennej x na y . Najogólniejszym unifikatorem $B(y)$ i $B(f(x))$ jest podstawienie $[y \leftarrow f(x)]$. Po wykonaniu tego podstawienia formuły $\neg B(y)$ i $B(f(x))$ stają się sprzeczne.

Zauważmy, że obie reguły w następującym sensie zachowują niesprzeczność:

Lemat 2.25 *Przyjmijmy, że $\mathcal{A}$ jest dowolną strukturą, φ – formuła bez kwantyfikatorów, a θ – dowolnym podstawieniem. Wtedy warunek $\mathcal{A} \models \varphi$ implikuje $\mathcal{A} \models \varphi\theta$. Jeżeli natomiast θ jest przemianowaniem zmiennych, to warunek $\mathcal{A} \models \varphi$ jest równoważny z $\mathcal{A} \models \varphi\theta$.*

Dowód. Trzeba pamiętać, że spełnianie formuły oznacza jej spełnianie przy każdym wartościowaniu zmiennych. $\square$

Lemat 2.26 *Przyjmijmy, że w strukturze $\mathcal{A}$ spełnione są formuły $\delta_1 \vee \varphi$ oraz $\neg\delta_2 \vee \psi$ i θ jest unifikatorem formuł δ_1 i δ_2 . Wtedy $\mathcal{A} \models (\varphi \vee \psi)\theta$. $\square$*

Wniosek 2.27 *Jeżeli $\psi_1, \dots, \psi_n$ są alternatywami literalów i posługując się podanymi regułami potrafimy wyprowadzić z nich sprzeczność, to formuła*

$$\forall \vec{x} \bigwedge_{i=1}^n \psi_i$$

jest sprzeczna. $\square$

Twierdzenie o pełności dla metody rezolucji. Aby uzasadnić korzystanie z metody rezolucji Robinson udowodnił

Twierdzenie 2.28 *Przypuśćmy, że mamy formułę $\forall \vec{x} (\psi_1 \wedge \dots \wedge \psi_m)$ klasy Π_1 w koniunkcyjnej postaci normalnej. Formuła ta jest sprzeczna wtedy i tylko wtedy, gdy z alternatyw literalów $\psi_1, \dots, \psi_m$ daje się wyprowadzić sprzeczność (czyli alternatywę bez członów) stosując jedynie przemianowanie zmiennych i regułę rezolucji.*

Dowód. Wobec wniosku 2.27 wystarczy pokazać, jak z formuły sprzecznej można wyprowadzić sprzeczność. W tym celu będziemy korzystać z aparatu i oznaczeń wprowadzonych w dowodzie twierdzenia 2.16 Herbranda.

Najpierw przyjmijmy, że $\sigma^0 = \neg\sigma$ oraz $\sigma^1 = \sigma$. To oznaczenie pozwala inaczej sformułować definicję modeli $\mathcal{H}_\alpha$ są to modele takie, że

$$\mathcal{H}_\alpha \models \bigwedge_{i \leq n} \sigma_i^{\alpha(i)}$$

dla wszystkich $n \in N$. Teraz zmodyfikujemy nieco definicję $m(\alpha)$. Założenie o sprzeczności danej formuły oznacza, że w każdym modelu Herbranda $\mathcal{H}_\alpha$ spełniona jest formuła

$$\exists \vec{x} (\neg\psi_1 \vee \dots \vee \neg\psi_m),$$

a więc także pewna formuła postaci $\neg\psi_i[\vec{x} \leftarrow \vec{t}]$. Przyjmijmy, że

$$m(\alpha) = \min\{k \in N : \exists j \leq m \exists \vec{t} \in \mathcal{T}_0 \text{ t.ż. } \bigwedge_{i \leq k} \sigma_i^{\alpha(i)} \Rightarrow \neg\psi_j[\vec{x} \leftarrow \vec{t}] \text{ jest tautologią}\}.$$

Liczby $m(\alpha)$ są dobrze zdefiniowane i mają własność podaną w dowodzie twierdzenia 2.16: jeżeli $\alpha, \beta \in 2^N$ oraz $\alpha(i) = \beta(i)$ dla wszystkich $i \leq m(\alpha)$, to $m(\alpha) = m(\beta)$. Dlatego też można powtórzyć rozumowanie z dowodu tego twierdzenia i wykazać, że zbiór

$$X = \{\langle \alpha(0), \dots, \alpha(m(\alpha)) \rangle : \alpha \in 2^N\}$$

jest skończony.

Zbiór X może być uważany za zbiór wszystkich maksymalnych gałęzi pewnego drzewa binarnego i może być z takim drzewem utożsamianym. To drzewo pozwala odtworzyć interesujące nas wyprowadzenie sprzeczności.

Niech $\langle \alpha(0), \dots, \alpha(m(\alpha)) \rangle$ będzie najdłuższym elementem zbioru X . Dla ustalenia uwagi przyjmijmy, że $\alpha(m(\alpha)) = 0$. Weźmy ciąg $\beta \in 2^N$ taki, że $\beta(i) = \alpha(i)$ dla $i < m(\alpha)$ oraz $\beta(m(\alpha)) = 1$. Z wyboru α wynika, że $m(\beta) \leq m(\alpha)$, ostra nierówność prowadzi do sprzeczności z przytoczoną własnością przyporządkowania m . Stąd $m(\alpha) = m(\beta)$ i w zbiorze X mamy dwa ciągi

$$\langle \alpha(0), \dots, \alpha(m(\alpha) - 1), 0 \rangle = \langle \alpha(0), \dots, \alpha(m(\alpha)) \rangle$$

oraz

$$\langle \alpha(0), \dots, \alpha(m(\alpha) - 1), 1 \rangle = \langle \beta(0), \dots, \beta(m(\beta)) \rangle.$$

Z należenia do X pierwszego z ciągów wynika, że dla pewnych j i $\vec{t}_0$ formuła

$$\left(\bigwedge_{i \leq m(\alpha)} \sigma_i^{\alpha(i)} \right) \Rightarrow \neg\psi_j[\vec{x} \leftarrow \vec{t}_0]$$

jest tautologią. Po uwzględnieniu, że $\psi_j[\vec{x} \leftarrow \vec{t}_0]$ jest alternatywą $\bigvee_v C_{j,v}$ literałów $C_{j,v}$ formułę tę można inaczej zapisać, w postaci

$$\bigwedge_v (\neg\sigma_0^{\alpha(0)} \vee \dots \vee \neg\sigma_{m(\alpha)-1}^{\alpha(m(\alpha)-1)} \vee \neg\sigma_{m(\alpha)}^0 \vee \neg C_{j,v}). \quad (1)$$

Ponieważ przytoczona formuła jest tautologią i w ciągu $\sigma_0, \sigma_1, \dots$ są same formuły atomowe, więc każdy literał $C_{j,v}$ jest równoważny $\neg\sigma_l^{\alpha(l)}$ dla pewnego $l \leq m(\alpha)$.

Z drugiej strony, na mocy definicji $m(\alpha)$, formuła

$$\left(\bigwedge_{i \leq m(\alpha)-1} \sigma_i^{\alpha(i)} \right) \Rightarrow \neg\psi_j[\vec{x} \leftarrow \vec{t}], \quad \text{czyli} \quad \bigwedge_v (\neg\sigma_0^{\alpha(0)} \vee \dots \vee \neg\sigma_{m(\alpha)-1}^{\alpha(m(\alpha)-1)} \vee \neg C_{j,v})$$

nie jest tautologią. Pewien literał $C_{j,v}$ nie jest więc postaci $\neg\sigma_l^{\alpha(l)}$ dla $l < m(\alpha)$. Nietrudno zauważyć, że ten literał jest równy $\sigma_{m(\alpha)}$. Spostrzeżenie to pozwala przedstawić formułę $\psi_j[\vec{x} \leftarrow \vec{t}_0]$ w postaci

$$\psi_j[\vec{x} \leftarrow \vec{t}_0] = \sigma_{\alpha(m)} \vee \varphi'_0 = \rho_0[\vec{x} \leftarrow \vec{t}_0] \vee \varphi_0[\vec{x} \leftarrow \vec{t}_0] \quad (2)$$

dla pewnej atomowej formuły ρ_0 i pewnej klauzuli φ_0 (alternatywy literałów z ψ_j różnych od ρ_0).

Na koniec zauważmy, że jeżeli w koniunkcji (1) opuścimy człon odpowiadający literałowi $\sigma_{m(\alpha)}$, to otrzymamy tautologię postaci

$$\left(\bigwedge_{i \leq m(\alpha)-1} \sigma_i^{\alpha(i)} \right) \Rightarrow \neg \varphi_0[\vec{x} \leftarrow \vec{t}_0]. \quad (3)$$

To samo rozumowanie możemy przeprowadzić dla należącego do X ciągu $\langle \alpha(0), \dots, \alpha(m(\alpha) - 1), 1 \rangle$. Przyjmijmy, że w tym przypadku dla pewnych j' i $\vec{t}_1$ tautologią okazuje się formuła

$$\left(\bigwedge_{i \leq m(\alpha)} \sigma_i^{\alpha(i)} \right) \Rightarrow \neg \psi_{j'}[\vec{x} \leftarrow \vec{t}_1].$$

W tym przypadku formuła $\psi_{j'}[\vec{x} \leftarrow \vec{t}_1]$ daje się przedstawić w postaci

$$\psi_{j'}[\vec{x} \leftarrow \vec{t}_1] = \neg \sigma_{\alpha(m)} \vee \varphi'_1 = \neg \rho_1[\vec{x} \leftarrow \vec{t}_1] \vee \varphi_1[\vec{x} \leftarrow \vec{t}_1] \quad (4)$$

dla pewnej atomowej formuły ρ_1 i pewnej klauzuli φ_1 (alternatywy literałów z $\psi_{j'}$ różnych od ρ_1). Podobnie, jak poprzednio, pokazujemy także, że tautologią jest formuła

$$\left(\bigwedge_{i \leq m(\alpha)-1} \sigma_i^{\alpha(i)} \right) \Rightarrow \neg \varphi_1[\vec{x} \leftarrow \vec{t}_1]. \quad (5)$$

Teraz możemy już zacząć konstruować rezolucyjny dowód sprzeczności klauzul $\psi_1, \dots, \psi_m$. Z tych klauzul wybieramy dwie: ψ_j oraz $\psi_{j'}$. Najpierw urozłączniamy zmienne zastępując $\psi_{j'}$ przez $\psi_{j'}[\vec{x} \leftarrow \vec{y}]$. Zgodnie z wzorami (2) i (4) mamy

$$\psi_j = \rho_0 \vee \varphi_0 \text{ oraz } \psi_{j'}[\vec{x} \leftarrow \vec{y}] = \neg \rho_1[\vec{x} \leftarrow \vec{y}] \vee \varphi_1[\vec{x} \leftarrow \vec{y}].$$

Weźmy podstawienie $\theta = [\vec{x} \leftarrow \vec{t}_0, \vec{y} \leftarrow \vec{t}_1]$. To podstawienie unifikuje formuły ρ_0 oraz $\rho_1[\vec{x} \leftarrow \vec{y}]$. Mamy bowiem

$$\rho_0 \theta = \rho_0[\vec{x} \leftarrow \vec{t}_0, \vec{y} \leftarrow \vec{t}_1] = \rho_0[\vec{x} \leftarrow \vec{t}_0] = \sigma_{\alpha(m)}$$

oraz

$$\rho_1[\vec{x} \leftarrow \vec{y}] \theta = \rho_1[\vec{x} \leftarrow \vec{y}][\vec{x} \leftarrow \vec{t}_0, \vec{y} \leftarrow \vec{t}_1] = \rho_1[\vec{x} \leftarrow \vec{y}][\vec{y} \leftarrow \vec{t}_1] = \rho_1[\vec{x} \leftarrow \vec{t}_1] = \sigma_{\alpha(m)}.$$

Niech więc θ_0 będzie najogólniejszym unifikatorem formuł ρ_0 i $\rho_1[\vec{x} \leftarrow \vec{y}]$, a θ' – takim podstawieniem, że

$$\psi_j \theta = \psi_j \theta_0 \theta' \text{ oraz } \psi_{j'}[\vec{x} \leftarrow \vec{y}] \theta = \psi_{j'}[\vec{x} \leftarrow \vec{y}] \theta_0 \theta'. \quad (6)$$

Reguła rezolucji pozwala z formuł ψ_j przez $\psi_{j'}[\vec{x} \leftarrow \vec{y}]$ wywnioskować formułę

$$\varphi_0 \theta_0 \vee \varphi_1[\vec{x} \leftarrow \vec{y}] \theta_0.$$

Dołączmy tę formułę do klauzul $\psi_1, \dots, \psi_m$. Gdyby z tego zbioru udało nam się wyprowadzić sprzeczność, to potrafilibyśmy ją wyprowadzić także ze zbioru $\psi_1, \dots, \psi_m$.

Dla nowego zbioru klauzul spróbujemy wyliczyć funkcję m' zdefiniowaną tak, jak funkcja m dla zbioru $\psi_1, \dots, \psi_m$. Dla dowolnego $\gamma \in 2^N$ wartości $m(\gamma)$ i

$m'(\gamma)$ zależą od możliwości wskazywania klauzul i konstruowania odpowiednich podstawień. Jeżeli definiując $m(\gamma)$ rozpatrywaliśmy klauzulę ψ_j z podstawieniem termów $\vec{t}$, to nadal możemy to robić. Wynika stąd, że nowe możliwości mogą doprowadzić najwyżej do zmniejszenia wartości $m(\gamma)$, a więc zachodzą nierówności $m'(\gamma) \leq m(\gamma)$. Jeżeli X jest zbiorem maksymalnych gałęzi pewnego drzewa binarnego T , to zbiór

$$X' = \{ \langle \alpha(0), \dots, \alpha(m'(\alpha)) \rangle : \alpha \in 2^N \}$$

opisuje drzewo, którego maksymalne gałęzie uległy pewnemu skróceniu, a więc poddrzewo drzewa T . Pokażemy teraz, że przejście do nowego drzewa powoduje skrócenie przynajmniej jednej gałęzi.

Przypomnijmy, że rozpoczęliśmy konstrukcję od wyboru ciągu $\alpha \in 2^N$ wyznaczającego maksymalną gałąź $\langle \alpha(0), \dots, \alpha(m(\alpha)) \rangle \in X$. Otóż dla tego ciągu α zachodzi nierówność $m'(\alpha) < m(\alpha)$. Aby się o tym przekonać, zgodnie z definicją $m'(\alpha)$, wystarczy wskazać odpowiednią klauzulę i podstawienie na przykład dla $k = m(\alpha) - 1$. Tą odpowiednią jest dołączona klauzula

$$\varphi_0 \theta_0 \vee \varphi_1 [\vec{x} \leftarrow \vec{y}] \theta_0,$$

a szukanym podstawieniem jest θ' . Interesująca nas nierówność wynika z tego, że formuła

$$\left(\bigwedge_{i \leq m(\alpha)-1} \sigma_i^{\alpha(i)} \right) \Rightarrow \neg(\varphi_0 \theta_0 \vee \varphi_1 [\vec{x} \leftarrow \vec{y}] \theta_0) \theta' \quad (7)$$

jest tautologią. Z wzorów (6) i definicji podstawienia θ wynika, że jest to formuła

$$\left(\bigwedge_{i \leq m(\alpha)-1} \sigma_i^{\alpha(i)} \right) \Rightarrow \neg(\varphi_0 [\vec{x} \leftarrow \vec{t}_0] \vee \varphi_1 [\vec{x} \leftarrow \vec{t}_1]).$$

Jest ona równoważna koniunkcji

$$\left(\left(\bigwedge_{i \leq m(\alpha)-1} \sigma_i^{\alpha(i)} \right) \Rightarrow \neg \varphi_0 [\vec{x} \leftarrow \vec{t}_0] \right) \wedge \left(\left(\bigwedge_{i \leq m(\alpha)-1} \sigma_i^{\alpha(i)} \right) \Rightarrow \neg \varphi_1 [\vec{x} \leftarrow \vec{t}_1] \right)$$

formuł (3) i (5), które są tautologiami. Stąd również formuła (7) jest tautologią.

Tak więc drzewo wyznaczone przez zbiór X' okazało się mniejsze od drzewa opisanego przez zbiór X . Jeżeli będziemy dowód prowadzić przez indukcję ze względu na wielkość takich drzew, to z założenia indukcyjnego otrzymamy wyprowadzenie sprzeczności z klauzul $\psi_1, \dots, \psi_m$ uzupełnionych o

$$\varphi_0 \theta_0 \vee \varphi_1 [\vec{x} \leftarrow \vec{y}] \theta_0.$$

Wyprowadzenie sprzeczności z $\psi_1, \dots, \psi_m$ nie stanowi problemu, ponieważ dodatkową klauzulę otrzymujemy z nich za pomocą reguły rezolucji.

Teraz pozostaje najwyżej sprawdzenie pierwszego kroku indukcyjnego. $\square$

2.8 Przykład

Przykład ten świadczy m. in. o tym, że dowodzenie twierdzeń prowadzi do zbiorów klauzul, które nie tworzą poprawnego programu prologowego.

Zauważmy najpierw, że zachodzi

Lemat 2.29 *Jeżeli φ i ψ są zdaniami, to koniunkcja $\varphi \wedge \psi$ jest sprzeczna wtedy i tylko wtedy, gdy sprzeczna jest koniunkcja $\varphi^* \wedge \psi^*$. $\square$*

Przykład 2.30 Za pomocą tego lematu i metody rezolucji pokażemy, że ze zdań

$$\varphi_1 = \forall x (E(x) \wedge \neg V(x) \Rightarrow \exists y (S(x, y) \wedge C(y))),$$

$$\varphi_2 = \exists x (P(x) \wedge E(x) \wedge \forall y (S(x, y) \Rightarrow P(y))),$$

$$\varphi_3 = \forall x (P(x) \Rightarrow \neg V(x))$$

wynika zdanie

$$\psi = \exists x (P(x) \wedge C(x)).$$

Na mocy twierdzenia o dedukcji mamy więc dowieść, że implikacja

$$\varphi_1 \wedge \varphi_2 \wedge \varphi_3 \Rightarrow \psi$$

jest tautologią. Jest to równoważne stwierdzeniu, że formuła

$$\varphi_1 \wedge \varphi_2 \wedge \varphi_3 \wedge \neg \psi$$

jest sprzeczna. Z przytoczonego lematu i wniosku 2.6 wynika, że wystarczy dowieść sprzeczność zdania

$$(\varphi_1)^* \wedge (\varphi_2)^* \wedge (\varphi_3)^* \wedge (\neg \psi)^*.$$

Formuła ta po wyciągnięciu na początek i usunięciu kwantyfikatorów, oraz po sprowadzeniu do postaci koniunkcyjnej jest koniunkcją następujących alternatyw:

$$\neg E(x) \vee V(x) \vee S(x, f(x)), \quad (1)$$

$$\neg E(x) \vee V(x) \vee C(f(x)), \quad (2)$$

$$P(a), \quad (3)$$

$$E(a), \quad (4)$$

$$\neg S(a, y) \vee P(y), \quad (5)$$

$$\neg P(x) \vee \neg V(x), \quad (6)$$

$$\neg P(x) \vee \neg C(x). \quad (7)$$

Zauważmy, że stosując regułę rezolucji możemy z tych alternatyw wyprowadzić

$$\neg V(a), \quad (8), \text{ z (3) i (6)}$$

$$V(a) \vee C(f(a)), \quad (9) \text{ z (2) i (4)}$$

$$C(f(a)), \quad (10) \text{ z (8) i (9)}$$

$$V(a) \vee S(a, f(a)), \quad (11) \text{ z (1) i (4)}$$

$$S(a, f(a)), \quad (12) \text{ z (8) i (11)}$$

$$P(f(a)), \quad (13) \text{ z (5) i (12)}$$

$$\neg C(f(a)), \quad (14) \text{ z (7) i (13)}$$

$$\square, \quad (15) \text{ z (10) i (14)}$$

2.9 Twierdzenie o pełności wg Herbranda

Wniosek 2.31 *Jeżeli formuła φ jest tautologią, to formuła $\neg(\neg\varphi)^*$ jest twierdzeniem rachunku kwantyfikatorów.*

Dowód. Jest to wniosek z twierdzenia Herbranda 2.17 i twierdzenia o pełności dla rachunku zdań. $\square$

Mając dowód formuły $\neg(\neg\varphi)^*$ możemy spróbować odtworzyć także dowód formuły φ . Gdyby to się udało, to można by dowieść twierdzenie o pełności. Dowód tego twierdzenia musi być jednak bardziej subtelny. Będzie w nim potrzebne

Twierdzenie 2.32 (o stałej) *Jeżeli stała c nie występuje aksjomatach teorii T , ani w formule φ oraz $T \vdash \varphi[x \leftarrow c]$, to $T \vdash \forall x \varphi$.*

Dowód. Wystarczy zauważyć, że wszędzie tam, gdzie w dowodzie mamy prawo użyć stałej c , możemy też użyć zmiennej tym bardziej, że stała ta nie występuje w aksjomatach. $\square$

Analiza dowodu twierdzenia o stałej pozwala zauważyć

Wniosek 2.33 *Jeżeli T jest teorią w języku L , φ jest formułą z tego języka, a c – stałą spoza L , to φ można dowieść w teorii T posługując się formułami języka $L \cup \{c\}$ wtedy i tylko wtedy, gdy φ można dowieść w teorii T posługując się formułami języka L . $\square$*

Opuszczanie kwantyfikatora egzystencjalnego. Metoda budowania dowodów zaproponowana przez Herbranda korzysta z reguły opuszczania kwantyfikatora egzystencjalnego wykorzystywanej w niektórych systemach logicznych, zwłaszcza w systemie dedukcji naturalnej, patrz część 2, rozdziały 6.1 i 6.3. Najpierw dowiedzimy w nieco ograniczonym zakresie, że w dowolnym systemie można z tej reguły korzystać.

Niech T będzie teorią w języku L . Rozszerzamy L o pewien zbiór stałych i powiększamy zbiór aksjomatów T tworząc w ten sposób teorię T_c . Robimy to dodając do L dla każdego zdania postaci $\exists x \varphi$ specjalną stałą c (wyznaczoną przez to zdanie) i do teorii T nowy aksjomat $(\exists x \varphi) \Rightarrow \varphi[x \leftarrow c]$. Teorię T_c konstruujemy tak długo aż dla każdej takiej formuły będzie istniała w języku odpowiednia stała, a w teorii znajdzie się odpowiedni aksjomat. Zwykle wymaga to nieskończenie wielu kroków. Czasem taką teorię nazywa się teorią Henkina.

Twierdzenie 2.34 *Jeżeli φ jest formułą zapisaną w języku L i $T_c \vdash \varphi$, to także $T \vdash \varphi$.*

Dowód. Twierdzenie to mówi, że w dowodach mamy prawo posługiwać się regułą opuszczania kwantyfikatora egzystencjalnego przynajmniej w szczególnych przypadkach, gdy przesłanka jest zdaniem.

Dowodząc zdanie φ w teorii T_c korzystamy ze skończonej liczby specjalnych aksjomatów. Niech $\exists x \psi \Rightarrow \psi[x \leftarrow c]$ będzie najpóźniej dodanym, wykorzystywanym w dowodzie specjalnym aksjomatem, a Ψ – koniunkcją pozostałych. Stała c nie występuje więc w Ψ . Na mocy twierdzenia o dedukcji

$$T \vdash ((\exists x \psi) \Rightarrow \psi[x \leftarrow c]) \Rightarrow (\Psi \Rightarrow \varphi).$$

Możemy też założyć, że zmienna x nie występuje w φ . Wtedy

$$((\exists x \psi) \Rightarrow \psi[x \leftarrow c]) \Rightarrow (\Psi \Rightarrow \varphi) = (((\exists x \psi) \Rightarrow \psi) \Rightarrow (\Psi \Rightarrow \varphi))[x \leftarrow c].$$

Stąd na mocy twierdzenia o stałej otrzymujemy, że

$$T \vdash \forall x (((\exists x \psi) \Rightarrow \psi) \Rightarrow (\Psi \Rightarrow \varphi)).$$

Korzystając jak zwykle z praw logiki możemy zauważyć, że także

$$T \vdash ((\exists x \psi) \Rightarrow \exists x \psi) \Rightarrow (\Psi \Rightarrow \varphi)$$

i w konsekwencji,

$$T \vdash \Psi \Rightarrow \varphi.$$

Powołując się na indukcję ze względu na liczbę potrzebnych w dowodzie aksjomatów specjalnych i wniosek 2.33 otrzymujemy, że nawet w języku bez specjalnych stałych

$$T \vdash \varphi. \quad \square$$

Wstępny przykład. Teoria T_c ma kilka ciekawych własności. Na przykład, jeżeli $\exists x \varphi$ jest zdaniem i c jest stałą specjalną dla tego zdania, to

$$T_c \vdash \exists x \varphi \Rightarrow \varphi[x \leftarrow c].$$

Implikacja odwrotna jest prawem logicznym, a więc także

$$T_c \vdash \exists x \varphi \Leftrightarrow \varphi[x \leftarrow c],$$

a nawet

$$T_c \vdash \forall x \neg \varphi \Leftrightarrow \neg \varphi[x \leftarrow c].$$

W teorii T_c dowodzenie zdań z kwantyfikatorem sprowadza się więc do dowodzenia zdań bez kwantyfikatora.

W szczególności, jeżeli $\exists x \neg \varphi$ jest zdaniem i c jest stałą specjalną dla tego zdania, to

$$T_c \vdash \varphi[x \leftarrow c] \Leftrightarrow \forall x \varphi.$$

Tego typu własność można skomplikować.

Przypuśćmy, że mamy formułę

$$\exists x \forall y \exists z \forall w \psi(x, y, z, w),$$

a a jest termem stałym, który w przyszłości zostanie odpowiednio dobrany. Dla tego termu istnieje stała specjalna $\tau(a)$ dla zdania

$$\exists y \neg \exists z \forall w \psi[x \leftarrow a].$$

Z przytoczonego spostrzeżenia otrzymujemy teraz

$$T_c \vdash \exists z \forall w \psi[x \leftarrow a][y \leftarrow \tau(a)] \Leftrightarrow \forall y \exists z \forall w \psi[x \leftarrow a],$$

a także

$$T_c \vdash \exists z \forall w \psi[x \leftarrow a][y \leftarrow \tau(a)] \Rightarrow \exists x \forall y \exists z \forall w \psi.$$

Powtórzmy to rozumowanie raz jeszcze. Dla termu stałego b (i termu a) bierzemy stałą specjalną $\sigma(a, b)$ dla zdania

$$\exists w \neg \psi[x \leftarrow a][y \leftarrow \tau(a)][z \leftarrow b].$$

Mamy więc

$$T_c \vdash \psi[x \leftarrow a][y \leftarrow \tau(a)][z \leftarrow b][w \leftarrow \sigma(a, b)] \Leftrightarrow \forall w \psi[x \leftarrow a][y \leftarrow \tau(a)][z \leftarrow b]$$

oraz

$$T_c \vdash \psi[x \leftarrow a][y \leftarrow \tau(a)][z \leftarrow b][w \leftarrow \sigma(a, b)] \Rightarrow \exists z \forall w \psi[x \leftarrow a][y \leftarrow \tau(a)].$$

Łącząc dwa udowodnione twierdzenia otrzymujemy, że

$$T_c \vdash \psi[x \leftarrow a][y \leftarrow \tau(a)][z \leftarrow b][w \leftarrow \sigma(a, b)] \Rightarrow \exists x \forall y \exists z \forall w \psi.$$

Gdyby teraz udało nam się tak dobrać termy a i b , aby

$$T_c \vdash \psi[x \leftarrow a][y \leftarrow \tau(a)][z \leftarrow b][w \leftarrow \sigma(a, b)],$$

to przedstawione rozumowanie dawałoby, że

$$T_c \vdash \exists x \forall y \exists z \forall w \psi.$$

Przedstawione rozumowanie, które może zostać przeprowadzone dla dowolnych formuł w postaci preneksowej, klasy Σ_{2k} zostanie wykorzystane do dowodu kolejnego twierdzenia Herbranda.

Dowód twierdzenia o pełności.

Twierdzenie 2.35 (Herbrand) *Przypuśćmy, że φ jest formułą i formuła $\neg(\neg\varphi)^*$ jest równoważna formule $\exists x \dots \psi$, gdzie ψ jest formułą otwartą. Jeżeli istnieją układy termów $\vec{t}_1, \dots, \vec{t}_m \in \mathcal{T}_0^n$ takie, że formuła*

$$\psi[\vec{x} \leftarrow \vec{t}_1] \vee \dots \vee \psi[\vec{x} \leftarrow \vec{t}_m]$$

jest tautologią, to φ jest prawem rachunku kwantyfikatorów.

Dowód. Dowód przedstawimy na przykładzie formuły klasy Σ_4 . Niech więc

$$\varphi = \exists x \forall y \exists z \forall w \psi(x, y, z, w).$$

Wtedy formuła $\neg(\neg\varphi)^*$ ma postać

$$\exists x \exists z \psi(x, f(x), z, g(x, z))$$

dla pewnych nowych symboli funkcyjnych f i g . Zakładamy więc, że istnieją układy termów stałych t_i oraz s_i takie, że formuła

$$\Phi = \bigvee_i \psi(t_i, f(t_i), s_i, g(t_i, s_i))$$

jest tautologią. Niech teraz T oznacza rachunek kwantyfikatorów, czyli teorię pozabawioną aksjomatów. Dla uproszczenia zakładamy także, że T nie zawiera nawet aksjomatów równości (dla teorii z równością są potrzebne dodatkowe konstrukcje, które nie zostaną przedstawione). Utwórzmy teorię T_c i weźmy zdefiniowane wyżej operacje τ i σ przyporządkowujące termom stałym specjalne stałe. Formułę Φ będziemy przekształcać wielokrotnie wykonując przekształcenie następującej postaci: sprawdzamy, czy w Φ występuje podterm postaci $f(a)$ dla pewnego termu stałego a nie zawierającego ani f , ani g , i jeżeli taki podterm znajdziemy, to zastępujemy go we wszystkich miejscach stałą $\tau(a)$. Analogicznie, wszystkie wystąpienia podtermów postaci $g(a, b)$ zastępujemy stałą $\sigma(a, b)$.

Zauważmy, że przekształcenia te zachowują „bycie tautologią”. Przypuśćmy, że po wykonaniu wszystkich możliwych takich przekształceń otrzymaliśmy formułę Φ^* . Tak więc Φ^* jest tautologią.

Podczas wykonywania opisanych przekształceń w jakiś sposób zmieniały się termy t_i oraz s_i . Przyjmimy, że zostały przekształcone w a_i oraz b_i odpowiednio. Nietrudno zauważyć, że formuła Φ^* ma postać

$$\Phi^* = \bigvee_i \psi(a_i, \tau(a_i), b_i, \sigma(a_i, b_i)).$$

Zgodnie z uwagami z poprzedniego paragrafu, człony tej alternatywy implikują zdanie φ . Wobec tego, także mamy

$$T_c \vdash \Phi^* \Rightarrow \varphi.$$

Na mocy twierdzenia o pełności dla rachunku zdań dowód ma także formuła Φ^* . Tak więc φ jest twierdzeniem teorii T_c , a ponieważ nie zawiera stałych specjalnych, jest też twierdzeniem rachunku kwantyfikatorów. $\square$

Przykład konstruowania dowodu. Weźmy tautologię

$$\varphi = \exists x \forall y A(x, y) \Rightarrow \forall y \exists x A(x, y)$$

i spróbujemy skonstruować jej dowód metodą Herbranda. Postacią normalną φ jest

$$\forall x \exists y \forall z \exists t (A(x, y) \Rightarrow A(t, z)),$$

a postacią normalną $\neg\varphi$ –

$$\exists x \forall y \exists z \forall t \neg(A(x, y) \Rightarrow A(t, z)).$$

Z tej formuły po skolemizacji otrzymujemy

$$\forall y \forall t \neg(A(c, y) \Rightarrow A(t, f(y))).$$

Ostatecznie mamy

$$\neg(\neg\varphi)^* \Leftrightarrow \exists y \exists t (A(c, y) \Rightarrow A(t, f(y))).$$

Zgodnie z twierdzeniem Herbranda, formuła φ jest tautologią, ponieważ tautologią jest

$$(A(c, d) \Rightarrow A(c, f(d))) \vee (A(c, f(d)) \Rightarrow A(d', f(f(d)))). \quad (8)$$

Pierwsza implikacja jest postaci $(A(c, y) \Rightarrow A(t, f(y)))[y \leftarrow d][t \leftarrow c]$, a druga $(A(c, y) \Rightarrow A(t, f(y)))[y \leftarrow f(d)][t \leftarrow d']$.

Weźmy teraz stałą specjalną α wprowadzoną aksjوماتem

$$\exists x \neg \exists y \forall z \exists t (A(x, y) \Rightarrow A(t, z)) \Rightarrow \neg \exists y \forall z \exists t (A(\alpha, y) \Rightarrow A(t, z)).$$

Z tego aksjomatu wynika następująca własność stałej α :

$$\exists y \forall z \exists t (A(\alpha, y) \Rightarrow A(t, z)) \Rightarrow \forall x \exists y \forall z \exists t (A(x, y) \Rightarrow A(t, z)).$$

Dla dowolnego termu b weźmy jeszcze stałą specjalną $\beta_\alpha(b)$ wprowadzoną aksjوماتem

$$\exists z \neg \exists t (A(\alpha, b) \Rightarrow A(t, z)) \Rightarrow \neg \exists t (A(\alpha, b) \Rightarrow A(t, \beta_\alpha(b))).$$

Mamy więc

$$\exists t (A(\alpha, b) \Rightarrow A(t, \beta_\alpha(b))) \Rightarrow \forall z \exists t (A(\alpha, b) \Rightarrow A(t, z)),$$

a także

$$\exists t (A(\alpha, b) \Rightarrow A(t, \beta_\alpha(b))) \Rightarrow \forall x \exists y \forall z \exists t (A(x, y) \Rightarrow A(t, z)).$$

Teraz w tautologii (8) symbole wprowadzone podczas skolemizacji zastępujemy przez odpowiednie stałe specjalne. Tym razem skolemizacja wymagała posłużenia się stałą c i symbolem f . Stałą c zastępujemy stałą specjalną α , a symbol f – operacją $\beta_\alpha(\cdot)$. Po wykonaniu tych zastępowań w formule (8) otrzymujemy

$$(A(\alpha, d) \Rightarrow A(\alpha, \beta_\alpha(d))) \vee (A(\alpha, \beta_\alpha(d)) \Rightarrow A(d', \beta_\alpha(\beta_\alpha(d)))).$$

Analizując postać tej formuły oraz formuły (8) można utworzyć taki oto dowód wprost formuły φ (a właściwie jej preneksowej postaci normalnej).

- | | | |
|------|---|-----------------|
| 1) | $\exists x \neg \exists y \forall z \exists t (A(x, y) \Rightarrow A(t, z)) \Rightarrow \neg \exists y \forall z \exists t (A(\alpha, y) \Rightarrow A(t, z))$ | aksjomat T_c |
| 2) | $\exists y \forall z \exists t (A(\alpha, y) \Rightarrow A(t, z)) \Rightarrow \forall x \exists y \forall z \exists t (A(x, y) \Rightarrow A(t, z))$ | kontrapoz. i 1) |
| 3) | $\exists z \neg \exists t (A(\alpha, d) \Rightarrow A(t, z)) \Rightarrow \neg \exists t (A(\alpha, d) \Rightarrow A(t, \beta_\alpha(d)))$ | aksjomat T_c |
| 4) | $\exists t (A(\alpha, d) \Rightarrow A(t, \beta_\alpha(d))) \Rightarrow \forall z \exists t (A(\alpha, d) \Rightarrow A(t, z))$ | kontrapoz. i 3) |
| 5) | $\exists z \neg \exists t (A(\alpha, \beta_\alpha(d)) \Rightarrow A(t, z)) \Rightarrow \neg \exists t (A(\alpha, \beta_\alpha(d)) \Rightarrow A(t, \beta_\alpha(\beta_\alpha(d))))$ | aksjomat T_c |
| 6) | $\exists t (A(\alpha, \beta_\alpha(d)) \Rightarrow A(t, \beta_\alpha(\beta_\alpha(d)))) \Rightarrow \forall z \exists t (A(\alpha, \beta_\alpha(d)) \Rightarrow A(t, z))$ | kontrapoz. i 5) |
| 7) | $A(\alpha, \beta_\alpha(d)) \vee \neg A(\alpha, \beta_\alpha(d))$ | aksjomat logiki |
| 8) | $(A(\alpha, d) \Rightarrow A(\alpha, \beta_\alpha(d))) \vee (A(\alpha, \beta_\alpha(d)) \Rightarrow A(d', \beta_\alpha(\beta_\alpha(d))))$ | wniosek z 7) |
| 1.1) | $A(\alpha, d) \Rightarrow A(\alpha, \beta_\alpha(d))$ | przypadek 1 |
| 1.2) | $A(\alpha, d) \Rightarrow A(t, \beta_\alpha(d)) [t \leftarrow \alpha]$ | 1.1) inaczej |
| 1.3) | $\exists t (A(\alpha, d) \Rightarrow A(t, \beta_\alpha(d)))$ | z 1.2) |
| 1.4) | $\forall z \exists t (A(\alpha, d) \Rightarrow A(t, z))$ | RO z 1.3) i 4) |
| 1.5) | $\exists y \forall z \exists t (A(\alpha, y) \Rightarrow A(t, z))$ | z 1.4) |
| 1.6) | $\forall x \exists y \forall z \exists t (A(x, y) \Rightarrow A(t, z))$ | RO z 1.5) i 2) |
| 2.1) | $A(\alpha, \beta_\alpha(d)) \Rightarrow A(d', \beta_\alpha(\beta_\alpha(d)))$ | przypadek 2 |
| 2.2) | $A(\alpha, \beta_\alpha(d)) \Rightarrow A(t, \beta_\alpha(\beta_\alpha(d))) [t \leftarrow d']$ | 2.1) inaczej |
| 2.3) | $\exists t A(\alpha, \beta_\alpha(d)) \Rightarrow A(t, \beta_\alpha(\beta_\alpha(d)))$ | z 2.2) |
| 2.4) | $\forall z \exists t (A(\alpha, \beta_\alpha(d)) \Rightarrow A(t, z))$ | RO z 2.3) i 4) |
| 2.5) | $\exists y \forall z \exists t (A(\alpha, y) \Rightarrow A(t, z))$ | z 2.4) |
| 2.6) | $\forall x \exists y \forall z \exists t (A(x, y) \Rightarrow A(t, z))$ | RO z 2.5) i 2) |
| 9) | $\forall x \exists y \forall z \exists t (A(x, y) \Rightarrow A(t, z))$ | z 1.6) i 2.6) |

Przedstawiony dowód jest sprawia wrażenie sztucznego. Zwykle nie posługujemy się specjalnymi aksjomatami teorii T_c , a raczej tzw. regułą odrywania kwantyfikatora egzystencjalnego. Naturalniejszą jest niżej przedstawiona wersja nie wprost tego dowodu. Należy pamiętać, że powstał na podstawie analizy tautologii (8).

1) $\neg \forall x \exists y \forall z \exists t (A(x, y) \Rightarrow A(t, z))$	założenie dowodu nie wprost
2) $\exists x \forall y \exists z \forall t \neg (A(x, y) \Rightarrow A(t, z))$	prawa de Morgana i 1)
3) $\forall y \exists z \forall t \neg (A(\alpha, y) \Rightarrow A(t, z))$	O $\exists$ i 2)
4) $\exists z \forall t \neg (A(\alpha, d) \Rightarrow A(t, z))$	O $\forall$ i 3)
5) $\forall t \neg (A(\alpha, d) \Rightarrow A(t, \beta))$	O $\exists$ i 4)
6) $\neg (A(\alpha, d) \Rightarrow A(\alpha, \beta))$	O $\forall$ i 5)
7) $\exists z \forall t \neg (A(\alpha, \beta) \Rightarrow A(t, z))$	O $\forall$ i 3)
8) $\forall t \neg (A(\alpha, \beta) \Rightarrow A(t, \gamma))$	O $\exists$ i 7)
9) $A(\alpha, \beta) \vee \neg A(\alpha, \beta)$	aksjomat logiki
10) $(A(\alpha, d) \Rightarrow A(\alpha, \beta)) \vee (A(\alpha, \beta) \Rightarrow A(d', \gamma))$	wniosek z 9)
11) $A(\alpha, \beta) \Rightarrow A(d', \gamma)$	wniosek z 9) i 6)
12) $\neg (A(\alpha, \beta) \Rightarrow A(d', \gamma))$	O $\forall$ i 8)
13) $\square$	sprzeczność 11) i 12)

2.10 Dodatek: Teorie Henkina i szczegółowy dowód twierdzenia o pełności

Jest to właściwie ten sam dowód co w poprzednim rozdziale, ale został on przedstawiony z wieloma dodatkowymi szczegółami.

Niech T będzie dowolną teorią zapisaną w języku L . Formuły zapisane w języku L , z jedną zmienną wolną ustawiamy w ciąg $\varphi_1(x_1), \varphi(x_2), \dots$. Bierzemy też ciąg nowych stałych $c_1, c_2, \dots$. Niech

$$T_1 = T \cup \{(\exists x_i \varphi_i(x_i)) \Rightarrow \varphi_i[x_i \leftarrow c_i] : i \in N\}$$

będzie teorią otrzymaną z T przez dodanie ciągu nowych aksjomatów. Stałą c_i użytą w konstrukcji tego ciągu będziemy nazywać stałą specjalną dla formuły $\exists x_i \varphi_i(x_i)$.

Z twierdzenia o stałej wynika, że jeżeli zapisana w języku L formuła φ ma dowód w teorii T_1 , to może też zostać dowiedziona w teorii T bez korzystania w dowodzie ze stałych specjalnych.

Przedstawioną konstrukcję będziemy teraz powtarzać. Niech T_{i+1} oznacza teorię, która powstała w wyniku zastosowania przedstawionej konstrukcji do teorii T_i (oczywiście za każdym razem bierzemy nowe stałe). Przyjmijmy, że T_∞ oznacza $\bigcup_{i \in N} T_i$. Teorie tej postaci nazywa się teoriami Henkina.

Lemat 2.36 *Jeżeli zapisana w języku L formuła φ ma dowód w teorii T_∞ , to może też zostać dowiedziona w teorii T bez korzystania w dowodzie ze stałych specjalnych użytych w konstrukcji T_∞ .*

Dowód. Aby dowieść ten lemat trzeba zauważyć dwa fakty. Po pierwsze, jeżeli formuła jest twierdzeniem T_∞ , to daje się też dowieść w teorii T_n dla pewnego n , korzystając z aksjomatów tej teorii i posługując się jej językiem. Następnie trzeba zauważyć, że daje się też dowieść w teorii T_{n-1} , a na mocy zasady indukcji, także w teorii T . $\square$

Twierdzenie 2.37 (Herbranda, o pełności) *Jeżeli zdanie Φ jest tautologią, to ma dowód w rachunku kwantyfikatorów, nie wymagający żadnych dodatkowych symboli.*

Dowód. Niech R oznacza rachunek kwantyfikatorów, a R_∞ teorię Henkina otrzymaną z rachunku kwantyfikatorów R . Skorzystamy z poprzedniego lematu i dowiedzimy tylko, że Φ jest twierdzeniem R_∞ . Przy okazji będziemy musieli przyjrzeć się trochę dokładniej skolemizacji.

Możemy dodatkowo założyć, że Φ jest w preneksowej postaci normalnej. Przyjmijmy, że

$$\Phi = \exists \vec{x} \forall y \Psi \quad \text{oraz} \quad \Psi = Q_1 z_1 \dots Q_k z_k \psi$$

dla pewnych kwantyfikatorów Q_i i dla bezkwantyfikatorowej formuły ψ . Dalsze rozumowanie zostanie przeprowadzone przez indukcję ze względu na liczbę kwantyfikatorów ogólnych w formule Φ (lub egzystencjalnych w $\neg\Phi$).

Równolegle będziemy definiować postać skolemowską formuły $\neg\Phi$, ciąg symboli funkcyjnych $f, \dots$ występujących w postaci skolemowskiej, pewne funkcje $F, \dots$, które stałym teorii R_∞ przyporządkowują inne stałe tej teorii, funkcje te będą wyznaczane przez wprowadzane symbole funkcyjne, a także ciąg podstawień η za zmienne występujące w $\neg\Phi$ i skwantyfikowane egzystencjalnie. Oprócz tego będziemy rozważać ciągi α postawień stałych specjalnych za zmienne występujące w $\neg\Phi$ i skwantyfikowane ogólnie.

Jeżeli η jest pewnym podstawieniem, to symbolem η_c będziemy oznaczać podstawienie, które za dowolną zmienną z , zamiast termu $\eta(z)$ podstawia stałą, która jest obliczana tak, jak wartość termu $\eta(z)$ przy interpretacji symboli funkcyjnych $f, \dots$ jako funkcji $F, \dots$ i zmiennych jako odpowiednio określonych stałych (szczegóły później staną się jasne).

Cała konstrukcja będzie miała następujące (dowodzone przez indukcję) własności:

- 1) Domknięcie kwantyfikatorami ogólnymi formuły $\neg\psi\eta$ jest postacią skolemowską $\neg\Phi$.
- 2) Dla każdego podstawienia α za wszystkie zmienne (występujące w $\exists y \neg\Psi$, a także w $\neg\psi\eta_c$) zdanie $(\exists y \neg\Psi\alpha) \Rightarrow \neg\psi\eta_c\alpha$ ma dowód w R_∞ korzystający z symboli formuły Ψ i stałych specjalnych.

Ponieważ zostały już wprowadzone oznaczenia implikujące konieczność przeprowadzenia drugiego kroku indukcyjnego, więc zaczynamy od niego.

Przeprowadzając skolemizację $\neg\Phi$ wprowadzamy symbol funkcyjny f . Mamy więc

$$(\neg\Phi)^* = (\forall \vec{x} \neg\Psi[y \leftarrow f(\vec{x})])^*.$$

Dla formuły $\Psi[y \leftarrow \dots]$ możemy skorzystać z założenia indukcyjnego. Wobec tego

$$(\forall \vec{x} \neg\Psi[y \leftarrow f(\vec{x})])^* = \forall \dots \neg\psi[y \leftarrow f(\vec{x})]\eta'$$

dla podstawienia η' konstruowanego przez indukcję. Możemy więc przyjąć, że

$$\eta = [y \leftarrow f(\vec{x})]\eta'.$$

Nietrudno zauważyć, że wtedy przeprowadzana konstrukcja ma własność 1), gdyż

$$(\neg\Phi)^* = \forall \dots \neg\psi[y \leftarrow f(\vec{x})]\eta' = \forall \dots \neg\psi\eta.$$

Teraz zdefiniujemy interpretację symbolu f . Dla stałych $\vec{a}$ symbolem $F(\vec{a})$ będziemy oznaczać stałą specjalną dla formuły

$$\exists y \neg \Psi[\vec{x} \leftarrow \vec{a}].$$

Wobec tego aksjomatem teorii R_∞ jest każda formuła

$$(\exists y \neg \Psi[\vec{x} \leftarrow \vec{a}]) \Rightarrow \neg \Psi[y \leftarrow F(\vec{a})][\vec{x} \leftarrow \vec{a}]$$

(zauważmy, że kolejność podstawień nie ma znaczenia).

Możemy już przystąpić do dowodu własności 2). Weźmy dowolne podstawienie α . Przyjmijmy też, że $(\neg \Psi)^-$ to formuła $\neg \Psi$ pozbawiona początkowego bloku kwantyfikatorów ogólnych. Formuła

$$(\exists y \neg \Psi[\vec{x} \leftarrow \alpha(\vec{x})]) \Rightarrow \neg \Psi[y \leftarrow F(\alpha(\vec{x}))][\vec{x} \leftarrow \alpha(\vec{x})],$$

czyli

$$(\exists y \neg \Psi \alpha) \Rightarrow \neg \Psi[y \leftarrow F(\alpha(\vec{x}))]\alpha$$

jest aksjomatem teorii R_∞ , formuła

$$\neg \Psi[y \leftarrow F(\alpha(\vec{x}))]\alpha \Rightarrow (\neg \Psi)^-[y \leftarrow F(\alpha(\vec{x}))]\alpha$$

jest prawem rachunku kwantyfikatorów, a formuła

$$(\neg \Psi)^-[y \leftarrow F(\alpha(\vec{x}))]\alpha \Rightarrow \neg \psi[y \leftarrow F(\alpha(\vec{x}))]\eta'_c \alpha$$

na mocy założenia indukcyjnego daje dowieść się w R_∞ . W końcu mamy

$$\neg \psi[y \leftarrow F(\alpha(\vec{x}))]\eta'_c \alpha = \neg \psi([y \leftarrow f(\vec{x})]\eta')_c \alpha = \neg \psi \eta_c \alpha.$$

Składając te trzy implikacje i korzystając z definicji η otrzymujemy

$$R_\infty \vdash (\exists y \neg \Psi \alpha) \Rightarrow \neg \psi \eta_c \alpha,$$

a więc własność 2).

Sprawdzenie pierwszego kroku indukcyjnego jest proste. W tym przypadku nie wprowadzamy symboli funkcyjnych, nie definiujemy ich interpretacji, a podstawienie η definiujemy jako puste. Tym samym dowód indukcyjny jest zakończony i obie podane na początku dowodu własności przeprowadzanej konstrukcji zachodzą dla wszystkich formuł, w tym dla rozważanej formuły Φ .

Skorzystamy teraz z twierdzenia 2.17 dla zdania Φ . Postacią skolemowską $\neg \Phi$ jest formuła $\forall \dots \neg \psi \eta$, jej negacja $\neg(\neg \Phi)^*$ jest równa $\exists \dots \psi \eta$. Ponieważ Ψ jest tautologią, na mocy twierdzenia 2.17 tautologią jest formuła postaci

$$\bigvee_i \psi \eta \gamma_i \tag{9}$$

dla pewnych podstawień, które zmiennym przyporządkowują termy stałe. W podstawieniach η i γ_i występują symbole funkcyjne wprowadzone podczas skolemizacji. Stopniowo zinterpretujemy je jako wskazane funkcje na stałych. Niech więc $\alpha_i(x)$ oznacza stałą będącą tak zinterpretowaną wartością termu γ_i . Po zastąpieniu wszystkich zmiennych w formule $\psi \eta$ stałymi analogicznie możemy interpretować symbole funkcyjne w termach podstawianych przez η . Po tych interpretacjach tautologia (9) przekształci się w formułę

$$\bigvee_i \psi \eta_c \alpha_i.$$

Ta formuła też jest tautologią. Zmiany dokonywane na termach nie wpływają na postać formuły, w ich wyniku może najwyżej dojść do utożsamienia niektórych formuł atomowych.

Teraz skorzystamy z własności 2) dla podstawień α_i . Wynika z niej, że

$$R_\infty \vdash \exists y \neg \Psi \alpha_i \Rightarrow \neg \psi \eta_c \alpha_i$$

dla wszystkich i . Ponieważ formuły

$$\neg \Phi \Rightarrow \exists y \neg \Psi \alpha_i$$

są prawami logicznymi, więc

$$R_\infty \vdash \neg \Phi \Rightarrow \neg \psi \eta_c \alpha_i.$$

Na mocy prawa kontrapozycji otrzymujemy, że

$$R_\infty \vdash \psi \eta_c \alpha_i \Rightarrow \Phi$$

dla wszystkich i . Stąd

$$R_\infty \vdash \left(\bigvee_i \psi \eta_c \alpha_i \right) \Rightarrow \Phi,$$

a ponieważ poprzednik tej implikacji jest tautologią, więc także

$$R_\infty \vdash \Phi,$$

a to kończy dowód. Można też zauważyć, że jeżeli znajdziemy konkretną tautologię (9), to na podstawie przedstawionego rozumowania można skonstruować efektywnie dowód formuły Φ w rachunku kwantyfikatorów. $\square$

3 Funkcje obliczalne według Herbranda i Gödla

Koncepcja funkcji rekurencyjnych zasugerowana przez Herbranda została przedstawiona przez Gödla podczas wykładów Institute for Advanced Study w 1934 roku i opublikowana w formie notatek z wykładu sporządzonych przez J.B. Rossera i S.C. Kleene'ego, doktorantów Alonzo Churcha.

3.1 Termy

Będziemy posługiwać się alfabetem, w którym jest nieskończenie wiele (przynajmniej potencjalnie) zmiennych i jest nieskończenie wiele symboli funkcyjnych każdej możliwej arności.

W tym alfabecie wyróżniamy jedną zmienną 0 i jeden symbol funkcyjny jednoargumentowy S . Zmienna 0 jest taką zmienną, pod którą nie wolno podstawiać.

Dalej rozważamy termy nad takim alfabetem. Symbolem $\underline{n}$ oznaczamy termy zdefiniowane rekurencyjnie równościami $\underline{0} = 0$ oraz $\underline{n+1} = S(\underline{n})$ zachodzącymi dla wszystkich liczb naturalnych n .

3.2 Równości i układy

Równości to pary termów. Równość dwóch termów zapisujemy w postaci $t_1 = t_2$. Od termu t_1 występującego w równości żądamy, aby nie zaczynał się symbolem S . Zwróćmy uwagę, że podany warunek świadczy o tym, że termy t_1 i t_2 odgrywają w równości inną rolę. Intuicyjnie równość $t_1 = t_2$ interpretujemy jako informację, że aby znaleźć wartość termu t_1 (czyli termu postaci $f(\dots)$, a więc wartość funkcji f dla pewnych argumentów) należy obliczyć wartość termu t_2 .

Układem równości nazywamy ciąg równości. Pierwszy symbol pierwszego termu ostatniej równości układu E nazywamy symbolem głównym układu E . Intuicyjnie, układ E definiuje funkcję oznaczaną symbolem głównym układu E .

3.3 Wyprowadzenia

Wyprowadzeniem z układu E nazywamy ciąg równości $r_1, \dots, r_n$ taki, że dla każdego $i \leq n$ równość r_i z tego ciągu

- 1) albo należy do E ,
- 2) albo powstała przez podstawienie w równości r_j dla $j < i$ za pewną zmienną x (różną od 0) termu postaci $\underline{n}$ dla pewnego naturalnego n ($r_i = r_j[x \leftarrow \underline{n}]$).
- 3) albo też powstała przez zastąpienie w pewnym miejscu w równości r_j dla $j < i$ podtermu postaci $f(\underline{n}_1, \dots, \underline{n}_k)$ termem $\underline{m}$ pod warunkiem, że dla pewnego $k < i$ równość r_k jest tożsama z $f(\underline{n}_1, \dots, \underline{n}_k) = \underline{m}$.

3.4 Analiza pojęcia wyprowadzenia

Teraz termy będziemy uważać także za drzewa z podwójnymi etykietami. Oprócz zwykłych etykiet poszczególne węzły będą etykietowane liczbami naturalnymi z tym, że te etykiety będą zdefiniowane lub nie. Etykiety będące liczbami naturalnymi będziemy nazywać wartościami i będziemy mówić o drzewach z wartościowaniem.

Drzewo będzie poprawnie zwartościowane (albo dwa drzewa będą poprawnie zwartościowane), jeżeli

- 1) dla każdej zmiennej x w sytuacji, gdy jeden z węzłów z etykietą x ma zdefiniowaną wartość, pozostałe też mają zdefiniowaną wartość i wartościami tych węzłów jest ta sama liczba,
- 2) każdy węzeł z etykietą 0 ma wartość 0,
- 3) węzeł z etykietą S ma wartość n wtedy i tylko wtedy, gdy $n > 0$ i węzeł potomny ma wartość $n - 1$.

Drzewo jest poprawnie zwartościowane ze względu na układ równości E , jeżeli jest poprawnie zwartościowane i dodatkowo

- 1) jeżeli węzeł z symbolem funkcyjnym $f \neq S$ ma wartość m , to wszystkie węzły potomne mają wartości,
- 2) jeżeli węzły potomne takiego węzła mają wartości $n_1, \dots, n_k$, to $E \vdash f(\underline{n}_1, \dots, \underline{n}_k) = \underline{m}$.

Niech T będzie drzewem z wartościami. Każdy węzeł w tego drzewa opisuje pewien term $T(w)$ zdefiniowany zgodnie ze wzorem

$$T(w) = \begin{cases} \underline{n} & \text{jeżeli } w \text{ ma określoną wartość i równa się ona } n, \\ x & \text{jeżeli } w \text{ nie ma wartości i jest etykietowany zmienną } x, \\ f(T(w_1), \dots) & \text{jeżeli } w \text{ nie ma wartości, ma etykietę } f \text{ i potomków } w_1, \dots \end{cases}$$

Będziemy mówić, że drzewo T opisuje term $T(w_0)$, gdzie w_0 jest korzeniem T .

Lemat 3.1 *Przypuśćmy, że $E \vdash t = s$. Wtedy istnieje równość $t' = s' \in E$, a jeżeli termy t' i s' potraktujemy jako drzewa $D(t')$ i $D(s')$ odpowiednio, to istnieją też poprawne ze względu na E wartościowania tych drzew takie, że $D(t')$ opisuje t i $D(s')$ opisuje s .*

3.5 Funkcje obliczalne

Częściowa funkcja $F : N^k \rightarrow N$ jest obliczalna według Herbranda-Gödla, jeżeli istnieje układ równości E z k -arnym symbolem głównym f taki, że

$$F(n_1, \dots, n_k) = m \iff E \vdash f(\underline{n_1}, \dots, \underline{n_k}) = \underline{m}$$

dla wszystkich liczb naturalnych $n_1, \dots, n_k, m$.

Lemat 3.2 *Dodawanie, mnożenie, funkcje $I_{n,k}$, funkcja charakterystyczna nierówności i funkcja Ackermanna są obliczalne według Herbranda-Gödla.*

Lemat 3.3 *Klasa funkcji obliczalnych według Herbranda-Gödla jest zamknięta ze względu na składanie oraz rekursję prostą (i inne, poprawne schematy rekursji).*

Lemat 3.4 *Klasa funkcji obliczalnych według Herbranda-Gödla jest zamknięta ze względu na operację minimum.*

Dowód. Przypuśćmy, że f jest funkcją obliczalną według Herbranda-Gödla, definiowaną układem równań E_f z symbolem głównym f . Weźmy następujący układ równań:

$$\begin{aligned} g(x, 0) &= S(0), \\ g(x, S(y)) &= g(x, y) \cdot f(x, y), \\ h(S(x), 0, y) &= y, \\ k(x) &= h(g(x, y), g(x, S(y)), y). \end{aligned}$$

Układ E , czyli podany układ uzupełniony o E_f i układ definiujący mnożenie definiuje funkcję μy ($f(x, y) = 0$).

Najpierw zauważmy, że dwie pierwsze równości definiują funkcję g , którą można też zdefiniować wzorem

$$g(x, y) = \prod_{z < y} f(x, z).$$

Zauważmy też, że

$$\mu y (f(x, y) = 0) = \mu y (g(x, S(y)) = 0).$$

Dalej wystarczy więc wykazać, że cały układ E definiuje funkcję

$$k(x) = \mu y (g(x, S(y)) = 0).$$

Przypuśćmy, że $E \vdash k(\underline{n}) = \underline{m}$. Równość ta może wynikać tylko z ostatniej równości E . Oznacza to, że w jej dowodzie zostały dowiedzione też równości pozwalające na zwartościowanie wszystkich podtermów termu $h(g(x, y), g(x, S(y)), y)$, a także w jakiś sposób została zwartościowana zmienna y . Tak więc istnieją liczby u, v, w takie, że w jest wartością zmiennej y oraz $E \vdash g(\underline{n}, \underline{w}) = \underline{u}$, $E \vdash g(\underline{n}, S(\underline{w})) = \underline{v}$ i $E \vdash h(\underline{u}, \underline{v}, \underline{w}) = \underline{m}$. Ponieważ układ E definiuje też opisaną wyżej funkcję g , więc $g(n, w) = u$ oraz $g(n, w + 1) = v$. Ostatnia z równości musiała natomiast zostać wyprowadzona z przedostatniej równości układu E . Tak więc $w = m$, $v = 0$ oraz $u > 0$. Mamy więc, że $g(n, m) > 0$ oraz $g(n, m + 1) = 0$. Funkcja g jest jednak specyficzna: warunek $g(n, m) > 0$ implikuje, że także $g(n, i) > 0$ dla wszystkich $i \leq m$. Stąd otrzymujemy, że $\mu y (g(n, S(y)) = 0) = m$. $\square$

Wniosek 3.5 *Wszystkie funkcje rekurencyjne są obliczalne według Herbranda-Gödla.*

Twierdzenie 3.6 *Klasa funkcji obliczalnych według Herbranda-Gödla jest równa klasie funkcji rekurencyjnych.*