

Jacques Herbrand i jego algorytm

Dowodzenie twierdzeń zawsze fascynowało ludzi, a współcześnie można próbować do tego wykorzystać komputery.

Antoni Kościelski

Najprostsza metoda automatycznego dowodzenia została przedstawiona w artykule *Logika z początku XX wieku*¹ i polega po prostu na poszukaniu dowodu. Implementacja tego algorytmu, wynikającego bezpośrednio z definicji prawa logicznego, zapewne generowałaby najpierw krótkie, następnie coraz dłuższe ciągi znaków i sprawdzała, czy przypadkiem nie został znaleziony dowód danej formuły. Przyjmijmy dla uproszczenia, że piszemy dowody formalne posługując się dziesięcioma znakami, a interesujące nas prawo ma dowód wymagający napisania przynajmniej 101 znaków. Taki algorytm, przed znalezieniem dowodu, powinien przebadać wszystkie stuznakowe teksty. Jest ich 10^{100} . Współczesne komputery bez problemów wykonują około miliarda rozkazów na sekundę. Przyjmijmy, że są w stanie przeanalizować w ciągu sekundy miliard potencjalnych dowodów. Łatwo ocenić, ile lat musiałoby trwać poszukiwanie dowodu. W ten sposób nie da się stworzyć sensownego programu.

Nadzieję na bardziej efektywne programy daje między innymi twierdzenie Herbranda charakteryzujące tautologie rachunku kwantyfikatorów. Dowód tego twierdzenia zawiera metodę sprawdzania, czy dana formuła jest taką tautologią, znaną jako algorytm Herbranda. Wobec twierdzenia o pełności w ten sam sposób możemy sprawdzać, czy formuła jest prawem rachunku kwantyfikatorów. Po roku 1959 algorytm Herbranda odegrał znaczącą rolę w rozwoju informatyki.



Jacques Herbrand podczas ostatniej wspinaczki.

Sylwetka Herbranda

Jacques Herbrand urodził się w Paryżu 12 lutego 1908 roku. Był jedynakiem, a rodzice dostrzegli jego ogromny talent i pomagali go rozwijać. Zwykle był najlepszy. W wieku 17 lat zdał najlepiej spośród kandydatów egzaminy do *École Normale Supérieure*. W 1928 roku zajął pierwsze miejsce w *Concours d'Agrégation* i uzyskał prawo nauczania we francuskich szkołach. Dwa lata poświęcił na przygotowanie doktoratu. Na przełomie 1929 i 1930 roku przerwał badania naukowe i odbył służbę wojskową, po czym otrzymał z najwyższymi honorami stopień doktora i został wyróżniony Nagrodą Rockefellera. Nagroda umożliwiła mu roczny pobyt na uniwersytetach niemieckich. Współpracował w Berlinie z późniejszym twórcą pierwszego komputera, Johnem von Neumannem i w Hamburgu z Emilem Artinem. Pracował też w Getyndze z Emmą Nöther. Niezwłocznie po powrocie do Francji wybrał się w góry. Ledwo rozpoczęte wakacje skończyły się szybko i tragicznie: zginął 27 lipca 1931 roku w pobliżu małej osady La Bérarde w Alpach, niedaleko Grenoble, podczas schodzenia z czterotysięcznika La Barre des Écrins (taki szczegół nie łatwo ustalić z całą pewnością). Ot po prostu lina, od której zależał los Herbranda, w jednym miejscu nie była dobrze zamocowana.

¹ Matematyka, 7/2008. Ten artykuł zawiera wiele informacji potrzebnych do zrozumienia tego tekstu.



La Barre des Écrins, 4102 m npm, jeden z najwyższych szczytów we Francji.

Zdażył napisać 10 artykułów poświęconych logice – zajmują ponad 250 stron druku² – i kilka innych, poświęconych zagadnieniom algebraicznym. Najważniejszym jest praca doktorska pod tytułem *Recherches sur la théorie de la démonstration*. Ciekawostką jest, że ukazała się w Polsce, w *Pracach Towarzystwa Naukowego Warszawskiego* (nr 33) z 1930 roku, wydawanych w liczącym się wówczas w świecie ośrodku naukowym prowadzącym badania logiczne. Zawiera konstruktywny dowód twierdzenia o pełności, w tym ogólną metodę dowodzenia praw rachunku kwantyfikatorów. Przedstawione w niej rozumowanie przypomina naszkicowany w *Logice z początku XX wieku* dowód twierdzenia o pełności dla rachunku zdań.

Przeglądając osiągnięcia Herbranda można zauważyć pewną paralelność z badaniami Kurta Gödla. Obaj uczeni zajmowali się podobnymi zagadnieniami, choć widać też pewne różnice. Największe osiągnięcia Gödla tamtego okresu mają wydźwięk negatywny: wykazywał on niemożność zrealizowania programu Hilberta. Herbrand – przeciwnie – starał się ten program usilnie realizować, a jego podobne rezultaty mają charakter pozytywny i są konstruktywne. Obaj dowiedli twierdzenie o pełności dla rachunku kwantyfikatorów, ale Herbrand zrobił to w sposób dający się zalgorytmizować.

Razem z twierdzeniem o zupełności arytmetyki Gödel opublikował też twierdzenie o niedowodliwości niesprzeczności arytmetyki. W odpowiedzi Herbrand dowiódł niesprzeczność fragmentu arytmetyki. W swoich rozważaniach Gödel posłużył się funkcjami pierwotnie rekurencyjnymi chyba nie zdając sobie sprawy, że wykorzystuje funkcje, które są obliczalne. W ostatniej pracy o niesprzeczności arytmetyki ukończonej tuż przed tragicznym wyjazdem w góry, która – tak się złożyło – dotarła do redakcji w dniu śmierci autora, Herbrand posłużył się – być może – już klasą funkcji rekurencyjnych i niewykluczone, że zdawał sobie sprawę z obliczalności tych funkcji. Wcześniej sygnalizował Gödłowi potrzebę i sensowność rozszerzenia definicji rekurencyjności.

Te ostatnie pomysły Herbranda są bardzo ciekawe dla osób interesujących się początkami informatyki. Jednak nie będziemy się nimi zajmować w tym artykule. Dokładniej zapoznamy się teraz z algorytmem pozwalającym automatycznie dowodzić twierdzenia.

Wprowadzenie do algorytmu

Swoje rozważania z pracy doktorskiej Herbrand rozpoczyna od charakteryzacji tautologii, w których wszystkie kwantyfikatory znajdują się na początku formuły i są to wyłącznie kwantyfikatory egzystencjalne. Takie formuły będziemy nazywać egzystencjalnymi. Herbrand zauważył, że

Lemat. Formuła egzystencjalna $\exists x \varphi$ (dla uproszczenia: z jednym kwantyfikatorem) jest tautologią wtedy i tylko wtedy, gdy tautologią rachunku zdań jest alternatywa

$$\varphi(t_1) \vee \varphi(t_2) \vee \dots \vee \varphi(t_n),$$

otrzymana w wyniku podstawienia w formule φ pewnych wyrażeń $t_1, t_2, \dots, t_n$, które tworzymy z symboli występujących w φ .

Trudno podać przykład ciekawej egzystencjalnej tautologii, choć za chwilę przeko-

² Jacques Herbrand Logical Writings, pod edycją Warrena Golfarba, D Reidel Publishing Company, 1971.

namy się, że jest ich bardzo dużo. Aby wyjaśnić spostrzeżenie Herbranda, zajmmy się tautologią ψ postaci

$$\exists x(A(x) \Rightarrow A(f(x))).$$

Formuła ta wyraża pewną własność bliżej nieokreślonych elementów i mówi coś o relacji oznaczanej symbolem A i funkcji nazywanej f .

Symbole A i f można interpretować na nieskończenie wiele sposobów. Można stworzyć nieskończenie wiele „światów”, w których formuła ψ – jako tautologia – powinna być prawdziwa. Herbrand zauważył, że wszystkie te „światy” można podzielić na skończenie wiele rodzajów i to tak, że w każdym ze „światów” ustalonego rodzaju jeden z elementów, których istnienie stwierdza formuła ψ , daje się wyrazić ustalonym wzorem, charakterystycznym dla „światów” tego rodzaju.

W przypadku formuły ψ każdy „świat” jest opisany przez wskazanie zbioru U elementów, których własności są wyrażane, jednoargumentowej relacji A_0 w zbiorze U , funkcji f_0 przekształcającej U w U oraz jednego z elementów x_0 należącego do U (będzie on najczęściej wartością zmiennej x). Są dwa rodzaje „światów”: takie, w których prawdziwa jest formuła $A(f(x))$ (w nich wartość funkcji f_0 dla argumentu x_0 jest w relacji A_0) i pozostałe. W „światach” pierwszego rodzaju jednym z elementów, których istnienie stwierdza formuła ψ , jest x_0 . Formuła $A(x) \Rightarrow A(f(x))$ okazuje się wtedy prawdziwa jako implikacja o prawdziwym następniku. W pozostałych „światach” jednym z takich elementów jest $f_0(x_0)$. Jeżeli zmienną x zinterpretujemy jako ten element, to formuła

$A(x) \Rightarrow A(f(x))$ będzie prawdziwa jako implikacja o fałszywym poprzedniku. Jeżeli jednak przyjmujemy, że wartością zmiennej x jest x_0 , to będzie prawdziwa formuła $A(f(x)) \Rightarrow A(f(f(x)))$. W każdym ze „światów” jest więc prawdziwa alternatywa

$$(A(x) \Rightarrow A(f(x))) \vee (A(f(x)) \Rightarrow A(f(f(x))))$$

Wobec tego jest to tautologia i to zarówno w sensie rachunku kwantyfikatorów, jak i rachunku zdań, podobnie jak każda formuła postaci $(p \Rightarrow q) \vee (q \Rightarrow r)$.

W dalszym ciągu swojej pracy doktorskiej Herbrand pokazuje, że dowolna formuła Ψ jest tautologią wtedy i tylko wtedy, gdy pewna, dająca się łatwo skonstruować formuła egzystencjalna jest tautologią (istnieje więc dużo egzystencjalnych tautologii). W dowodzie tej równoważności, po sprowadzeniu formuły Ψ do postaci normalnej (z kwantyfikatorami na początku formuły) Herbrand wykorzystuje przekształcenie zwane skolemizacją. Pierwszy raz tego typu przekształcenie zastosował norweski matematyk Thoralf Skolem.

Skolemizacja formuły $\forall x \exists y \varphi(x, y)$ polega na usunięciu pierwszego kwantyfikatora egzystencjalnego i zastąpieniu w formule $\varphi(x, y)$ zmiennej y wyrażeniem $f(x)$, gdzie f jest nowym, dotychczas niewykorzystywanym symbolem. Po wykonaniu tego przekształcenia z formuły $\forall x \exists y \varphi(x, y)$ otrzymujemy $\forall x \varphi(x, f(x))$. Ogólniej, między nawiasami w podstawianym wyrażeniu $f(x)$ umieszczamy wszystkie zmienne związane kwantyfikatorami ogólnymi poprzedzającymi usuwany. Opisaną operację powtarzamy tak długo, aż znikną wszystkie kwantyfikatory egzystencjalne. Jeżeli usuwany kwantyfikator znajduje się na początku formuły, to zamiast wyrażenia $f(x)$ podstawiamy po prostu f .

Skolemizacja ma pewne logiczne uzasadnienie. Zastępuje zdanie stwierdzające, że dla każdego x istnieje y o własności φ przez zdanie stwierdzające, że dla każdego x pewien, zależny od x element $f(x)$ ma własność φ . Zauważmy też, że po skolemizacji otrzymujemy zdanie zapisane w bogatszym języku, za pomocą większej liczby symboli.

Algorytm Herbranda

Dane: formuła, w naszym przypadku

$$(a) \quad \forall x(A(x) \Rightarrow B(x)) \Rightarrow \forall x(\neg B(x) \Rightarrow \neg A(x)).$$

Pytanie: czy dana formuła jest tautologią?

Krok 1: sprowadzamy formułę (a) do postaci normalnej.

Eliminujemy z niej wszystkie implikacje i skorzystamy z praw de Morgana:

$$\exists x(A(x) \wedge \neg B(x)) \vee \forall x(B(x) \vee \neg A(x)).$$

Zastępujemy zmienną x przez y :

$$\exists x(A(x) \wedge \neg B(x)) \vee \forall y(B(y) \vee \neg A(y)).$$

Przesuwamy kwantyfikatory na początek formuły:

$$(b) \quad \exists x \forall y((A(x) \wedge \neg B(x)) \vee B(y) \vee \neg A(y)).$$

Krok 2: negujemy formułę (b):

$$(c) \quad \forall x \exists y \neg((A(x) \wedge \neg B(x)) \vee B(y) \vee \neg A(y))$$

Krok 3: przeprowadzamy skolemizację formuły (c):

$$\forall x \neg((A(x) \wedge \neg B(x)) \vee B(f(x)) \vee \neg A(f(x))),$$

Krok 4: raz jeszcze negujemy wynik i otrzymujemy:

$$(d) \quad \exists x((A(x) \wedge \neg B(x)) \vee B(f(x)) \vee \neg A(f(x))),$$

Krok 5: w formule (d) opuszczamy kwantyfikator, a pozostałą część sprowadzamy do koniunkcyjnej postaci normalnej:

$$(e) \quad ((A(x) \vee B(f(x)) \vee \neg A(f(x))) \wedge (\neg B(x) \vee B(f(x)) \vee \neg A(f(x))))).$$

Krok 6: szukamy odpowiednich podstawień (w nieokreślony na razie sposób, na przykład przeglądając wszystkie możliwości): w rozważanym przykładzie podstawiamy za zmienną x samą zmienną x oraz wyrażenie $f(x)$ otrzymując

$$(f) \quad ((A(x) \vee B(f(x)) \vee \neg A(f(x))) \wedge (\neg B(x) \vee B(f(x)) \vee \neg A(f(x))))$$

oraz

$$(g) \quad ((A(f(x)) \vee B(f(f(x))) \vee \neg A(f(f(x)))) \wedge (\neg B(f(x)) \vee B(f(f(x))) \vee \neg A(f(f(x))))).$$

Krok 7: sprawdzamy, czy alternatywa koniunkcji (f) i (g) jest tautologią, na przykład sprowadzając ją do koniunkcyjnej postaci normalnej. Aby przekonać się, że jest tautologią, wystarczy zaobserwować, że cztery alternatywy utworzone z członów koniunkcji (f) i (g) są tautologiami.

Krok 8: Jeżeli przekonamy się, że alternatywa formuł (f) i (g) jest tautologią (rachunku zdań), to stwierdzamy, że formuła (a) jest tautologią (rachunku kwantyfikatorów).

Łącząc te dwa elementy otrzymujemy

Algorytm Herbranda

Został on przedstawiony na rysunku 1 na przykładzie formuły (a)

$$\forall x(A(x) \Rightarrow B(x)) \Rightarrow \forall x(\neg B(x) \Rightarrow \neg A(x)),$$

która jest prostszą wersją jednej z tautologii przytoczonych w artykule *Logika z początku XX wieku*. Aby zbadać, czy rzeczywiście jest to tautologia rachunku kwantyfikatorów, najpierw sprowadzamy ją do postaci normalnej, w której wszystkie kwantyfikatory znajdują się na początku. Formuła (b) jest postacią normalną formuły (a).

Istotnym krokiem algorytmu jest skolemizacja. W przypadku formuły (c) polega ona na usunięciu pierwszego kwantyfikatora egzystencjalnego i zastąpieniu w reszcie formuły związanej tym kwantyfikatorem zmiennej y wyrażeniem $f(x)$.

Formuła (a) jest tautologią wtedy i tylko wtedy, gdy formuła (d) jest tautologią. Nie podejmuję się uzasadnić teraz tego faktu. Może on budzić wątpliwości. Wykonywane przekształcenia są bardzo formalne, a ich sens logiczny nie jest widoczny. Fakt ten jest jednak prawdziwy. Sam algorytm też można skrócić usuwając dwukrotne negowanie. Dzięki negowaniu został przedstawiony za pomocą skolemizacji. Przy okazji zauważmy, że w podany sposób można skonstruować wiele egzystencjalnych tautologii.

Dalej musimy jeszcze zbadać, czy formuła (d) jest tautologią. W tym celu korzystamy ze spostrzeżenia Herbranda charakteryzującego tautologie będące formułami egzystencjalnymi. Wymaga to znalezienia tautologii określonej postaci. Metoda szukania takiej tautologii nie została podana, ale w ten sposób Herbrand sprowadził badanie, czy formuła jest tautologią rachunku kwan-

tyfikatorów, do sprawdzania, czy pewna szczególna forma zdaniowa jest tautologią.

Spróbujmy jeszcze przekonać się metodą Herbranda, że formuła $\exists x \forall y R(x, y)$ nie jest tautologią. Nie trzeba jej sprowadzać do postaci normalnej. Po zanegowaniu, skolemizacji i ponownym negowaniu otrzymujemy formułę $\exists x R(x, f(x))$. Wykonując podstawienia i tworząc alternatywę podstawień otrzymujemy formuły takie, jak

$$R(x, f(x)) \vee R(ff(x), fff(x)) \vee R(fff(x), ffff(x))$$

(pomiąłem niektóre nawiasy). Aby przekonać się, że formuła $\exists x \forall y R(x, y)$ nie jest tautologią, wystarczy zauważyć, że w tak otrzymywanych alternatywach nie pojawi się negacja elementarnego członu, a w tautologiach – przeciwnie – musi wystąpić³.

Twierdzenie o pełności wg Herbranda

Swój algorytm Herbrand zastosował w dowodzie twierdzenia o pełności. Istotną część tego twierdzenia polega na wykazaniu, że tautologie są prawami rachunku kwantyfikatorów. Mając tautologię Ψ , Herbrand wykonuje algorytm i najpierw tworzy równoważną jej formułę egzystencjalną $\exists x \varphi$. Następnie znajduje alternatywę $\varphi(t_1) \vee \varphi(t_2) \vee \dots \vee \varphi(t_n)$ będącą tautologią rachunku zdań. Zgodnie z twierdzeniem o pełności dla rachunku zdań ta alternatywa ma dowód. Nietrudno zauważyć, że wynika z niej formuła $\exists x \varphi$. Dalej Herbrand konstruktywnie pokazał, jak z tej formuły wynika formuła Ψ . Ten fragment dowodu wymaga posłużenia się odpowiednimi własnościami skolemizacji.

Twierdzenie o pełności pozwala inaczej spojrzeć na algorytm Herbranda. Algorytm ten nie tylko bada, czy dana formuła jest

³ Także ta kwestia została wyjaśniona w artykule *Logika z początku XX wieku*.

tautologią rachunku kwantyfikatorów. Umożliwia także sprawdzanie, czy interesująca nas formuła jest prawem rachunku kwantyfikatorów. Wobec twierdzenia o dedukcji (znalezione go także przez Herbranda), pozwala w wielu przypadkach badać, czy dana formuła jest twierdzeniem.

Nierozstrzygalność rachunku kwantyfikatorów

Zarówno algorytm Herbranda, jak i opisany na początku artykułu naiwny algorytm nie pozwalają rozstrzygać, czy dana formuła jest twierdzeniem. Herbrand nie podał, jak w 6. kroku algorytmu efektywnie znaleźć potrzebne podstawienia $t_1, t_2, \dots, t_n$. Najprostsza metoda znowu polega na przeszukaniu wszystkich możliwych podstawień, a jest ich nieskończenie wiele. Takie poszukiwanie może się nigdy nie zakończyć. Mimo to Herbrand bardzo starannie przeanalizował pojęcie twierdzenia i uzyskał wszystko, co można było uzyskać. Wspomnianej wady obu algorytmów nie da się poprawić.

Właściwie powinno być to wiadomo już w momencie opublikowania pracy Herbranda. Wtedy jednak nikt nie potrafił skojarzyć ustalonych faktów, nikt też nie zdecydował się na sformułowanie twierdzenia o nierozstrzygalności rachunku kwantyfikatorów. Zgodnie z dzisiejszą wiedzą, z pracy Kurta Gödla z 1931 roku o niezupełności arytmetyki i z twierdzenia Herbranda o dedukcji opublikowanego w 1930 roku wynika, że zbiór praw rachunku kwantyfikatorów jest nierozstrzygalny. W tamtych czasach tego twierdzenia nie można było jednak dowieść. Głównie dlatego, że nikt nie potrafił precyzyjnie zdefiniować ani pojęcia algorytmu, ani problemu rozstrzygalnego.

Te pojęcia dopiero się kształtowały. Kurt Gödel posłużył się już funkcjami pierwotnie rekurencyjnymi, ale prawdopodobnie nie zdawał sobie sprawy z tego, że był bliski zdefiniowania pojęcia funkcji obliczal-

nej i dowiódł ważne zastosowania obliczalności. Jacques Herbrand zaproponował rozszerzenie definicji Gödla i był bliski zdefiniowania klasy funkcji rekurencyjnych, która jest jedną z formalizacji obliczalności (to także jest znaczącym osiągnięciem tego matematyka). Mało znane badania obliczalności prowadził jeszcze w dwudziestych latach XX w. Emil Post. Dopiero w 1936 roku Alonzo Church i niezależnie Alan Turing podali precyzyjną definicję obliczalności⁴.

Sformalizowanie pojęcia obliczalności umożliwiło Alonzo Churchowi inne odczytanie pracy Gödla o niezupełności arytmetyki. Zauważył on, że rozumowanie Gödla dowodzi nie tylko niezupełności arytmetyki Peano, ale także uzasadnia nierozstrzygalność zbioru twierdzeń tej arytmetyki z nieskończonym zbiorem aksjomatów. I to właśnie jest powodem jej niezupełności: teoria zupełna ma bowiem rozstrzygalny zbiór twierdzeń. Nieco później w dość zawiły sposób wywnioskował stąd nierozstrzygalność rachunku kwantyfikatorów, a więc dowiódł, że nie ma algorytmu, który po przeanalizowaniu formuły udziela poprawnej i jednoznacznej odpowiedzi na pytanie, czy ta formuła jest, czy też nie jest prawem rachunku kwantyfikatorów.

Ostateczny kształt dowodowi twierdzenia o nierozstrzygalności rachunku kwantyfikatorów nadał Raphael Mitchel Robinson, który zauważył, że dowód twierdzenia Gödla o niezupełności arytmetyki, a także dowód twierdzenia o nierozstrzygalności arytmetyki można powtórzyć dla pewnej arytmetyki ze skończonym zbiorem aksjomatów. Aksjomaty tej arytmetyki można znaleźć w artykule *Początki informatyki teoretycznej*⁵, a Robinson oznaczał ją po prostu literą Q . Znając osiągnięcia Robinsona twierdzenie o nierozstrzygalności rachunku kwantyfikatorów możemy łatwo

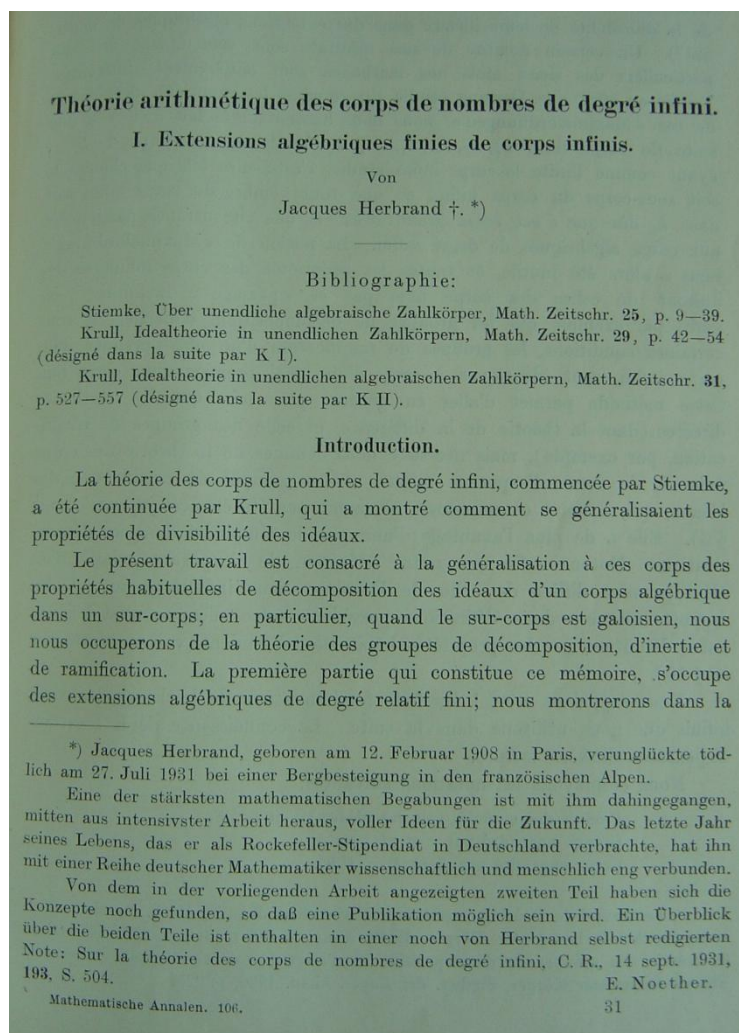
⁴ Definicja Turinga jest przedstawione w artykule *Turing i jego maszyna* zamieszczonym w „Matematyce” 6/2006

⁵ „Matematyka” 1/2006

wyprowadzić z pracy Gödla posługując się twierdzeniem o dedukcji.

Mimo że algorytm Herbranda nie w każdym przypadku rozstrzyga, czy dana formuła jest prawem logiki, jest wykorzystywany w programach umożliwiającym automatyczne dowodzenie twierdzeń. Jedne z

pierwszych eksperymentów z tym algorytmem zostały przeprowadzone w IBM Research Center w 1959 roku przez Paula Gilmora i odegrały ważną rolę w rozwoju informatyki.



Strona tytułowa ostatniej algebraicznej pracy Herbranda z notatką o autorze pióra Emmy Nöther