Podstawy matematyki dla informatyków

Materiały do wykładu dla I roku informatyki

P. Urzyczyn
urzy@mimuw.edu.pl

27 września 2017, godzina 13:51

1 Język logiki matematycznej

Zadaniem matematyki jest badanie rozmaitych abstrakcyjnych obiektów, odkrywanie ich własności i analizowanie związków pomiędzy tymi własnościami. Formułowanie myśli i stwierdzeń oraz wiązanie ich ze sobą w sposób ścisły i jednoznaczny, a jednocześnie zrozumiały, ma tu pierwszorzędne znaczenie. Aby *odpowiednie dać rzeczy słowo* posługujemy się umownymi skrótami i konwencjami. Jedną z takich konwencji jest używanie ustalonych *spójników logicznych* i *kwantyfikatorów* do budowania osądów (stwierdzeń, zdań) złożonych z prostszych wyrażeń. W ten sposób znaczenie złożonego osądu (zdania) jest jednoznacznie określone przez znaczenie jego składowych. Skróty stosowane przy tej okazji nazywamy *notacją logiczną*.

Na przykład, jeśli A i B są pewnymi stwierdzeniami, to wyrażenie „A *i* B" (zapisywane w skrócie jako $A \wedge B$) orzeka, że ma miejsce zarówno stan rzeczy opisany przez A jak i przez B. Wyrażenie to nazywamy *koniunkcją* osądów A i B. W języku polskim koniunkcji odpowiada słowo i, ale także każde ze słów *oraz*, *a*, *ale*, różniących się odcieniami znaczeniowymi. Te różnice znaczeniowe znikają w języku matematyki, gdzie koniunkcja stanowi tylko suche stwierdzenie koincydencji dwóch faktów.

Alternatywa osądów A i B to wyrażenie „A *lub* B" (w skrócie $A \vee B$). Stwierdza ono zajście co najmniej jednej z możliwości, może A, może B, a być może obu. W języku polskim alternatywie odpowiadają słowa *lub* i *albo*[1] ale także na przykład zwrot „*A, chyba że B*".

Sens koniunkcji i alternatywy jest dosyć oczywisty i na ogół zgodny ze sposobem w jaki w języku polskim używamy słów *i* oraz *lub*. Znacznie mniej jasne jest jak należy ściśle interpretować stwierdzenie postaci „*jeśli A to B*", czyli *implikację* (w skrócie $A \to B$). Chcemy oczywiście powiedzieć, że $A \to B$ wyraża wynikanie stwierdzenia B ze stwierdzenia A, ale co to naprawdę znaczy „wynikanie"?

[1] Czasami słowo *albo* używane jest w znaczeniu tzw. alternatywy wykluczającej, ale my nie będziemy stosować tej zasady.

Wynikanie w matematyce to zwykle tzw. *implikacja materialna*, wyrażająca jedynie „obserwacyjną" zależność pomiędzy przesłanką i konkluzją: implikacja $A \to B$ zachodzi wtedy, gdy zajściu A z pewnością towarzyszy B. Przy tym, jeśli A nie ma miejsca, to implikację akceptujemy „walkowerem". Ścisłą definicję implikacji materialnej można podać, odwołując się do koncepcji *wartości logicznej*. Otóż przyjmujemy, że każde poprawnie zbudowane i jednoznacznie sformułowane wyrażenie o charakterze orzekającym (zdanie logiczne) jest albo prawdziwe albo fałszywe.[2] Inaczej, każde zdanie logiczne ma *wartość logiczną*: jest nią prawda (oznaczana zwykle przez 1) lub fałsz (oznaczany przez 0). W naszej dwuwartościowej logice (którą nazywamy też *logiką klasyczną*) znaczenie spójników logicznych można opisywać dobrze znanymi tabelkami. Dla koniunkcji i alternatywy tabelka wygląda tak:

A	B	$A \wedge B$	$A \vee B$
0	0	0	0
0	1	0	1
1	0	0	1
1	1	1	1

Natomiast idea implikacji materialnej może być wyrażona tak:

A	B	$A \to B$
0	0	1
0	1	1
1	0	0
1	1	1

Implikacja jest więc fałszywa tylko wtedy, gdy przesłanka jest prawdziwa, a konkluzja fałszywa. W pozostałych przypadkach musimy uznać implikację za prawdziwą.

Jak już powiedzieliśmy, wartość logiczna, którą przypisujemy implikacji $A \to B$ zależy wyłącznie od wartości logicznych przypisanych jej przesłance A i konkluzji B. Wartość ta nie zależy natomiast od samej treści tych wyrażeń, czy też jakichkolwiek innych związków pomiędzy A i B. W szczególności, wypowiedzi A i B mogą mówić o zajściu jakichś zdarzeń i wtedy wartość logiczna implikacji materialnej $A \to B$ nie ma nic wspólnego z ich ewentualnym następstwem w czasie, lub też z tym, że jedno z tych zdarzeń spowodowało drugie. W języku polskim stwierdzenie *„jeśli A to B"* oczywiście sugeruje taki związek, np. w zdaniu:

Jeśli zasilanie jest włączone, to drukarka działa.

Ale przecież implikacja materialna nie zachodzi, o czym dobrze wiedzą użytkownicy drukarek. Co więcej, zwykle materialną prawdą jest stwierdzenie odwrotne:

Jeśli drukarka działa, to zasilanie jest włączone.

Natomiast zdanie

Drukarka działa, ponieważ zasilanie jest włączone,

[2] Wyrażenie zawierające zmienne, jak np. „$3 + x = y$", można zinterpretować jako prawdziwe lub fałszywe, gdy określone są wartości zmiennych.

stwierdza związek przyczynowo-skutkowy, a ponadto faktyczne zajście wymienionych zdarzeń, a to nie daje się wyrazić za pomocą materialnej implikacji.

Następny ważny spójnik logiczny to *negacja*. Mówimy „*nieprawda, że A*", i piszemy w skrócie $\neg A$, gdy chcemy powiedzieć, że A nie ma miejsca, tj. że przypuszczenie A prowadzi do sprzeczności (fałszu, absurdu). Jeśli sam absurd (zdanie wzorcowo fałszywe) oznaczymy przez $\bot$, to negacja $\neg A$ jest tym samym co implikacja $A \to \bot$. W logice dwuwartościowej negację opisujemy tabelką:

A	$\neg A$
0	1
1	0

W myśl tej tabeli, jedno ze stwierdzeń A i $\neg A$ musi być prawdziwe; zasada ta, zapisywana jako $A \vee \neg A$, nosi nazwę prawa wyłączonego środka (*tertium non datur*).

Pozostaje jeszcze *równoważność* $A \leftrightarrow B$, którą czytamy „A wtedy i tylko wtedy, gdy B". Równoważność wyraża tę samą myśl co koniunkcja dwóch implikacji: $(A \to B) \wedge (B \to A)$, a więc uznamy ją za prawdziwą, gdy wartości logiczne A i B są takie same. Na przykład równoważność $A \leftrightarrow \neg\neg A$ jest zawsze prawdziwa.

Zwróćmy jeszcze uwagę na utarte w matematyce znaczenie pewnych zwrotów języka polskiego:

- Zdanie „*A, tylko (wtedy) gdy B*" odpowiada implikacji $A \to B$, natomiast zdaniem „*A wtedy, gdy B*" stwierdzamy implikację $B \to A$.
- Implikacja $A \to B$ jest nazywana implikacją *odwrotną* do $B \to A$.
- Gdy zachodzi implikacja $A \to B$, to mówimy, że A jest warunkiem *wystarczającym* na B, natomiast B nazywamy warunkiem *koniecznym* dla A.
- Jeśli zaś stwierdzamy, że $A \leftrightarrow B$, to możemy powiedzieć, że A jest warunkiem *koniecznym i wystarczającym* na B. (Oczywiście wtedy B jest też warunkiem koniecznym i wystarczającym na A.)

Nawiasy: Zasady użycia nawiasów w wyrażeniach logicznych są określone przez *priorytety*, które przypisujemy spójnikom. Najsilniej wiąże negacja, potem (równorzędnie) koniunkcja i alternatywa, a najniższy priorytet mają implikacja i równoważność. Zatem na przykład wyrażenie $\neg A \vee B \to C$ oznacza to samo co $((\neg A) \vee B) \to C$, a napis $A \vee B \wedge C$ jest niepoprawny, bo dwuznaczny.

Kwantyfikatory

Słowo *predykat* oznacza wyrażoną w jakimś języku „własność" lub „relację" odnoszącą się do pewnych obiektów. Na przykład w zdaniu „*Liczba 5 jest parzysta*" rolę predykatu odgrywa słowo *parzysta*, a w zdaniu „$2 < 3 \to 2 = 1$" rolę tę pełnią relacje równości i mniejszości oznaczone znakami $=$ i $<$. Predykat może mieć dowolną liczbę argumentów i może być złożony: formuła $x < 3 \to x = 1$ określa pewien predykat odnoszący się do zmiennej x.

Z *logiką predykatów* mamy do czynienia wtedy, gdy pytamy o własności zdań zawierających predykaty. Zdania takie mogą być budowane z pomocą spójników logicznych $\wedge$, $\vee$, $\rightarrow$, $\neg$ i $\leftrightarrow$, ale także przy użyciu *kwantyfikatorów*.

Przypuśćmy, że $A(x)$ wyraża pewną własność obiektów x należących do dziedziny $\mathcal{D}$. Jeśli chcemy stwierdzić, że wszystkie elementy x dziedziny $\mathcal{D}$ mają własność $A(x)$, to możemy napisać $\forall x{\in}\mathcal{D}\, A(x)$ lub $\forall x{:}\mathcal{D}\, A(x)$. Czytamy to zwykle tak: „*Dla każdego x należącego do $\mathcal{D}$ zachodzi $A(x)$*" albo tak: „*Dla każdego x typu $\mathcal{D}$ zachodzi $A(x)$*". Znak $\forall$ nazywamy *kwantyfikatorem ogólnym* lub *uniwersalnym*.

Natomiast *kwantyfikator szczegółowy*, inaczej *egzystencjalny*, $\exists$, służy do wyrażania stwierdzeń postaci „*Dla pewnego x z dziedziny $\mathcal{D}$ zachodzi $A(x)$*", które zapisujemy tak: $\exists x{\in}\mathcal{D}\, A(x)$ lub tak: $\exists x{:}\mathcal{D}\, A(x)$.

Dziedzinę $\mathcal{D}$, którą przebiegają wartości zmiennej x, często traktujemy jako domyślną i po prostu piszemy $\exists x\, A(x)$. Podobnie postępujemy z kwantyfikatorem $\forall$. Inne popularne uproszczenie polega na pisaniu np. $\forall xy{:}\mathcal{D}\ldots$ zamiast $\forall x{:}\mathcal{D}\forall y{:}\mathcal{D}\ldots$

Wartości logicznych wyrażeń kwantyfikatorowych nie da się zdefiniować za pomocą tabelek. Na przykład dlatego, że dziedzina $\mathcal{D}$ może być nieskończona. Ale też dlatego, że znaczenie danego zdania zależy tutaj od znaczenia wszystkich występujących w nim symboli (predykatów, nazw obiektów i funkcji). Na przykład sens zdania $\forall x{:}\mathcal{D}(f(x) < 0 \rightarrow x < 1)$ zależy od dziedziny $\mathcal{D}$ i od tego co oznaczają symbole f, $<$, 0, 1. Możemy tylko powiedzieć, że:

- Stwierdzenie $\forall x{:}\mathcal{D}\, \varphi(x)$ jest spełnione wtedy i tylko wtedy, gdy wszystkie elementy $d \in \mathcal{D}$ mają własność $\varphi(d)$.
- Stwierdzenie $\exists x{:}\mathcal{D}\, \varphi(x)$ jest spełnione wtedy i tylko wtedy, gdy przynajmniej jeden element $d \in \mathcal{D}$ ma własność $\varphi(d)$.

Nawiasy: Istnieją dwie tradycje nawiasowania wyrażeń z kwantyfikatorami. Pierwsza nadaje kwantyfikatorom najwyższy priorytet, tj. formułę $\forall x\, P(x) \rightarrow R(y)$ należy rozumieć tak samo jak $(\forall x\, P(x)) \rightarrow R(y)$. W *zasięgu kwantyfikatora* znajduje się jedynie najbliższy człon formuły. Druga tradycja rozciąga zasięg kwantyfikatora najdalej jak to możliwe, wtedy jednak po zmiennej zwiazanej kwantyfikatorem należy postawić kropkę, która niejako zastępuje nawias. Można więc napisać $\forall x.\, P(x) \rightarrow R(y)$ zamiast $\forall x(P(x) \rightarrow R(y))$.

Zmienne wolne i związane: Jak już mówiliśmy, znaczenie zmiennych występujących w danym stwierdzeniu ma wpływ na jego wartość logiczną. Na przykład warunek „$x > 4$" jest spełniony, gdy wartością x jest liczba 5. Ale ocena prawdziwości każdego ze zdań „$\forall x{:}\mathbb{N}.\, x > 4$" i „$\exists x{:}\mathbb{N}.\, x > 4$" nie wymaga określenia wartości x. W tych zdaniach zmienna x została *związana* kwantyfikatorem. Zdanie „$\exists x{:}\mathbb{N}.\, x > 4$" nie wyraża już żadnej własności liczby x, a raczej własność relacji większości w zbiorze $\mathbb{N}$. Równie dobrze zamiast „$\exists x{:}\mathbb{N}.\, x > 4$" moglibyśmy przecież napisać „$\exists z{:}\mathbb{N}.\, z > 4$". Albo powiedzieć po polsku „pewna liczba naturalna jest większa od 4".

A zatem wartość logiczna formuły zależy tylko od zmiennych, które nie są związane kwantyfikatorami. Nazywamy je *zmiennymi wolnymi*. Na przykład w formule $\exists x{:}\mathbb{N}(x < 4 \wedge y \leq x)$ zmienna x jest związana a zmienna y jest wolna. Formuła ta wyraża więc pewien predykat odnoszący się do y (ten sam, co $\exists u{:}\mathbb{N}.u < 4 \wedge y \leq u$ i ten sam, co $y \leq 3$).

Zauważmy jednak, że wiązanie zmiennych przez kwantyfikator odnosi się tylko do tego pod-

wyrażenia, którego dotyczy ten kwantyfikator. Wystąpienia zmiennej poza *zasięgiem* kwantyfikatora pozostają wolne. A więc ta sama zmienna może być zarówno wolna jak i związana, jak np. zmienna x w formule $(\forall x{:}\mathbb{N}.\, x > 0 \vee x \leq y) \to x = 1$. Tutaj wolne są zarówno zmienne y jak i x (w swoim trzecim wystąpieniu), a więc wartość formuły zależy od nich obu.

Aby uniknąć problemów z interpretacją wyrażeń zawierających zmienne wolne i związane, pamiętajmy o tym, że znaczenie takiego wyrażenia nie zależy od wyboru zmiennych związanych. Można więc dobrać zmienne związane w ten sposób, aby się nie myliły ze zmiennymi wolnymi, np. naszą formułę napiszemy tak: $(\forall z{:}\mathbb{N}.\, z > 0 \vee z \leq y) \to x = 1$.

Zjawisko wiązania zmiennych występuje nie tylko w wyrażeniach o charakterze logicznym. Na przykład całkę $\int_x^{x+1} x^2\ dx$ powinniśmy rozumieć tak samo jak $\int_x^{x+1} y^2\ dy$, bo zmienna x w wyrażeniu x^2 (ale nie w granicach całkowania) jest związana przez dx. Identyfikatory lokalne w programowaniu to także nic innego jak zmienne związane (swoimi deklaracjami), a identyfikatory globalne odpowiadają zmiennym wolnym.

Konfuzje składniowe

Język formuł matematycznych rządzi się nieco innymi prawami niż język polski (i każdy inny język naturalny). Ma swoje własne reguły składniowe, dopuszczające znacznie mniejszą dowolność interpretacyjną. Tłumacząc zdania języka polskiego na język matematyki (i odwrotnie), należy o tym pamiętać. Na przykład te dwa zdania mają bardzo podobną budowę:

Każdy kot ma wąsy.
Pewien kot ma wąsy.

Można je „przetłumaczyć" na język logiki tak:

$\forall x{:}Kot.\, MaWąsy(x);$
$\exists x{:}Kot.\, MaWąsy(x),$

ale czasem robi się to inaczej, i wtedy podobieństwo znika:

$\forall x(Kot(x) \to MaWąsy(x));$
$\exists x(Kot(x) \wedge MaWąsy(x)).$

Dość częstym błędem jest właśnie mylenie koniunkcji z implikacją w zasięgu działania kwantyfikatora. A oto inny przykład. Zdania:

Liczba 6 *jest parzysta;*

Liczba 6 *jest dwukrotnością pewnej liczby,*

oznaczają to samo. Zaprzeczeniem pierwszego z nich jest oczywiście zdanie

Liczba 6 *nie jest parzysta,*

ale zaprzeczeniem drugiego nie jest zdanie

Liczba 6 *nie jest dwukrotnością pewnej liczby,*

otrzymane przecież przez analogiczną operację „podstawienia". Użycie słowa „pewnej" powoduje bowiem, że to zdanie rozumiemy jako $\exists x(\neg 6 = 2x)$, a nie jako $\neg\exists x(6 = 2x)$.

Innym popularnym błędem jest mylenie koniunkcji z alternatywą w przesłance implikacji, zwłaszcza gdy występuje tam negacja. Mamy bowiem skłonność do powtarzania słowa „nie" w obu członach założenia i nie razi nas zdanie

Kto nie ma biletu lub nie jest pracownikiem teatru, ten nie wejdzie na przedstawienie.

Ale od tekstu matematycznego oczekujemy więcej ścisłości i w takim tekście zdanie:

Jeśli x *nie jest równe* 2 *lub nie jest równe* 3*, to* $x^2 - 5x + 6$ *nie jest zerem.*

może wprowadzić czytelnika w błąd. „Dosłowne" tłumaczenie tego zdania na język logiki predykatów, to przecież formuła

$$\neg(x = 2) \vee \neg(x = 3) \to \neg(x^2 - 5x + 6 = 0),$$

a nie formuła

$$\neg(x = 2 \vee x = 3) \to \neg(x^2 - 5x + 6 = 0).$$

Wielu takich dwuznaczności unikniemy, gdy przypomnimy sobie, że w języku polskim istnieją takie słowa jak *ani* i *żaden*.

Cena jaką płacimy za ścisłość języka matematyki, to często pewne ograniczenia i utrudnienia. Zastanówmy się jak w języku logiki predykatów wyrazić stwierdzenie:

Jeśli Joe ma osła, to go bije.

Chciałoby się napisać tak: $(\exists x{:}\mathit{Osioł}.\, \mathit{Ma}(\mathit{Joe}, x)) \to \mathit{Bije}(\mathit{Joe}, x)$. Ale ta formuła jest niedobra, bo zasięg kwantyfikatora obejmuje tylko przesłankę implikacji i zmienna x jest wolna w konkluzji. Zdanie $\exists x{:}\mathit{Osioł}\,(\mathit{Ma}(\mathit{Joe}, x) \to \mathit{Bije}(\mathit{Joe}, x))$ jest tym bardziej nie na temat. Jest ono „walkowerem" prawdziwe, bo na pewno istnieją osły, których Joe nie ma. Aby rozwiązać problem bitego osła musimy (wbrew intuicji) użyć kwantyfikatora... ogólnego:

$$\forall x{:}\mathit{Osioł}\,(\mathit{Ma}(\mathit{Joe}, x) \to \mathit{Bije}(\mathit{Joe}, x)).$$

2 Typy i zbiory

Matematycy chętnie posługują się językiem teorii zbiorów. Trudno sobie wyobrazić współczesny tekst matematyczny, w którym zbiory nie pojawiają się w taki, czy inny sposób. Dlatego w naszym wykładzie też poświęcimy im dużo miejsca. Nie sposób jest podać ścisłej definicji tak pierwotnego pojęcia jakim jest zbiór. Georg Cantor, twórca teorii zbiorów (zwanej też teorią mnogości) próbował zrobić to tak:

Zbiorem nazywamy zgromadzenie w jedną całość wyraźnie wyróżnionych przedmiotów naszej intuicji lub naszej myśli.

Sens definicji Cantora jest taki: Jeśli potrafimy wyodrębnić pewne przedmioty za pomocą jakiegoś kryterium wyboru (predykatu), to te przedmioty tworzą dobrze określony zbiór. A więc zbiór to w istocie „upostaciowienie", albo „materializacja" pewnego predykatu. Jest to wygodny skrót myślowy: zamiast mówić o wszystkich przedmiotach x, spełniających kryterium $K(x)$, wygodniej rozważać tylko jeden przedmiot – zbiór z nich złożony. Na oznaczenie tego zbioru można użyć notacji $\{x \mid K(x)\}$. Zauważmy, że x jest tu związane, tj. $\{x \mid K(x)\}$ to to samo co $\{y \mid K(y)\}$.

Kłopoty ze zbiorami

Na co dzień pojęcie zbioru służy nam właśnie jako wygodny skrót myślowy. Ale jeśli raz zgodziliśmy się traktować zbiory tak jak wszystkie inne przedmioty, musimy się też zgodzić na konsekwencje, na przykład na zbiory zbiorów. W „naiwnej teorii mnogości" można było na przykład rozważać zbiór wszystkich zbiorów: $Z = \{x \mid x \text{ jest zbiorem}\}$. Oczywiście taki zbiór musiałby być swoim własnym elementem (co zapiszemy tak: $Z \in Z$). To jeszcze nic złego, ale co począć z takim zbiorem:

$$R = \{x \mid x \text{ jest zbiorem i } x \notin x\} \ ?$$

Niebezpieczne pytanie: czy $R \in R$? Jeśli $R \in R$, to R musi spełniać warunek:

$$\text{„}R \text{ jest zbiorem i } R \notin R\text{".} \qquad (*)$$

A jeśli $R \notin R$, to (*) nie zachodzi i mamy $R \in R$. Tak czy owak, jest źle!

Powyższe rozumowanie, zwane *antynomią Russella*, wskazuje na to, że „naiwne" pojmowanie zbiorów prowadzi do sprzeczności. Ale nie wynika stąd, że cała teoria zbiorów jest bezużyteczna. Przeciwnie, pojęcie zbioru jest wygodne i potrzebne, kłopoty pojawiają się wtedy, gdy go nadużywamy.

Gdy mówimy o kryterium odróżniającym jakieś obiekty od innych, musimy bowiem pamiętać, że nie każde kryterium $K(x)$ ma sens dla dowolnego x. Wartości zmiennej x w sposób jawny lub domyślny przebiegają zawsze jakąś konkretną dziedzinę $\mathcal{D}$ (powiemy, że x jest *typu* $\mathcal{D}$). Zamiast $\{x \mid K(x)\}$ powinniśmy więc raczej napisać $\{x{:}\mathcal{D} \mid K(x)\}$ lub $\{x \in \mathcal{D} \mid K(x)\}$.

Przy tej okazji zauważmy, że cała dziedzina $\mathcal{D}$ też jest zbiorem (wyróżnionym z siebie samej przez trywialne, zawsze spełnione, kryterium). Przestrzeń, z której wyodrębnia się zbiory, jest jednak na ogół tworem bardziej „pierwotnym" niż jakikolwiek predykat odnoszący się do

elementów tej przestrzeni. Dlatego takie dziedziny będziemy często nazywać *typami*. Można na przykład mówić o typie $\mathbb{N}$ liczb naturalnych, typie $\mathbb{R}$ liczb rzeczywistych, czy też o typie Bool wartości logicznych, którego elementami są *prawda* i *fałsz*.

Uwaga: Każdemu obiektowi powinniśmy w zasadzie jednoznacznie przypisać jego typ. Czasem jednak naturalne jest to, że obiekt danego typu $\mathcal{D}$ w pewnych sytuacjach może być uważany za obiekt innego typu $\mathcal{E}$. Tak jest na przykład z liczbami naturalnymi, które mogą być też uważane za liczby rzeczywiste. Wówczas mówimy o tym, że typ $\mathcal{D}$ jest *podtypem* typu $\mathcal{E}$, i o (domyślnej lub jawnej) *koercji* lub *konwersji* elementów typu $\mathcal{D}$ w elementy typu $\mathcal{E}$.

Podzbiory

Jak powiedziano wyżej, zbiór złożony dokładnie z tych elementów typu $\mathcal{D}$, które spełniają warunek $K(x)$, oznaczamy przez $\{x : \mathcal{D} \mid K(x)\}$ lub $\{x \in \mathcal{D} \mid K(x)\}$. Napis „$y \in A$" czytamy „$y$ jest elementem zbioru A". A więc dla $y : \mathcal{D}$ mamy równoważność:

$$K(y) \;\leftrightarrow\; y \in \{x{:}\mathcal{D} \mid K(x)\}.$$

Zamiast $\{x : \mathcal{D} \mid x \in A \wedge K(x)\}$ piszemy po prostu $\{x \in A \mid K(x)\}$. Napis $\{x \mid K(x)\}$ ma zaś sens wtedy, gdy wiadomo jakiego typu są elementy.

Innym sposobem zdefiniowania zbioru (ale tylko skończonego) jest bezpośrednie wyliczenie: na przykład zbiór, którego elementami są liczby 1,2 i 7, oznaczymy przez $\{1, 2, 7\}$. Podobnie, napis $\{\{0\}, \{1, 2\}, \{1, 7\}\}$ oznacza zbiór o trzech elementach $\{0\}, \{1, 2\}, \{1, 7\}$, które same są zbiorami. Ogólnie, zbiór o elementach $x_1, \ldots, x_n$ oznaczymy przez $\{x_1, \ldots, x_n\}$. W szczególności $\{x\}$ oznacza *singleton* x, tj. zbiór, którego jedynym elementem jest x. Zgodnie z tą konwencją, napis $\{\ \}$ oznacza zbiór, który nie ma żadnego elementu, czyli *zbiór pusty*. Taki zbiór jest jednak częściej oznaczany symbolem $\varnothing$.

Czasami definiujemy też zbiory przez *zastępowanie*. Jeśli każdemu elementowi x jakiegoś zbioru A potrafimy jednoznacznie przypisać jakiś obiekt a_x ustalonego typu, to zbiór wszystkich takich obiektów oznaczymy przez $\{a_x \mid x \in A\}$. Na przykład zbiór $\{|x| \mid x \in (-1, 3)\}$ to przedział $[0, 3)$.

Uwaga: z tego, że $a_y \in \{a_x \mid x \in A\}$ nie wynika, że $y \in A$. Na przykład liczba $|-2|$ należy do zbioru $\{|x| \mid x \in (-1, 3)\}$, bo $|-2| = |2|$, ale przecież $-2 \notin (-1, 3)$.

Mówimy, że zbiór A jest *zawarty* w zbiorze B (lub, że jest jego *podzbiorem*) wtedy i tylko wtedy, gdy zachodzi warunek $\forall z(z \in A \to z \in B)$. Piszemy wówczas „$A \subseteq B$". W szczególności, każdy zbiór złożony z elementów typu $\mathcal{D}$ jest podzbiorem typu $\mathcal{D}$. Podzbiory takie tworzą typ $\mathsf{P}(\mathcal{D})$ nazywany *typem potęgowym* typu $\mathcal{D}$. Ogólniej, jeśli $A \subseteq \mathcal{D}$, to zbiór wszystkich podzbiorów A, czyli zbiór

$$\mathsf{P}(A) = \{X : \mathsf{P}(\mathcal{D}) \mid X \subseteq A\}$$

nazywamy *zbiorem potęgowym* zbioru A. Na przykład $\mathsf{P}(\{1, 2\}\}) = \{\varnothing, \{1\}, \{2\}, \{1, 2\}\}$ oraz $\mathsf{P}(\{\{1, 2\}, \{1\}\}) = \{\varnothing, \{\{1, 2\}\}, \{\{1\}\}, \{\{1, 2\}, \{1\}\}\}$. Oczywiście zachodzi równoważność:

$$A \in \mathsf{P}(B) \quad\leftrightarrow\quad A \subseteq B.$$

Używamy następujących skrótów:

$$A \not\subseteq B \text{ oznacza } \neg(A \subseteq B);$$

$$A \subsetneq B \text{ oznacza } A \subseteq B \wedge A \neq B.$$

Uwaga: Należy odróżniać *zawieranie* ($\subseteq$) od *należenia* ($\in$).

Równość

Napis „$x = y$" oznacza, że x i y są nazwami tego samego przedmiotu. Napis taki ma sens wtedy, gdy obiekty oznaczone przez x i y są tego samego typu. Sposób ustalenia czy x i y oznaczają to samo zależy oczywiście od tego jaki to jest typ. Często spotykane sformułowanie „*istnieje dokładnie jeden element* $x : \mathcal{D}$ *o własności* $W(x)$" możemy zapisać z pomocą równości jako $\exists x : \mathcal{D}(W(x) \wedge \forall y : \mathcal{D}(W(y) \to y = x))$. W skrócie piszemy $\exists! x : \mathcal{D}.\, W(x)$.

Równość zbiorów. Jak powiedzieliśmy, zbiór to pewien skrót myślowy. W istocie chodzi o elementy spełniające pewne kryterium wyboru. Inaczej mówiąc, zbiór jest jednoznacznie określony przez swoje elementy. Sposób w jaki określamy elementy zbioru (np. ich kolejność) nie ma znaczenia, ważne jest jedynie to, czy dany przedmiot należy do naszego zbioru, czy nie. Wyrażamy tę własność za pomocą następującej *zasady jednoznaczności*: Dla $A, B : \mathsf{P}(\mathcal{D})$,

$$A = B \leftrightarrow \forall z(z \in A \leftrightarrow z \in B).$$

Aby udowodnić, że dwa zbiory A i B są równe, postępujemy więc zwykle tak: pokazujemy, że każdy element zbioru A należy też do B, a każdy element zbioru B należy do A. A zatem równość zbiorów to ich wzajemne zawieranie.

Fakt 2.1 $\forall A, B{:}\mathsf{P}(\mathcal{D})(A = B \leftrightarrow A \subseteq B \wedge B \subseteq A)$.

Przykład: Zgodnie z zasadą jednoznaczności napisy $\{a, b\}$, $\{b, a\}$, $\{b, a, b\}$ i $\{a, b, b, a\}$ (a jeśli $a = b$, to także napisy $\{a\}$ i $\{b\}$) oznaczają ten sam zbiór.

Mówimy, że zbiór jest *pusty*, gdy nie ma żadnego elementu.

Fakt 2.2 *Każdy typ* $\mathcal{D}$ *ma dokładnie jeden pusty podzbiór.*

Dowód: Przypuśćmy, że $A_1, A_2 : \mathsf{P}(\mathcal{D})$ oraz $\forall x{:}\mathcal{D}(x \notin A_1)$ oraz $\forall x{:}\mathcal{D}(x \notin A_2)$. Wtedy

$$\forall x{:}\mathcal{D}(x \in A_1 \leftrightarrow x \in A_2)$$

co oznacza (z jednoznaczności), że $A_1 = A_2$. □

Zbiór pusty oznaczamy symbolem $\varnothing$.

Uwaga: Mówiąc, że zbiór A jest pusty, zaprzeczamy stwierdzeniu $\exists x.\, x \in A$; zauważmy, że znaczy to tyle samo, co stwierdzenie $\forall x.\, x \notin A$. Inaczej:

$$\neg\exists x.\, x \in A \qquad \text{wtedy i tylko wtedy, gdy} \qquad \forall x.\, x \notin A.$$

Powyższa równoważność jest przykładem zastosowania *prawa De Morgana:*

$$\neg\exists x.\, W(x) \qquad \text{wtedy i tylko wtedy, gdy} \qquad \forall x.\, \neg W(x).$$

Analogicznie, zaprzeczeniem tezy uniwersalnej jest teza egzystencjalna:

$$\neg\forall x\, W(x) \qquad \text{wtedy i tylko wtedy, gdy} \qquad \exists x.\, \neg W(x).$$

Działania na zbiorach

Niech $A, B : \mathsf{P}(\mathcal{D})$. Wówczas:

- *Sumą* zbiorów A i B nazywamy zbiór $A \cup B = \{x : \mathcal{D} \mid x \in A \vee x \in B\}$.
- *Iloczyn* lub *przecięcie* zbiorów A i B to zbiór $A \cap B = \{x : \mathcal{D} \mid x \in A \wedge x \in B\}$.
- *Różnicą* zbiorów A i B nazywamy zbiór $A - B = \{x : \mathcal{D} \mid x \in A \wedge x \notin B\}$.
- *Dopełnienie* zbioru A (do typu $\mathcal{D}$) to zbiór $-A = \{x : \mathcal{D} \mid x \notin A\}$ (czyli różnica $\mathcal{D} - A$).
- *Różnica symetryczna* zbiorów A i B to zbiór $A \dot{-} B = (A - B) \cup (B - A)$.

Dla odróżnienia od sumy prostej (patrz niżej), „zwykłą" sumę nazywamy czasem sumą *mnogościową* lub *teoriomnogościową*. Także o przecięciu zbiorów mówimy „iloczyn mnogościowy".

Uwaga: Definicja dopełnienia zbioru A zależy od typu $\mathcal{D}$. Jeśli typ nie jest ustalony, pojęcie dopełnienia nie ma sensu. Ale typ zazwyczaj wynika z kontekstu.

Ćwiczenie 2.3 Przypuśćmy, że zbiór A ma n elementów, a zbiór B ma m elementów. Ile elementów mogą mieć zbiory $A \cup B$, $A \cap B$, $A - B$?

Ćwiczenie 2.4 *Udowodnić, że dla dowolnych A i B jeśli $A - B = \varnothing$ to $A \subseteq B$.*

Rozwiązanie: Mamy udowodnić, że dla każdych A i B, z założenia $A - B = \varnothing$ wynika teza $A \subseteq B$. Przypuśćmy więc, że A i B są zbiorami spełniającymi warunek $A - B = \varnothing$. Naszym zadaniem jest wykazanie, że $A \subseteq B$, czyli że $\forall x\,(x \in A \to x \in B)$. Inaczej: każdy element x zbioru A ma należeć do B. Rozważmy więc jakiś element $x \in A$, pokażemy, że $x \in B$.

Posłużymy się wnioskowaniem przez zaprzeczenie: przypuśćmy, że $x \notin B$. Skoro $x \in A$ i $x \notin B$, to $x \in A - B$. Ale $A - B = \varnothing$, więc $x \in \varnothing$, co jest niemożliwe. Hipoteza $x \notin B$ okazała się fałszywa, czyli faktycznie $x \in B$.

A zatem udowodniliśmy implikację $x \in A \to x \in B$, a ponieważ x był zupełnie dowolny, więc możemy stwierdzić, że zachodzi warunek $\forall x\,(x \in A \to x \in B)$, czyli $A \subseteq B$. Ostatecznie widzimy, że zawieranie $A \subseteq B$ musi zachodzić zawsze wtedy, gdy $A - B = \varnothing$.

Powyższe rozwiązanie jest oczywiście nieco „przegadane". Zwykle taki dowód zapiszemy w zwięzły sposób:

Niech $A - B = \varnothing$ oraz $x \in A$. Gdyby $x \notin B$, to $x \in A - B = \varnothing$; sprzeczność. Zatem $x \in B$.

Przyjrzyjmy się bliżej konstrukcji naszego dowodu. Zauważmy na przykład różnicę pomiędzy założeniami $x \in A$ i $x \notin B$. Pierwsze z tych założeń obowiązuje wszędzie tam, gdzie mowa o hipotetycznym przedmiocie x, drugie założenie potrzebne nam było tylko „lokalnie" dla wykazania jego fałszywości.

Także sama nazwa x ma sens tylko w „wewnętrznej" części rozumowania (nie ma przecież w tezie twierdzenia mowy o żadnym x). Przypomina to zjawisko znane z programowania: lokalnie zadeklarowanego identyfikatora używamy tylko w bloku zawierającym jego deklarację.

Strukturę blokową naszego dowodu przedstawimy na rysunku z pomocą *pudełek Jaśkowskiego*. (Uczynione kursywą adnotacje o „celu" nie są częścią dowodu, ale komentarzem.)

Załóżmy, że $A - B = \varnothing$.	*(Cel 1: $A \subseteq B$)*
Weźmy dowolne $x \in A$.	*(Cel 2: $x \in B$)*
Załóżmy, że $x \notin B$.	*(Cel 3: sprzeczność)*
Skoro $x \in A$ i $x \notin B$, to $x \in A - B$.	
Ale $A - B = \varnothing$, więc $x \in \varnothing$; sprzeczność.	*(Cel 3 osiągnięty)*
Zatem $x \in B$.	*(Cel 2 osiągnięty)*
Zatem $\forall \in x\,(x \in A \to x \in B)$, czyli $A \subseteq B$.	*(Cel 1 osiągnięty)*

Zatem jeśli $A - B = \varnothing$ to $A \subseteq B$.

Na dobrą sprawę należałoby cały rysunek włożyć do jeszcze jednego dużego pudełka, zaczynającego się od „*Niech A i B będą dowolnymi zbiorami...*

Ćwiczenie 2.5 *Udowodnić, że dla dowolnych A, B, C, jeśli $A - B \subseteq C$ to $A \subseteq B \cup C$.*

Rozwiązanie: Załóżmy, że $A - B \subseteq C$. Aby wykazać $A \subseteq B \cup C$, przypuśćmy, że $x \in A$. Jeśli $x \in B$ to oczywiście $x \in B \cup C$, w przeciwnym razie $x \in A - B \subseteq C$, skąd $x \in C \subseteq B \cup C$.

W powyższym zwięzłym dowodzie występuje wnioskowanie przez przypadki. Bardziej szczegółową wersję tego dowodu przedstawimy z pomocą pudełek:

Załóżmy, że $A - B \subseteq C$.	*(Cel 1: $A \subseteq B \cup C$)*
Weźmy dowolne $x \in A$.	*(Cel 2: $x \in B \cup C$)*
Wiadomo, że $x \in B$ lub $x \notin B$.	
Przypuśćmy, że $x \in B$.	*(Cel 3: $x \in B \cup C$)*
Wtedy $x \in B \cup C$	*(Cel 3 osiągnięty)*
Przypuśćmy, że $x \notin B$.	*(Cel 4: $x \in B \cup C$)*
Ponieważ $x \in A$ i $x \notin B$, więc $x \in A - B$.	
Ponieważ $x \in A - B$ oraz $A - B \subseteq C$, więc $x \in C$.	
Wtedy $x \in B \cup C$	*(Cel 4 osiągnięty)*
Ponieważ $x \in B$ lub $x \notin B$, więc $x \in B \cup C$.	*(Cel 2 osiągnięty)*
Zatem $\forall \in x\,(x \in A \to x \in B \cup C)$	*(Cel 1 osiągnięty)*

Zatem jeśli $A - B \subseteq C$ to $A \subseteq B \cup C$.

Działania nieskończone

Pojęcie sumy i iloczynu można uogólnić. Przypuśćmy, że mamy rodzinę[3] zbiorów $\mathcal{R} : \mathsf{P}(\mathsf{P}(\mathcal{D}))$, inaczej mówiąc $\mathcal{R} \subseteq \mathsf{P}(\mathcal{D})$. Wtedy *sumą* (lub *sumą uogólnioną*) rodziny $\mathcal{R}$ nazywamy zbiór

$$\bigcup \mathcal{R} = \{x{:}\mathcal{D} \mid \exists A(x \in A \wedge A \in \mathcal{R})\}.$$

Suma rodziny $\mathcal{R} : \mathsf{P}(\mathsf{P}(\mathcal{D}))$ jest zawarta w $\mathcal{D}$, czyli jest typu $\mathsf{P}(\mathcal{D})$. Zapamiętajmy taką zasadę:

$$x \in \bigcup \mathcal{R} \quad \Leftrightarrow \quad \exists A(x \in A \wedge A \in \mathcal{R}).$$

[3] *Rodzina zbiorów* to zbiór, którego elementami są zbiory.

Jeśli $\mathcal{R} : \mathsf{P}(\mathsf{P}(\mathcal{D}))$ jest rodziną niepustą ($\mathcal{R} \neq \varnothing$) to określamy *uogólniony iloczyn* rodziny $\mathcal{R}$:

$$\bigcap \mathcal{R} = \{x{:}\mathcal{D} \mid \forall A(A \in \mathcal{R} \to x \in A)\}.$$

Dla $\mathcal{R} \neq \varnothing$ mamy równoważność:[4]

$$x \in \bigcap \mathcal{R} \quad \Leftrightarrow \quad \forall A(A \in \mathcal{R} \to x \in A).$$

Ćwiczenie 2.6 Udowodnić, że:

- jeśli $A \in \mathcal{R}$, to $\bigcap \mathcal{R} \subseteq A \subseteq \bigcup \mathcal{R}$;
- jeśli $\mathcal{R} \neq \varnothing$, to $\bigcap \mathcal{R} \subseteq \bigcup \mathcal{R}$;
- jeśli $\mathcal{R} = \varnothing$, to $\bigcup \mathcal{R} = \varnothing$;
- jeśli $\varnothing \in \mathcal{R}$, to $\bigcap \mathcal{R} = \varnothing$.

[4]Powyższa definicja ma sens także dla rodziny pustej; mielibyśmy wtedy $\bigcap \varnothing = \mathcal{D}$. Choć moralnie słuszna, definicja taka prowadzi jednak do pewnej niezręczności: iloczyn byłby większy od sumy.

3 Produkty, sumy proste, relacje

Jak już powiedzieliśmy wcześniej, każdy zbiór składa się z elementów tego samego rodzaju, czy też tego samego *typu*. Przez *typ* rozumiemy tu więc pewną naturalnie określoną dziedzinę matematyczną. Oczywiście wszystkie obiekty danego typu także tworzą zbiór, a więc każdy typ jest zbiorem. Na odwrót niekoniecznie – aby mówić o typie musimy mieć po temu dostatecznie dobre powody, zwykle zależne od kontekstu matematycznego.

Są pewne naturalne sposoby tworzenia nowych typów z typów już znanych. Oprócz typu potęgowego to na przykład iloczyn kartezjański, suma prosta i typ funkcyjny.

Iloczyn kartezjański zbiorów A i B, to zbiór oznaczany przez $A \times B$, który składa się z *par uporządkowanych* postaci $\langle a, b\rangle$, gdzie $a \in A$ oraz $b \in B$. Para uporządkowana $\langle a, b\rangle$ to abstrakcyjny obiekt zadany przez wybór *pierwszej współrzędnej* a i *drugiej współrzędnej* b. Inaczej mówiąc, dwie pary uważamy za równe, gdy ich odpowiednie współrzędne są takie same.

$$\langle a, b\rangle = \langle x, y\rangle \quad \text{wtedy i tylko wtedy, gdy} \quad a = x \ \text{ oraz } \ b = y.$$

Powyższa równoważność wyraża zasadniczą własność par uporządkowanych. Można uważać ją za pośrednią (aksjomatyczną) definicję pojęcia pary uporządkowanej. Rzeczywiście, dalsze nasze rozważania dotyczące par uporządkowanych będą się wyłącznie na tej własności opierać. (Tak naprawdę nie jest ważne *czym w istocie są* pary uporządkowane, ważne że zachowują się zgodnie ze swoją „specyfikacją".)

Jeśli $a : \mathcal{D}$ i $b : \mathcal{E}$, to para uporządkowana $\langle a, b\rangle$ jest typu $\mathcal{D} \times \mathcal{E}$. A zatem, dla $A : \mathsf{P}(\mathcal{D})$ i $B : \mathsf{P}(\mathcal{E})$ mamy $A \times B : \mathsf{P}(\mathcal{D} \times \mathcal{E})$.

Pojęcie produktu można uogólnić na trzy i więcej wymiarów. Można też *zdefiniować* produkt $A \times B \times C$ jako $(A \times B) \times C$, przyjmując, że trójka uporządkowana $\langle a, b, c\rangle$ to para $\langle\langle a, b\rangle, c\rangle$. Czwórka uporządkowana $\langle a, b, c, d\rangle$ to para $\langle\langle a, b, c\rangle, d\rangle$ i tak dalej. Równie dobrze moglibyśmy nawiasy rozstawić inaczej, ale nie ma to znaczenia. Istotne własności produktu $(A \times B) \times C$ są takie same jak własności produktu $A \times (B \times C)$ i na dobrą sprawę można je utożsamiać.

Na oznaczenie produktu postaci $A \times A$ piszemy często A^2. Podobnie A^3 oznacza $A \times A \times A$ i ogólnie A^k to produkt postaci $A \times \cdots \times A$, gdzie A występuje k razy.

Uporządkowane pary, trójki, czwórki itd. nazywamy po polsku *krotkami*.

Suma prosta zbiorów A i B, którą oznaczymy przez $A \oplus B$, zwana jest też *koproduktem* lub *sumą rozłączną*. Elementami $A \oplus B$ są „kopie" elementów A i „kopie" elementów B. Ściślej, każdy element sumy prostej $A \oplus B$ jest

- albo postaci $\langle a\rangle_1$, gdzie $a \in A$ (lewa kopia elementu a);
- albo postaci $\langle b\rangle_2$, gdzie $b \in B$ (prawa kopia elementu b).

Przyjmujemy przy tym, że lewe i prawe kopie są zawsze różne, czyli:

$$\langle x\rangle_i = \langle y\rangle_j \quad \text{wtedy i tylko wtedy, gdy} \quad x = y \text{ oraz } i = j.$$

Jeśli $A : \mathsf{P}(\mathcal{D})$ i $B : \mathsf{P}(\mathcal{E})$, to elementy sumy prostej są typu $\mathcal{D} \oplus \mathcal{E}$, a sam zbiór $A \oplus B$ jest typu $\mathsf{P}(\mathcal{D} \oplus \mathcal{E})$.

Suma prosta umożliwia częściowe (ale kontrolowane) „obejście" zasady, że elementy jednego zbioru muszą być tego samego typu. Często zaniedbujemy różnicę pomiędzy elementem a zbioru A i elementem $\langle a\rangle_1$ zbioru $A \oplus B$ (i tak samo z prawej), traktując składowe sumy prostej jak zwykłe jej (rozłączne) podzbiory.[5] Stosowanie tej konwencji wymaga jednak pewnej ostrożności: na przykład $A \oplus A$ jest sumą mnogościową dwóch *rozłącznych kopii* tego samego zbioru A. W szczególności $A \oplus A$ to co innego niż A.

Suma prosta trzech (i więcej) składników może być definiowana podobnie jak w przypadku produktu: $A \oplus B \oplus C = (A \oplus B) \oplus C$.

Ćwiczenie 3.1 Przypuśćmy, że zbiór A ma n elementów, a zbiór B ma m elementów. Ile elementów mają zbiory $A \oplus B$, $A \times B$, $\mathsf{P}(A)$?

Relacje

Relacja to to samo, co wieloargumentowy predykat. Skoro zaś zbiór to nic innego jak „zmaterializowany" predykat, więc wygodnym uściśleniem pojęcia relacji jest taka definicja: relacja dwuargumentowa to po prostu zbiór wszystkich uporządkowanych par tych przedmiotów, pomiędzy ktorymi relacja zachodzi.[6] Istotnie, znając ten zbiór, wiemy wszystko o relacji.

Definicja 3.2 Dowolny podzbiór r iloczynu kartezjańskiego $A\times B$ nazywamy *relacją z A do B*. Jeśli $A = B$, to mówimy, że r jest relacją *w zbiorze A*. Piszemy często „$x\,r\,y$" albo „$r(x, y)$" zamiast „$\langle x, y\rangle \in r$".

Definicja 3.3 Pewne własności relacji dwuargumentowych mają swoje nazwy. Oto niektóre z nich. Mówimy, że relacja r w A jest

$$\begin{array}{rll} \textit{zwrotna w } A, & \text{gdy} & \forall x \in A\,(x\,r\,x);\\ \textit{symetryczna}, & \text{gdy} & \forall x, y\,(x\,r\,y \to y\,r\,x);\\ \textit{przechodnia}, & \text{gdy} & \forall x, y, z\,(x\,r\,y \wedge y\,r\,z \to x\,r\,z);\\ \textit{antysymetryczna}, & \text{gdy} & \forall x, y\,(x\,r\,y \wedge y\,r\,x \to x = y);\\ \textit{spójna w } A, & \text{gdy} & \forall x, y \in A\,(x\,r\,y \vee y\,r\,x). \end{array}$$

Na przykład relacja prostopadłości prostych na płaszczyźnie jest symetryczna, ale nie jest zwrotna,[7] antysymetryczna, przechodnia ani spójna. Natomiast relacja równoległości prostych jest zwrotna, przechodnia i symetryczna, ale nie jest antysymetryczna ani spójna.

Relacja równoważności to relacja zwrotna, symetryczna i przechodnia, jak na przykład wspomniana właśnie równoległość prostych. Natomiast relację zwrotną, antysymetryczną i przechodnią nazywamy *częściowym porządkiem* (lub relacją *częściowo porządkującą*). Jeśli częściowy porządek jest na dodatek spójny, to przysługuje mu tytuł *liniowego* porządku. Na przykład zawieranie zbiorów wyznacza częściowy porządek w $\mathsf{P}(\mathcal{D})$, a zwykła relacja $\leq$ jest liniowym porządkiem w $\mathbb{N}$ (podobnie w $\mathbb{R}$).

[5] Uważamy więc $\mathcal{D}$ i $\mathcal{E}$ za *podtypy* koproduktu $\mathcal{D} \oplus \mathcal{E}$.

[6] Tu ograniczamy się do relacji dwuargumentowych. Relacje wieloargumentowe utożsamia się ze zbiorami odpowiednich krotek.

[7] Zamiast „zwrotna w A", czy „spójna w A" zwykle mówimy „zwrotna" czy „spójna". Pamiętajmy jednak, że jeśli $A \subsetneq B$ to relacja zwrotna (spójna) w A nie jest już zwrotna (spójna) w B.

Definicja 3.4 *Relacją odwrotną* do danej relacji $r \subseteq A \times B$ nazywamy zbiór

$$r^{-1} = \{\langle y, x\rangle \in B \times A \mid \langle x, y\rangle \in r\}.$$

Oczywiście r^{-1} jest relacją z B do A. Na przykład relacją odwrotną do relacji $\leq$ w zbiorze liczb rzeczywistych jest relacja $\geq$.

Jeśli $r \subseteq A \times B$ oraz $s \subseteq B \times C$, to *złożeniem relacji r z relacją s* nazywamy relację $r \cdot s \subseteq A \times C$, oznaczaną też przez $(r\,;s)$, a określoną tak:

$$x\,(r \cdot s)\,y \quad \text{wtedy i tylko wtedy, gdy} \quad \exists z{\in}B\,(x\,r\,z \wedge z\,s\,y).$$

Na przykład złożenie relacji $\leq$ z relacją $<$ to relacja $<$.

Relacja *identycznościowa* w zbiorze A to relacja $\mathbf{1}_A = \{\langle a, a\rangle \mid a \in A\}$. Zauważmy, że dowolna relacja r jest zwrotna w A wtedy i tylko wtedy, gdy $\mathbf{1}_A \subseteq r$.

Zauważmy, że użycie jedynki na oznaczenie relacji identycznościowej nie jest przypadkowe, bo jeśli r jest relacją w A to $r \cdot \mathbf{1}_A = \mathbf{1}_A \cdot r = r$. Następujący łatwy fakt stanowi ładną charakteryzację przechodniości:

Fakt 3.5 *Relacja r jest przechodnia wtedy i tylko wtedy, gdy $r \cdot r \subseteq r$.*

Lemat 3.6 *Iloczyn dowolnej niepustej rodziny relacji przechodnich jest relacją przechodnią.*

Dowód: Niech $\mathcal{R}$ będzie niepustą rodziną relacji przechodnich w jakimś typie $\mathcal{D}$. Oznacza to, że każdy element $r \in \mathcal{R}$ jest relacją przechodnią w $\mathcal{D}$, w szczególności $r \subseteq \mathcal{D} \times \mathcal{D}$. Wtedy także $\bigcap \mathcal{R} \subseteq \mathcal{D} \times \mathcal{D}$, tj. $\bigcap \mathcal{R}$ jest relacją w $\mathcal{D}$. Mamy udowodnić, że jest to relacja przechodnia. Niech więc $\langle a, b\rangle, \langle b, c\rangle \in \bigcap \mathcal{R}$. Z definicji iloczynu wynika, że $\langle a, b\rangle, \langle b, c\rangle \in r$ dla wszystkich elementów $r \in \mathcal{R}$. Ale elementy rodziny $\mathcal{R}$ są relacjami przechodnimi, więc para $\langle a, c\rangle$ należy do każdej z nich. Stąd $\langle a, c\rangle \in \bigcap \mathcal{R}$, a tegośmy właśnie chcieli. □

Fakt 3.7 *Dla dowolnej relacji r istnieje taka relacja przechodnia r^+, że*

- $r \subseteq r^+$;
- *jeśli $r \subseteq s$ i s przechodnia to $r^+ \subseteq s$.*

Dowód: Skorzystamy z lematu 3.6. Zakładając, że r jest relacją w zbiorze A, można zdefiniować $r^+ = \bigcap\{s \subseteq A \times A \mid s \text{ jest przechodnia oraz } r \subseteq s\}$. □

Ćwiczenie 3.8 Uzupełnić dowód faktu 3.7. Dlaczego definicja relacji r^+ jest w ogóle poprawna? Dlaczego ta relacja spełnia oczekiwane warunki?

Definicja 3.9 Relacja r^+ z faktu 3.7 to najmniejsza relacja przechodnia zawierająca r. Nazywamy ją *domknięciem przechodnim* relacji r. Podobnie, *domknięcie przechodnio-zwrotne* r to najmniejsza relacja przechodnia i zwrotna zawierająca r, czyli relacja $r^* = \mathbf{1}_A \cup r^+$.

Jeśli używamy symbolu strzałki do oznaczenia relacji, to często zamiast $\to^*, \Rightarrow^*$ itp. piszemy odpowiednią strzałkę z podwójnym grotem, czyli $\twoheadrightarrow, \Rightarrow\!\!\!\!\Rightarrow$, itp.

Fakt 3.10 Dla dowolnej relacji r w zbiorze A zachodzą równości

$$r^+ = r \cdot r^* = r^* \cdot r.$$

Dowód: Udowodnimy pierwszą równość. Inkluzja $\subseteq$ wynika stąd, że relacja $r \cdot r^*$ jest przechodnia i zawiera r (to łatwe). Dla dowodu inkluzji odwrotnej, zauważmy najpierw[8], że składanie relacji jest monotoniczne (jeśli $r \subseteq r'$ i $s \subseteq s'$, to $r \cdot s \subseteq r' \cdot s'$) i rozdzielne względem sumy: $r \cdot (s \cup s') = r \cdot s \cup r \cdot s'$. Zatem $r \cdot r^* = r \cdot (\mathbf{1}_A \cup r^+) = r \cdot \mathbf{1}_A \cup r \cdot r^+ \subseteq r \cup r^+ \cdot r^+ \subseteq r^+$, na mocy faktu 3.5. □

[8]Sprawdzenie tych własności proponujemy Czytelnikowi jako ćwiczenie.

4 Funkcje

O *funkcji* z A do B mówimy wtedy, gdy każdemu elementowi zbioru A potrafimy jednoznacznie przypisać pewien element zbioru B. Inaczej, definicja funkcji f z A do B polega na określeniu *wartości funkcji* $f(x) \in B$ dla każdego *argumentu* $x \in A$. Wtedy można napisać

$$f : A \to B.$$

Zbiór wszystkich funkcji z A do B oznaczamy przez $A \to B$ albo przez B^A. Funkcję można zdefiniować na kilka sposobów. Najprościej jest napisać równanie postaci $f(x) = E(x)$, gdzie $E(x)$ jest wyrażeniem (nie zawierającym symbolu f), którego wartość należy do B, gdy $x \in A$. Na przykład równanie $f(x) = 3x^2 + 2x - 1$ określa pewną funkcję z $\mathbb{R}$ do $\mathbb{R}$. Tę samą definicję można wypowiedzieć przy pomocy *notacji lambda*, pisząc $f = \lambda x.3x^2 + 2x - 1$, albo $f = \lambda x{:}\mathbb{R}.3x^2+2x-1$. Ogólnie, napis postaci $\lambda x.E(x)$ czytamy: „funkcja zmiennej x określona wyrażeniem $E(x)$." A więc:

$$(\lambda x.E(x))(a) = E(a).$$

Notacja lambda jest przydatna np. wtedy, gdy nie chcemy wprowadzać dodatkowych symboli na oznaczenie funkcji, np. w zdaniu „Funkcja $\lambda x.2xy + y$ jest pochodną funkcji $\lambda x.x^2y + xy$."

Częstym sposobem określania funkcji jest definicja warunkowa, ta na przykład

$$f(n) = \begin{cases} n/2, & \text{jeśli } n \text{ jest parzyste;} \\ 3n+1, & \text{w przeciwnym przypadku,} \end{cases}$$

określa funkcję z $\mathbb{N}$ do $\mathbb{N}$. Można ją też zapisać inaczej:

$$f(n) = \texttt{if } n \text{ jest parzyste } \texttt{then } n/2 \texttt{ else } 3n+1.$$

Taka definicja też bezpośrednio określa wartość funkcji dla każdego argumentu.

Jednak nie zawsze definicja „wprost" jest możliwa. Wtedy czasem używamy definicji „implicite". Na przykład dzielenie całkowite liczb naturalnych przez trzy można określić, mówiąc, że wartością funkcji dla danego $n : \mathbb{N}$ jest liczba $m : \mathbb{N}$, spełniająca warunek

$$(3 \cdot m \leq n) \wedge (n < 3 \cdot (m+1)).$$

Ważne, że taka liczba m zawsze istnieje i że jest tylko jedna. Tego typu definicje można zapisywać z pomocą (niesłusznie trochę dziś zapomnianej) *notacji jota*: wyrażenie $\iota y.W(y)$ czytamy „jedyne y o własności $W(y)$". Wartość tego wyrażenia[9] jest określona wtedy i tylko wtedy, gdy istnieje dokładnie jeden taki element y, że $W(y)$. Notacja jota może też wskazywać na typ y lub zbiór do którego y ma należeć. Zatem dla $a \in A$ mamy:

$$a = \iota y \in A.W(y) \qquad \text{wtedy i tylko wtedy, gdy} \qquad W(a) \wedge \forall y \in A(W(y) \to y = a).$$

A więc nasze dzielenie całkowite to funkcja $\lambda n{:}\mathbb{N}\, \iota m{:}\mathbb{N}.\, (3 \cdot m \leq n) \wedge (n < 3 \cdot (m+1))$.

Funkcje częściowe: Ponieważ nie zawsze istnieje dokładnie jedna wartość y spełniająca żądany warunek, więc wyrażenie $\lambda x \in A\, \iota y \in B.\, W(x, y)$ może nie być dobrą definicją funkcji z A do B. Nawet definicja funkcji „wprost" może być niebezpieczna, np. taka: $\lambda x{:}\mathbb{R}.1/x$.

[9] Zauważmy, że zarówno lambda jak jota powoduje wiązanie zmiennej.

Dlatego mówi się też o *funkcjach częściowych*, tj. takich, których wartość nie zawsze musi być określona. („Zwykłe" funkcje, nazywane też funkcjami *całkowitymi*, stanowią szczególny przypadek funkcji częściowych.) Napis $f : A \rightharpoonup B$ stwierdza, że f jest funkcją częściową z A do B. Zbiór $\mathrm{Dom}(f) = \{x \in A \mid f(x) \text{ jest określone}\}$ nazywamy wtedy *dziedziną* funkcji f. Zachodzi równoważność:

$$a \in \mathrm{Dom}(\lambda x \in A.\, \imath y \in B.\, W(x,y)) \qquad \Leftrightarrow \qquad \exists! y \in B.\, W(a,y). \tag{4.1}$$

Dziedziną funkcji całkowitej $f : A \to B$ jest oczywiście A.

Równość funkcji: Powiedzieliśmy wyżej, że aby jednoznacznie określić funkcję $f : A \to B$ potrzeba i wystarcza określić wartość $f(x) \in B$ dla dowolnego $x \in A$. Funkcje, które tym samym argumentom przypisują te same wartości uznajemy więc za identyczne.

$$f = g \quad \Leftrightarrow \quad (\mathrm{Dom}(f) = \mathrm{Dom}(g)) \wedge \forall x (x \in \mathrm{Dom}(f) \to f(x) = g(x)). \tag{4.2}$$

Dla funkcji o ustalonej dziedzinie A możemy to napisać prościej:

$$\begin{array}{ll} f = g & \text{wtedy i tylko wtedy, gdy } \forall x{:}A.\, f(x) = g(x); \\ f \neq g & \text{wtedy i tylko wtedy, gdy } \exists x{:}A.\, f(x) \neq g(x). \end{array}$$

Wykresem funkcji $f : A \to B$ nazywamy zbiór (relację) $\mathcal{W}(f) = \{\langle x, y\rangle \mid f(x) = y\} \subseteq A \times B$. A zatem zasada równości dla funkcji może zostać wypowiedziana tak: funkcje o tych samych wykresach są równe. W aksjomatycznej teorii mnogości funkcję wręcz utożsamia się z jej wykresem (uważa się, że funkcja to po prostu zbiór par).

Inne definicje: *Zbiorem wartości* funkcji $f : A \to B$ nazywamy zbiór

$$\mathrm{Rg}(f) = \{y{:}B \mid \exists x{:}A\, f(x) = y\}.$$

Napis $f : A \to B$ oznacza, że $\mathrm{Dom}(f) = A$ oraz $\mathrm{Rg}(f) \subseteq B$. Jeśli $f : A \to B$ i $B \subseteq B'$ to poprawny jest też napis $f : A \to B'$. Ale dla $A' \neq A$ nieprawdą jest, że $f : A' \to B$.

Czasami chcemy ograniczyć dziedzinę jakiejś funkcji do interesującego nas podzbioru dziedziny. Jeśli $f : A \to B$, to *obcięciem* funkcji f do podzbioru C zbioru A nazywamy funkcję $f\restriction_C : C \to B$, określoną tak samo jak f, tj. $f\restriction_C(a) = f(a)$ dla $a \in C$. Ściśle rzecz biorąc, funkcje f i $f\restriction_C$ są różne, ale często dla uproszczenia zamiast $f\restriction_C$ piszemy po prostu f.

Funkcja określona na iloczynie kartezjańskim $A \times B$ nazywana jest *funkcją dwuargumentową*. Zwykle zamiast $f(\langle x, y\rangle)$ piszemy po prostu $f(x, y)$. Podobnie postępujemy przy większej liczbie argumentów. (Funkcja zeroargumentowa to stała.)

Zbiór wszystkich funkcji o dziedzinie A i wartościach w B oznaczamy przez $A \to B$ lub B^A.

Jeśli $A : \mathsf{P}(\mathcal{D})$ i $B : \mathsf{P}(\mathcal{E})$ to funkcja $f : A \to B$ jest funkcją częściową z $\mathcal{D}$ do $\mathcal{E}$. Mówimy wtedy, że taka funkcja jest *typu* $\mathcal{D} \rightharpoonup \mathcal{E}$, a typ $\mathcal{E}$ (ale nie zbiór B) możemy nazwać *przeciwdziedziną* funkcji f. Powiemy oczywiście, że funkcja określona na całym typie $\mathcal{D}$ jest typu $\mathcal{D} \to \mathcal{E}$. Każda funkcja całkowita typu $\mathcal{D} \to \mathcal{E}$ może być uważana za funkcję częściową o dziedzinie $\mathcal{D}$. Inaczej mówiąc typ $\mathcal{D} \to \mathcal{E}$ jest podtypem typu $\mathcal{D} \rightharpoonup \mathcal{E}$.

Uwaga: Przypomnijmy, że typ $\mathcal{D}$ jest podtypem typu $\mathcal{E}$, gdy obiekty typu $\mathcal{D}$ można traktować tak jak gdyby należały do $\mathcal{E}$. Dotychczas zauważyliśmy trzy przykłady tego zjawiska:

- Liczby naturalne i rzeczywiste;
- Funkcje całkowite i częściowe;
- Składowe sumy prostej i cała suma.

Mamy wtedy *koercję* czyli (*włożenie*) z $\mathcal{D}$ do $\mathcal{E}$, czasami jawnie określoną, ale często traktowaną jako domyślną. Przykładem koercji jest włożenie $\lambda x{:}\mathcal{D}_1.\langle x\rangle_1 : \mathcal{D}_1 \xrightarrow{1-1} \mathcal{D}_1 \oplus \mathcal{D}_2$. W zależności od potrzeb możemy element d typu $\mathcal{D}_1$ utożsamiać z jego obrazem $\langle d\rangle_1$, który jest typu $\mathcal{D}_1 \oplus \mathcal{D}_2$ albo odróżniać te dwa obiekty. Podobna koercja występuje np. wtedy gdy liczbę naturalną 1 utożsamiamy z liczbą rzeczywistą 1.0.

Injekcje, surjekcje, bijekcje

- Funkcja $f : A \to B$ jest *różnowartościowa* (co zapisujemy $f : A \xrightarrow{1-1} B$) wtedy i tylko wtedy, gdy zachodzi warunek $\forall x, y \in A\,(x \neq y \to f(x) \neq f(y))$, lub równoważnie, gdy $\forall x, y \in A\,(f(x) = f(y) \to x = y)$.
- Funkcja $f : A \to B$ jest *na* B wtedy i tylko wtedy, gdy $\forall y \in B\,\exists x \in A\,(f(x) = y)$, lub równoważnie, gdy $B = \mathrm{Rg}(f)$. Używamy wtedy zapisu $f : A \xrightarrow{\text{na}} B$.
- Funkcję różnowartościową nazywamy też *injekcją*, funkcję „na" nazywamy *surjekcją*, a funkcję, która jest różnowartościowa i „na" nazywamy *bijekcją*. W przypadku bijekcji stosujemy notację $f : A \xrightarrow[\text{na}]{1-1} B$.

Analogiczne pojęcia można sformułować dla funkcji częściowych. Powiemy na przykład, że funkcja $f : A \rightharpoonup B$ jest różnowartościowa, gdy $\forall x, y \in \mathrm{Dom}(f)\,(x \neq y \to f(x) \neq f(y))$.

Ważne przykłady

Funkcjami różnowartościowymi są *włożenia* $\mathsf{in}_1 : A \xrightarrow{1-1} A \oplus B$ i $\mathsf{in}_2 : B \xrightarrow{1-1} A \oplus B$, określone wzorem $\mathsf{in}_i(z) = \langle z\rangle_i$, dla $i = 1, 2$. Jeśli $A, B \neq \varnothing$, to przykładami surjekcji są *rzutowania* $\pi_1 : A \times B \to A$ oraz $\pi_2 : A \times B \to B$ określone równaniami $\pi_1(\langle x, y\rangle) = x$ i $\pi_2(\langle x, y\rangle) = y$. Trywialnym przykładem bijekcji typu $A \to A$ jest funkcja identycznościowa $\mathrm{id}_A = \lambda x \in A.x$.

Odwracanie i składanie funkcji

Jeżeli $f : A \xrightarrow{1-1} B$ to możemy określić funkcję częściową $f^{-1} : B \rightharpoonup A$, przyjmując dla $y : B$

$$f^{-1}(y) = \boldsymbol{\imath}\, x \in A.\, f(x) = y.$$

Na przykład $(\lambda x{:}\mathbb{R}.\, 2x)^{-1} = \lambda x{:}\mathbb{R}.\, \frac{x}{2}$. Funkcję f^{-1} nazywamy *funkcją odwrotną* do funkcji f. Mamy do zapamiętania równoważność:

$$f^{-1}(y) = x \qquad \text{wtedy i tylko wtedy, gdy} \qquad f(x) = y.$$

Fakt 4.1 *Jeśli* $f : A \xrightarrow{1-1} B$, *to* $f^{-1} : \mathrm{Rg}(f) \xrightarrow[\text{na}]{1-1} A$. *Jeśli* f *jest bijekcją z* A *do* B *to* f^{-1} *jest bijekcją z* B *do* A.

Dowód: Zacznijmy od tego, ze dziedziną f^{-1} jest $\mathrm{Rg}(f)$. Istotnie, na mocy równoważności (4.1) mamy $y \in \mathrm{Dom}(f^{-1})$ wtedy i tylko wtedy, gdy $\exists! x\, f(x) = y$. Stąd natychmiast otrzymujemy inkluzję $\mathrm{Dom}(f^{-1}) \subseteq \mathrm{Rg}(f)$. Aby wykazać inkluzję w przeciwną stronę, załóżmy, że $y \in \mathrm{Rg}(f)$. Wtedy $y = f(x)$ dla pewnego x. Takie x jest tylko jedno, bo gdyby $f(x') = y$ to $x = x'$ z różnowartościowości funkcji f.

Oczywiście jeśli $f^{-1}(y) = x$ to $f(x) = y$, w szczególności $x \in A$. Zatem $f^{-1} : \mathrm{Rg}(f) \to A$.

Funkcja f^{-1} jest różnowartościowa, bo gdyby $f^{-1}(x) = f^{-1}(y) = z$ to $x = f(z) = y$. Jest ona także na A, bo dla dowolnego $x \in A$ mamy $f(x) = f(x)$, a stąd $x = f^{-1}(f(x))$. □

Definicja 4.2 Niech $f : A \to B$ oraz $g : B \to C$. *Złożeniem funkcji f i g* nazywamy funkcję $g \circ f : A \to C$ określoną równaniem $(g \circ f)(x) = g(f(x))$ dla $x \in A$. Na przykład złożenie $(\lambda x.\, x + 1) \circ (\lambda x.\, 2x)$ to funkcja $(\lambda x.\, 2x + 1)$.

Dowody poniższych faktów pozostawione są jako ćwiczenie:

Fakt 4.3

1) *Jeśli $f : A \to B$, $g : B \to C$ i $h : C \to D$, to $h \circ (g \circ f) = (h \circ g) \circ f$.*

2) *Jeśli $f : A \xrightarrow[\text{na}]{1-1} B$, to $f^{-1} \circ f = \mathrm{id}_A$ oraz $f \circ f^{-1} = \mathrm{id}_B$.*

3) *Jeśli $f : A \to B$, to $f \circ \mathrm{id}_A = f = \mathrm{id}_B \circ f$.*

Fakt 4.4

1) *Jeśli $f : A \xrightarrow{1-1} B$ oraz $g : B \xrightarrow{1-1} C$ to $g \circ f : A \xrightarrow{1-1} C$.*

2) *Jeśli $f : A \xrightarrow{\text{na}} B$ oraz $g : B \xrightarrow{\text{na}} C$ to $g \circ f : A \xrightarrow{\text{na}} C$.*

Sumowanie funkcji: Jeśli $f, g : A \rightharpoonup B$ mają tę własność, że $f(x) = g(x)$ dla dowolnego $x \in \mathrm{Dom}(f) \cap \mathrm{Dom}(g)$, to mówimy, że funkcje f i g są *zgodne*. Najprostszy przypadek zgodności funkcji zachodzi wtedy, gdy ich dziedziny są rozłączne. Inny szczególny przypadek ma miejsce, gdy $\mathrm{Dom}(f) \subseteq \mathrm{Dom}(g)$, oraz $f(x) = g(x)$ dla wszystkich $x \in \mathrm{Dom}(f)$. Piszemy wtedy $f \subseteq g$. (W istocie wtedy wykres funkcji f jest zawarty w wykresie funkcji g.) Jeśli f i g są zgodne, to można określić funkcję $h : A \rightharpoonup B$ wzorem

$$h(x) = \begin{cases} f(x), & \text{jeśli } x \in \mathrm{Dom}(f); \\ g(x), & \text{jeśli } x \in \mathrm{Dom}(g). \end{cases}$$

Funkcję tę nazywamy oczywiście *sumą* funkcji f i g i zwykle oznaczamy przez $f \cup g$.

Podobna sytuacja ma miejsce przy definiowaniu funkcji na sumie prostej. Jeśli $f : A_1 \to B$ i $g : A_2 \to B$, to *sumą prostą* funkcji f i g nazwiemy funkcję $f \oplus g : A_1 \oplus A_2 \to B$ określoną wzorem

$$(f \oplus g)(\langle x \rangle_i) = \begin{cases} f(x), & \text{jeśli } i = 1; \\ g(x), & \text{jeśli } i = 2. \end{cases}$$

Sumowanie funkcji ma sens także dla dowolnej rodziny funkcji częściowych $\mathcal{F}$, o ile każde dwie funkcje z tej rodziny są zgodne. *Sumą* rodziny $\mathcal{F}$ nazwiemy wtedy funkcję $\bigcup\mathcal{F}$ określoną na $\bigcup\{\mathrm{Dom}(f) \mid f \in \mathcal{F}\}$ warunkiem

$$(\bigcup\mathcal{F})(x) = f(x), \text{ gdzie } f \in \mathcal{F} \text{ oraz } x \in \mathrm{Dom}(f).$$

Ściślej, $(\bigcup\mathcal{F})(x) = \imath y \exists f(f \in \mathcal{F} \wedge x{\in}\mathrm{Dom}(f) \wedge y = f(x))$.

Obrazy i przeciwobrazy

Definicja 4.5 Niech $f : A \multimap B$. *Obraz* zbioru $C \subseteq A$ przy przekształceniu f to zbiór

$$f(C) = \{b \in B \mid \exists a \in \mathrm{Dom}(f)\,(a \in C \wedge f(a) = b)\}.$$

Inaczej można napisać:

$$f(C) = \{f(a) \mid a \in C\}.$$

Przeciwobrazem zbioru $D \subseteq B$ przy przekształceniu f nazywamy zbiór

$$f^{-1}(D) = \{a \in A \mid a \in \mathrm{Dom}(f) \wedge f(a) \in D\}.$$

Na przykład niech $f : \mathbb{N} \to \mathsf{P}(\mathbb{N})$ będzie funkcją przyporządkowującą każdej liczbie $n \in \mathbb{N}$ zbiór jej właściwych (różnych od 1 i od n) dzielników pierwszych, przy czym przyjmijmy, że zero nie ma dzielników pierwszych. Wtedy $f(\{1,3,4,6,9,12\}) = \{\varnothing,\{2\},\{3\},\{2,3\}\}$, oraz $f^{-1}(\{\{2\},\{1,2,27,36\}\}) = \{2^k \mid k \in \mathbb{N} - \{0,1\}\}$.

Uwaga: (1) Oznaczenie $f^{-1}(C)$ jest w istocie dwuznaczne. Może tu chodzić o przeciwobraz C przy przekształceniu f lub o obraz C przy przekształceniu f^{-1} (jeśli jest określone). Szczęśliwie, w obu wypadkach chodzi o ten sam zbiór (ćwiczenie).

(2) Także napis $f(C)$ może budzić wątpliwości: chodzi o wartość funkcji f w punkcie C czy o obraz zbioru C? Dlatego obraz oznacza się czasem przez $\vec{f}(C)$, lub $f[C]$. Ponieważ jednak elementy dziedziny funkcji i podzbiory tej dziedziny są *innego typu*, znaczenie napisu $f(C)$ jest zwykle oczywiste.

Rodziny indeksowane

O *zbiorze indeksowanym* $\{A_t\}_{t\in T}$ mówimy wtedy, gdy rozważamy pewne obiekty A_t, identyfikowane (indeksowane) przez elementy zbioru T, a przy tym możliwe są powtórzenia (możliwe, że $A_t = A_s$ dla $t \neq s$). Chcemy odróżnić taki indeksowany zbiór od zwykłego zbioru $\{A_t \mid t \in T\}$, bo tak naprawdę mamy tu do czynienia z funkcją, która indeksom $t \in T$ przypisuje wartości A_t.

Definicja 4.6 *Zbiór indeksowany* $\{A_t\}_{t\in T}$ elementów $\mathcal{D}$, to taka funkcja $A : T \to \mathcal{D}$, że $A(t) = A_t$, dla dowolnego $t \in T$.

Jeśli A_t są zbiorami, to mówimy o *indeksowanej rodzinie zbiorów*. *Sumę* takiej rodziny indeksowanej $\{A_t\}_{t\in T}$ definiujemy jako sumę uogólnioną jej zbioru wartości, czyli zwykłej rodziny zbiorów $\{A_t \mid t \in T\}$. Podobnie definiujemy iloczyn rodziny indeksowanej. A więc:

$$\bigcup_{t\in T} A_t = \bigcup\{A_t \mid t \in T\} \qquad \text{oraz} \qquad \bigcap_{t\in T} A_t = \bigcap\{A_t \mid t \in T\}.$$

Produkt uogólniony: Iloczyn kartezjański (produkt) $A \times B$ składa się z par. Elementami produktu skończonej liczby zbiorów są zaś krotki odpowiedniej długości. To podsuwa pomysł jak można zdefiniować produkt rodziny zbiorów indeksowanej liczbami naturalnymi: produktem rodziny $\{A_n\}_{n\in\mathbb{N}}$ powinien być zbiór wszystkich ciągów nieskończonych $a_0, a_1, \ldots$ spełniających warunek $a_n \in A_n$ dla dowolnego $n \in \mathbb{N}$. No dobrze, ale co to jest „ciąg nieskończony"? Funkcja o dziedzinie $\mathbb{N}$. Po tej obserwacji poniższa definicja powinna być oczywista.

Definicja 4.7 *Produktem uogólnionym* (lub po prostu „produktem" albo „iloczynem kartezjańskim") rodziny indeksowanej $\{A_t\}_{t\in T}$ podzbiorów $\mathcal{D}$ nazywamy zbiór

$$\prod_{t\in T} A_t = \{f{:}T \to \mathcal{D} \mid \forall t \in T.\, f(t) \in A_t\}$$

Zapiszmy inaczej to, co najważniejsze w tej definicji:

$$f \in \prod_{t\in T} A_t \quad \Leftrightarrow \quad \mathrm{Dom}(f) = T \quad \wedge \quad \forall t \in T.\, f(t) \in A_t.$$

Przykład: $\prod_{n\in\mathbb{N}}\{0,\ldots,n\} = \{f : \mathbb{N} \to \mathbb{N} \mid \forall n.\, f(n) \leq n\}$.

Uwaga: jeśli $A_t = A$, dla wszystkich $t \in T$, to produkt $\prod_{t\in T} A_t$ jest zbiorem A^T wszystkich funkcji z T do A.

5 Relacje równoważności

Relacja równoważności jest zazwyczaj zadana przez jakieś kryterium klasyfikacji przedmiotów ze względu na pewną cechę. Przedmioty są w relacji jeśli mają tę cechę wspólną, tj. kryterium ich nie rozróżnia. Zwykle prowadzi to do utożsamiania przedmiotów „nierozróżnialnych" i tworzenia pojęć abstrakcyjnych, np. „wektor swobodny", „kierunek". W tym przypadku słowo „abstrakcja" należy rozumieć jako oderwanie od pozostałych cech przedmiotów, które są nieistotne z punktu widzenia naszego kryterium.

Definicja 5.1 Dwuargumentowa relacja r w zbiorze A jest *relacją równoważności* wtedy i tylko wtedy, gdy jest zwrotna, symetryczna i przechodnia (definicja 3.3), to jest:

- $\forall x{\in}A\,(x\,r\,x)$;
- $\forall x, y{\in}A\,(x\,r\,y \to y\,r\,x)$;
- $\forall x, y, z{\in}A(x\,r\,y \wedge y\,r\,z \to x\,r\,z)$.

Relacja symetryczna i przechodnia jest nazywana *częściową relacją równoważności*. Relacja taka może zajść tylko pomiędzy elementami należącymi do jej *dziedziny*, tj. do zbioru

$$\mathrm{Dom}(r) = \{a \in A \mid \exists b{\in}A.\, a\,r\,b\}.$$

Jeśli r jest relacją równoważności w A, to oczywiście $\mathrm{Dom}(r) = A$.

Fakt 5.2 *Jeśli r jest częściową relacją równoważności w A, oraz $a \in \mathrm{Dom}(r)$, to $\langle a, a\rangle \in r$.*

Dowód: Jeśli $a \in \mathrm{Dom}(r)$ to $a\,r\,b$ dla pewnego $b \in A$. Wtedy $b\,r\,a$ na mocy symetrii, a stąd $a\,r\,a$, z przechodniości. □

Wniosek 5.3 *Każda częściowa relacja równoważności r jest relacją równoważności w* $\mathrm{Dom}(r)$.

Przykładami relacji równoważności są równoległość prostych, podobieństwo figur geometrycznych, przystawanie wektorów. Skrajne przykłady relacji równoważności to relacja identycznościowa $\mathbf{1}_A = \{\langle x, x\rangle \mid x \in A\}$ i relacja pełna (totalna) $A \times A$. Szczególnym przykładem (częściowej) relacji równoważności jest *jądro* dowolnego (częściowego) przekształcenia f, czyli relacja $\ker(f)$ zadana (dla $x, y \in \mathrm{Dom}(f)$) przez warunek

$$\langle x, y\rangle \in \ker(f) \quad \Leftrightarrow \quad f(x) = f(y).$$

Akt *abstrakcji*, polegający na utożsamieniu elementów zbioru A pozostających ze sobą w pewnej (częściowej) relacji równoważności, powołuje do życia nowe abstrakcyjne obiekty przedstawiające elementy zbioru $\mathrm{Dom}(r)$ „z dokładnością" do r: kierunek prostej, wektor swobodny... Jeśli przez $a/_r$ oznaczymy taki *abstrakt* elementu $a \in A$, to zachodzi równoważność:

$$a/_r = b/_r \qquad \text{wtedy i tylko wtedy, gdy} \qquad a\,r\,b.$$

Jeśli r jest (częściową) relacją równoważności w A, to abstrakty elementów zbioru A wyznaczone przez relację r tworzą zbiór (a właściwie nowy typ) zwany *zbiorem ilorazowym* relacji r.

$$A/_r = \{a/_r \mid a \in \mathrm{Dom}(r)\}$$

Zasada abstrakcji

Definicja 5.4 *Klasą abstrakcji* (częściowej) relacji równoważności r w zbiorze A, wyznaczoną przez element $x \in A$, nazywamy zbiór $[x]_r = \{y \in A \mid x\, r\, y\}$.

Na przykład relacja przystawania modulo 3 w zbiorze liczb naturalnych ma trzy klasy abstrakcji, każda złożona z liczb o innej reszcie z dzielenia przez 3. A klasami abstrakcji relacji identycznościowej są wszystkie singletony.

Fakt 5.5 *Niech $r \subseteq A \times A$ będzie częściową relacją równoważności w A.*

1) *Jeśli $x \in \mathrm{Dom}(r)$ to $x \in [x]_r$.*

2) *Jeśli $x, y \in \mathrm{Dom}(r)$ to następujące warunki są równoważne:*

   a) $x\, r\, y$;
   b) $x \in [y]_r$;
   c) $y \in [x]_r$;
   d) $[x]_r = [y]_r$;
   e) $[x]_r \cap [y]_r \neq \varnothing$.

Dowód: Część (1) wynika natychmiast z faktu 5.2. W części (2) równoważność warunków (a), (b) i (c) wynika wprost z definicji i z tego, że relacja jest symetryczna.

(a)⇒(d) Załóżmy, że $x\, r\, y$ i niech $t \in [x]_r$. Wtedy $x\, r\, t$, więc też $y\, r\, t$ z przechodniości i symetrii. A więc pokazaliśmy inkluzję $[x]_r \subseteq [y]_r$. Inkluzji odwrotnej dowodzimy analogicznie.

(d)⇒(e) Skoro $x \in [x]_r = [y]_r$, na mocy części (1), to $x \in [x]_r \cap [y]_r$.

(e)⇒(a) Jeśli $t \in [x]_r \cap [y]_r$, to $x\, r\, t$ oraz $y\, r\, t$. Z symetrii i przechodniości wynika $x\, r\, y$. □

Z powyższego natychmiast otrzymujemy równoważność:

$$[x]_r = [y]_r \qquad \text{wtedy i tylko wtedy, gdy} \qquad x/_r = y/_r.$$

Oznacza to, że mamy naturalną odpowiedniość pomiędzy elementami $A/_r$ i klasami abstrakcji, zadaną przez bijekcję i ze zbioru klas abstrakcji r do zbioru $A/_r$. Bijekcja ta jest określona warunkiem $i([x]_r) = x/_r$. Można powiedzieć, że klasa abstrakcji $[x]_r$ stanowi „implementację" pojęcia abstraktu $x/_r$ w języku teorii zbiorów. Taka implementacja z powodzeniem może odgrywać rolę abstraktu, a typ ilorazowy $A/_r$ można utożsamiać ze zbiorem klas abstrakcji. Od tej pory skwapliwie korzystamy z tej możliwości i uważamy, że zachodzą równości:

$$x/_r = [x]_r \qquad \text{oraz} \qquad A/_r = \{[a]_r \mid a \in \mathrm{Dom}(r)\}.$$

Z faktu 5.5 wynika też, że każda (częściowa) relacja równoważności wyznacza podział swojej dziedziny na rozłączne zbiory. Na odwrót to też prawda: każdy podział określa jednoznacznie relację równoważności. Aby uściślić tę obserwację zaczniemy od definicji podziału.

Definicja 5.6 *Podziałem* zbioru A nazywamy rodzinę $P \subseteq \mathsf{P}(A)$, która spełnia warunki:

- $\forall p (p \in P \to p \neq \varnothing)$;

- $\forall p, q(p, q \in P \to (p = q \vee p \cap q = \varnothing))$;
- $\bigcup P = A$, czyli $\forall x(x{\in}A \to \exists p{\in}P\,(x \in p))$.

Twierdzenie 5.7 (Zasada abstrakcji)

1) Jeżeli r jest relacją równoważności w A to $A/_r$ jest podziałem zbioru A.

2) Jeżeli P jest podziałem zbioru A, to istnieje taka relacja równoważności r w A, że $P = A/_r$.

Dowód: Część (1) wynika łatwo z faktu 5.5. Dla dowodu części (2), rozpatrzmy dowolny podział P zbioru A i niech r będzie taką relacją:

$$r = \{\langle x, y\rangle \in A \times A \mid \exists p{\in}P\,(x \in p \wedge y \in p)\}$$

Najpierw zauważmy, że r jest relacją równoważności w zbiorze A. Zwrotność wynika z warunku $\bigcup P = A$, a symetria bezpośrednio z definicji r. Pozostaje przechodniość. Przypuśćmy więc, że $x\,r\,y$ i $y\,r\,z$. Wtedy są takie $p, q \in P$, że $x, y \in p$ oraz $y, z \in q$. Ale wtedy $p \cap q \neq \varnothing$, co implikuje $p = q$. Skoro więc $x \in p$ i $z \in q = p$, to $x\,r\,z$.

Następna obserwacja jest taka:

$$\textit{Jeśli } x \in p \in P \textit{ to } [x]_r = p. \qquad (*)$$

Dla dowodu (*) przypuśćmy, że $x \in p \in P$ i niech $t \in [x]_r$. Wtedy $x, t \in q$ dla pewnego $q \in P$. Ale $q = p$ bo $x \in p \cap q$. Zatem $t \in p$ i wykazaliśmy już, że $[x]_r \subseteq p$. Na odwrót, jeśli $t \in p$, to $t\,r\,x$ (bo $x \in p$) więc $t \in [x]_r$.

Teraz wreszcie pokażemy, że $P = A/_r$.

($\subseteq$): Jeśli $p \in P$, to $p \neq \varnothing$, więc jest $x \in p$. Wtedy $p = [x]_r$ na mocy (*), więc $p \in A/_r$.

($\supseteq$): Dla dowolnego $x \in A$ istnieje takie $p \in P$, że $x \in p$. Wtedy $[x]_r = p$. A zatem każda klasa $[x]_r \in A/_r$ należy do P. □

Sens zasady abstrakcji jest taki: relacje równoważności i podziały zbioru to w istocie to samo. Jedno determinuje drugie i na odwrót.

Pewnik wyboru

Niech r będzie częściową relacją równoważności w typie $\mathcal{D}$. Z typem ilorazowym $\mathcal{D}/_r$ wiążemy dwa naturalne przekształcenia:

- *kanoniczną surjekcję* $\kappa : \mathrm{Dom}(r) \to \mathcal{D}/_r$, określoną wzorem $\kappa(a) = [a]_r$;
- *funkcję wyboru* $\sigma : \mathcal{D}/_r \to \mathcal{D}$, o własności $\sigma([a]_r) \in [a]_r$, dla dowolnego $a \in \mathcal{D}$.

Zauważmy, że $[\sigma(x)]_r = x$ dla dowolnego $x \in \mathcal{D}/_r$, czyli $\kappa \circ \sigma = \mathrm{id}_{\mathcal{D}/_r}$.

Kanoniczna surjekcja jest wyznaczona jednoznacznie, funkcji wyboru może być wiele. Założenie, że dla każdego typu ilorazowego istnieje funkcja wyboru nazywane jest *aksjomatem wyboru* (lub *pewnikiem wyboru*). To założenie nie jest wcale oczywiste, bo nie zawsze jest możliwe określenie konkretnej funkcji wyboru. Z tego powodu mówi się o „niekonstruktywnym" charakterze aksjomatu wyboru. Na dodatek pewnik wyboru ma różne zaskakujące konsekwencje i dlatego czasami budzi kontrowersje. Niemniej życie bez pewnika wyboru byłoby bardzo uciążliwe, co się niebawem okaże.

Własność $\sigma(K) \in K$, dla $K \in \mathrm{Dom}(\sigma)$, sugeruje następujące uogólnienie pojęcia funkcji wyboru. Jeśli $\mathcal{R}$ jest rodziną podzbiorów $\mathcal{D}$ (niekoniecznie rozłączną), to *funkcją wyboru dla* $\mathcal{R}$ nazywamy dowolną funkcję $f : \mathcal{R} \to \mathcal{D}$ spełniającą dla wszystkich $A \in \mathcal{R}$ warunek $f(A) \in A$.

Twierdzenie 5.8 *Dla dowolnej rodziny $\mathcal{R}$ zbiorów niepustych istnieje funkcja wyboru.*

Dowód: Załóżmy, że $\mathcal{R} \subseteq \mathsf{P}(\mathcal{D})$ i niech $\mathcal{T} = \mathcal{D} \times \mathsf{P}(\mathcal{D})$. Dla $A \in \mathcal{R}$ niech X_A oznacza zbiór $\{\langle x, A\rangle \mid x \in A\}$ (jest to podzbiór $\mathcal{T}$). Zauważmy, że zbiory X_A i X_B są rozłączne zawsze gdy tylko $A \neq B$. Dalej niech $Z : \mathsf{P}(\mathsf{P}(\mathcal{T}))$ będzie rodziną wszystkich zbiorów postaci X_A, tj. niech $Z = \{X_A \mid A \in \mathcal{R}\}$. Dla formalistów możemy napisać:

$$Z = \{V : \mathsf{P}(\mathcal{T}) \mid \exists A{:}\mathsf{P}(\mathcal{D})(A \in \mathcal{R} \wedge \forall b{:}\mathcal{D}\forall B{:}\mathsf{P}(\mathcal{D})(\langle b, B\rangle \in V \leftrightarrow B = A \wedge b \in A))\}.$$

Rodzina Z składa się z niepustych i parami rozłącznych zbiorów, stanowi więc podział swojej własnej sumy $\bigcup Z$. Na mocy zasady abstrakcji mamy więc częściową relację równoważności r w typie $\mathcal{T}$, dla której zachodzi $Z = \mathcal{T}/_r$. Dalej, jeśli $\sigma : \mathcal{T}/_r \to \mathcal{T}$ jest funkcją wyboru dla rodziny $\mathcal{T}/_r$, to funkcja wyboru dla rodziny $\mathcal{R}$ może być określona jako $\lambda A.\, \pi_1(\sigma(X_A))$. □

Oto jeszcze inna wersja aksjomatu wyboru. Zbiór $S \subseteq \bigcup X$ nazywamy *selektorem* dla rodziny zbiorów X, jeżeli S ma dokładnie po jednym elemencie wspólnym z każdym zbiorem rodziny X, tj.:

$$\forall a \in X \exists t \in a\, (S \cap a = \{t\}).$$

Na przykład zbiór $\{1, 3, 4\}$ jest selektorem dla rodziny $\{\{1, 2\}, \{3, 5\}, \{4, 5\}\}$, natomiast rodzina $\{\{1\}, \{2\}, \{1, 2\}\}$ nie ma selektora.

Fakt 5.9 *Dla dowolnej rodziny X niepustych zbiorów parami rozłącznych*[10] *istnieje selektor.*

Dowód: Rodzina X stanowi podział zbioru $\bigcup X$, wyznacza więc pewną częściową relację równoważności. Zbiór wartości funkcji wyboru dla tej relacji jest żądanym selektorem. □

Uwaga: Fakt 5.9 podkreśla niekonstruktywność aksjomatu wyboru. Selektor to przecież zbiór, a zbiór powinien być określony przez pewien predykat (kryterium przynależności). Ale kryterium definiujące selektor pozostaje nieustalone!

Następujące dwa twierdzenia demonstrują znaczenie aksjomatu wyboru. Jest on niezbędny do uzasadnienia pewnych intuicyjnie oczywistych własności.

Twierdzenie 5.10 *Jeśli $\{A_t\}_{t\in T}$ jest rodziną indeksowaną zbiorów niepustych, to produkt $\Pi_{t\in T} A_t$ jest niepusty.*

Dowód: Niech φ będzie funkcją wyboru dla $\{A_t \mid t \in T\}$, gdzie $A_t : \mathsf{P}(\mathcal{D})$ i niech $f : T \to \mathcal{D}$ będzie określona przez równanie $f(t) = \varphi(A_t)$, dla $t \in T$. Oczywiście $f \in \prod_{t\in T} A_t$. □

Twierdzenie 5.11 *Załóżmy, że $A \neq \varnothing$. Wtedy:*

[10] Mówimy, że rodzina X jest *rozłączna* lub jest rodziną zbiorów *parami rozłącznych*, gdy zachodzi warunek $\forall a, b \in X(a \neq b \to a \cap b = \varnothing)$.

1) Jeśli $f : A \xrightarrow{1-1} B$ *to istnieje taka funkcja* $g : B \xrightarrow{\text{na}} A$*, że* $g \circ f = \mathrm{id}_A$*.*

2) Jeśli $g : B \xrightarrow{\text{na}} A$ *to istnieje taka funkcja* $f : A \xrightarrow{1-1} B$*, że* $g \circ f = \mathrm{id}_A$*.*

Dowód: (1) Skoro $A \neq \varnothing$, to mamy jakiś element $\alpha \in A$. A skoro funkcja f jest różnowartościowa, to istnieje funkcja odwrotna $f^{-1} : \mathrm{Rg}(f) \xrightarrow[\text{na}]{1-1} A$. Możemy więc tak zdefiniować $g(b)$, dla $b \in B$:

$$g(b) = \begin{cases} f^{-1}(b), & \text{jeśli } b \in \mathrm{Rg}(f); \\ \alpha, & \text{w przeciwnym przypadku.} \end{cases}$$

(2) Dla $a \in A$, niech $F_a = g^{-1}(\{a\})$. Zbiory F_a są niepuste, więc produkt $\Pi_{a \in A} F_a$ jest niepusty. Jeśli funkcja f jest elementem tego produktu, to dla dowolnego $a \in A$ mamy $g(f(a)) = a$, bo $f(a) \in F_a$. Ponadto $f : A \xrightarrow{1-1} B$, bo zbiory $F_a = g^{-1}(\{a\})$ są rozłączne i zawarte w B. □

Wniosek 5.12 *Jeśli* $A \neq \varnothing$*, to następujące warunki są równoważne:*

1) Istnieje funkcja $f : A \xrightarrow{1-1} B$*;*

2) Istnieje funkcja $g : B \xrightarrow{\text{na}} A$*.*

6 Liczby naturalne

Mimo że pojęcie liczby naturalnej wydaje się intuicyjnie oczywiste, ścisła definicja liczb naturalnych nie jest wcale łatwa. Jednym ze sposobów jest podejście aksjomatyczne, z którym najczęściej wiążemy nazwisko Giuseppe Peano. Aksjomaty Peana liczb naturalnych są takie:

- *Zero* jest liczbą naturalną.[11]
- Każda liczba naturalna ma *następnik*, który jest liczbą naturalną.
- Liczby o tych samych następnikach są równe.
- Zero nie jest następnikiem żadnej liczby naturalnej.
- Jeśli zero ma pewną własność W, oraz
  - z tego że jakaś liczba naturalna ma własność W wynika, że jej następnik też ma własność W,

  to każda liczba naturalna ma własność W.

Podobną charakteryzację liczb naturalnych podał Richard Dedekind. W obu przypadkach idea jest następująca. Liczba naturalna to albo zero („element pierwotny"), albo następnik $s(n)$ innej liczby naturalnej n (którą należy znać wcześniej). Inaczej: zaczynając od zera, każde kolejne użycie operacji następnika tworzy nową liczbę naturalną, różną od wszystkich poprzednich. Na przykład, liczba 1 to $s(0)$, liczba 2 to $s(s(0))$, i tak dalej. W ten sposób otrzymujemy wszystkie liczby naturalne, innych nie ma.

Nasze założenia o typie liczb naturalnych $\mathbb{N}$ to są zgodne z duchem aksjomatów Peana i z myślą Dedekinda. Poniżej, symbol 0 oznacza zero, a następnik liczby n jest oznaczany przez $s(n)$.

1. Zakładamy, że $0 : \mathbb{N}$ oraz, że
2. jeśli $n : \mathbb{N}$ to także $s(n) : \mathbb{N}$.

Mamy więc operację następnika $s : \mathbb{N} \to \mathbb{N}$. Przyjmujemy, że ma ona następujące własności:

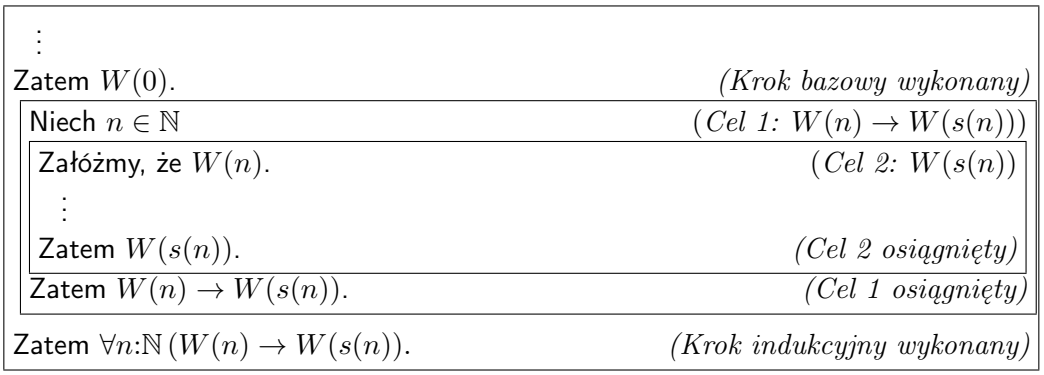
3. Jest różnowartościowa: $\forall mn{:}\mathbb{N}(s(n) = s(m) \to n = m)$.
4. Zero nie należy do jej zbioru wartości: $\forall m{:}\mathbb{N}, \neg s(m) = 0$.

Piąty warunek zwany jest *schematem* (lub *zasadą*) *indukcji* i wyraża w sposób pośredni stwierdzenie, że wszystkie liczby naturalne powstają z zera przez iterowanie następnika (nie ma żadnych innych liczb naturalnych).

5. Jeśli $W(0)$ oraz $\forall n{:}\mathbb{N}(W(n) \to W(s(n)))$ to $\forall n{:}\mathbb{N}.\, W(n)$.

Zasada indukcji jest podstawową metodą wnioskowania o własnościach liczb naturalnych. Schemat logiczny wnioskowania przez indukcję przedstawimy z pomocą pudełek Jaśkowskiego:

[11] Peano zaczynał liczyć od jedynki.



Oto pierwsze przykłady. Niech $\mathbf{s}$ oznacza relację następnika w $\mathbb{N}$, tj. niech $n\,\mathbf{s}\,m$ zachodzi wtedy i tylko wtedy, gdy $m = s(n)$. Przypomnijmy, że symbol $\mathbf{s}^+$ oznacza domknięcie przechodnie relacji $\mathbf{s}$, a symbol $\mathbf{s}^*$ jej domknięcie przechodnio-zwrotne.

Fakt 6.1 *Dla każdej liczby naturalnej n zachodzi związek $0\,\mathbf{s}^*n$.*

Dowód: Mamy udowodnić, że $\forall n : \mathbb{N}.\, W(n)$, gdzie $W(n)$ oznacza $0\,\mathbf{s}^*n$. Krok bazowy: ponieważ relacja $\mathbf{s}^*$ jest zwrotna, więc $0\,\mathbf{s}^*0$, czyli $W(0)$ zachodzi. Krok indukcyjny: przypuśćmy, że liczba n spełnia założenie indukcyjne $W(n)$, czyli że $0\,\mathbf{s}^*n$. Ponieważ $n\,\mathbf{s}^*s(n)$ i relacja $\mathbf{s}^*$ jest przechodnia, więc $0\,\mathbf{s}^*s(n)$. Udowodniliśmy więc, że $W(n)$ implikuje $W(s(n))$. □

Fakt 6.2 *Dla żadnej liczby naturalnej n nie zachodzi związek $n\,\mathbf{s}^+n$. W szczególności zawsze mamy $s(n) \neq n$.*

Dowód: Gdyby $0\,\mathbf{s}^+0$ to, na mocy faktu 3.10, mielibyśmy $0\ \mathbf{s}^*\,y\ \mathbf{s}\ 0$, czyli $0 = s(y)$ dla pewnego y – sprzeczność. Załóżmy więc, że $n\ \mathbf{s}^+n$ nie zachodzi i przypuśćmy, że $s(n)\ \mathbf{s}^+s(n)$. Mamy znowu $s(n)\ \mathbf{s}^*\,y\ \mathbf{s}\ s(n)$ dla pewnego y, ale wtedy $s(y) = s(n)$ więc $y = n$. A zatem $n\ \mathbf{s}\ s(n)\ \mathbf{s}^*\,y = n$, czyli $\langle n, n\rangle$ należy do złożenia relacji $\mathbf{s}$ z relacją $\mathbf{s}^*$. Z faktu 3.10 wiemy, że tym złożeniem jest $\mathbf{s}^+$ i znowu dostajemy sprzeczność. □

Strukturę naszego dowodu indukcyjnego widzimy na następnym rysunku:

Załóżmy, że $0\,\mathbf{s}^+0$. *(Cel 1: sprzeczność)*
⋮
Sprzeczność. *(Cel 1 osiągnięty)*

Zatem $\neg\,0\,\mathbf{s}^+0$. *(Krok bazowy wykonany)*

Niech $n \in \mathbb{N}$ *(Cel 2:* $\neg(n\,\mathbf{s}^+n) \to \neg(s(n)\,\mathbf{s}^+s(n))$*)*

Załóżmy, że $\neg(n\,\mathbf{s}^+n)$. *(Cel 3:* $\neg(s(n)\,\mathbf{s}^+s(n))$*)*

Załóżmy, że $s(n)\,\mathbf{s}^+s(n)$. *(Cel 4: sprzeczność)*
⋮
Sprzeczność. *(Cel 4 osiągnięty)*

Zatem $\neg\,s(n)\,\mathbf{s}^+s(n)$. *(Cel 3 osiągnięty)*

Zatem $\neg\,(n\,\mathbf{s}^+n) \to \neg\,(s(n)\,\mathbf{s}^+s(n))$ *(Cel 2 osiągnięty)*

Zatem $\forall n{:}\mathbb{N}\,(\neg(n\,\mathbf{s}^+n) \to \neg(s(n)\,\mathbf{s}^+s(n)))$ *(Krok indukcyjny wykonany)*

Zatem $\forall n{:}\mathbb{N}\,\neg(n\,\mathbf{s}^+n)$

Ćwiczenie 6.3 *Udowodnić, że* $\forall n{:}\mathbb{N}(n = 0 \vee \exists m{:}\mathbb{N}.\, n = s(m))$.

Definicja 6.4 Relację (nieostrej) nierówności $\leq$ pomiędzy liczbami naturalnymi można zdefiniować jako domknięcie przechodnio-zwrotne $\mathbf{s}^*$ relacji następnika:

$$m \leq n \quad \text{wtedy i tylko wtedy, gdy} \quad m\,\mathbf{s}^*n.$$

Nierówność ostra jest pojęciem wtórnym w stosunku do relacji $\leq$:

$$m < n \quad \text{wtedy i tylko wtedy, gdy} \quad m \leq n \text{ ale } m \neq n.$$

Fakt 6.5 *Relacje* $\mathbf{s}^+$ *i* $<$ *pokrywają się.*

Dowód: Wystarczy pokazać, że warunek $m\,\mathbf{s}^+n$ zachodzi wtedy i tylko wtedy, gdy $m\,\mathbf{s}^*n$ oraz $m \neq n$. Implikacja z lewej do prawej wynika z faktu 6.2, w przeciwną stronę z tego, że relacja $\mathbf{s}^*$ jest sumą $\mathbf{s}^+$ i identyczności (fakt 3.10). □

Fakt 6.6 *Relacja* $\leq$ *jest relacją liniowego porządku w* $\mathbb{N}$*, tj. jest zwrotna, przechodnia, antysymetryczna i spójna. Zero jest elementem najmniejszym, tj.* $\forall m.\, 0 \leq m$.

Dowód: Zwrotność i przechodniość wynikają wprost z definicji. Aby wykazać antysymetrię, przypuśćmy, że $n \leq m \leq n$, gdzie $m \neq n$. Z wniosku 6.5 wynika, że w istocie zachodzi $n\ \mathbf{s}^+\ m\ \mathbf{s}^+\ n$, skąd $n\ \mathbf{s}^+\ n$, co jest niemożliwe (fakt 6.2).

Własność $\forall m.\, 0 \leq m$ wynika z faktu 6.1.

Spójność, czyli warunek $\forall nm \in \mathbb{N}(m \leq n \vee n \leq m)$ udowodnimy przez indukcję *ze względu na* n, tj. pokażemy, że każde $n \in \mathbb{N}$ ma własność

$$\forall m \in \mathbb{N}(m \leq n \vee n \leq m).$$

Dla $n = 0$ mamy zawsze $n \leq m$, załóżmy więc, że $\forall m \in \mathbb{N}(m \leq n \vee n \leq m)$ i pokażmy, że wtedy także $\forall m \in \mathbb{N}(m \leq s(n) \vee s(n) \leq m)$. Niech $m \in \mathbb{N}$. Jeśli $m \leq n$, to tym bardziej

$m \leq s(n)$. W przeciwnym razie $n \leq m$, ale przypadek $n = m$ już jest rozpatrzony, więc możemy przyjąć, że w istocie $n\ \mathbf{s}^{+} m$. Z faktu 3.10 wynika $n\ \mathbf{s}\ s(n) \leq m$. □

Twierdzenie 6.7 (Zasada minimum) *Każdy niepusty podzbiór $A \subseteq \mathbb{N}$ ma element najmniejszy, tj. taki element $a \in A$, że $\forall b\,(b \in A \to a \leq b)$.*

Dowód: Przypuśćmy, że $A \subseteq \mathbb{N}$ nie ma najmniejszego elementu. Przez indukcję ze względu na n dowodzimy, że $\forall n \forall k (k \in A \to n < k)$. Stąd już wynika, że $A = \varnothing$.

Najpierw zauważmy, że $0 \notin A$. W przeciwnym razie 0 byłoby oczywiście najmniejszym elementem. A więc $\forall k (k \in A \to 0 < k)$.

Załóżmy, że $\forall k (k \in A \to n < k)$. Wtedy $\forall k (k \in A \to s(n) \leq k)$, bo na mocy faktów 3.10 i 6.5 nierówność $n < k$ to to samo co $s(n) \leq k$. Gdyby więc $s(n) \in A$ to $s(n)$ byłoby najmniejszym elementem A. No to $s(n) \notin A$ i warunek można wzmocnić: $\forall k (k \in A \to s(n) < k)$. □

Najmniejszy element zbioru $A \subseteq \mathbb{N}$ oznaczamy przez $\min A$.

Wniosek 6.8 (Trochę inna zasada indukcji)

$$\textit{Jeśli} \quad \forall n{:}\mathbb{N}(\forall m{:}\mathbb{N}(m < n \to W(m)) \to W(n)) \quad \textit{to} \quad \forall n{:}\mathbb{N}.\, W(n).$$

Dowód: Niech $A = \{n : \mathbb{N} \mid \neg W(n)\}$. Jeśli teza nie zachodzi, to zbiór A jest niepusty, ma więc element najmniejszy n. Wtedy zachodzi $\forall m{:}\mathbb{N}(m < n \to W(m))$, ale nie jest spełniony warunek $W(n)$, co jest sprzeczne z założeniem. □

A więc, aby udowodnić, że każda liczba naturalna spełnia pewien warunek (należy do pewnego zbioru B), wystarczy stwierdzić taką prawidłowość:

Jeśli wszystkie liczby mniejsze od pewnego n należą do B, to także $n \in B$.

Definiowanie przez indukcję

Indukcja to nie tylko metoda dowodzenia twierdzeń, to także metoda definiowania pojęć i obiektów, zwłaszcza funkcji. Jeśli chcemy określić funkcję f o dziedzinie $\mathbb{N}$ to powinniśmy wskazać jej wartości dla wszystkich argumentów $n : \mathbb{N}$. Możemy oczywiście zacząć od zera. Przypuśćmy dalej, że znając wartość $f(x)$ dla jakiegokolwiek x, potrafimy zawsze wskazać wartość $f(s(x))$. Na przykład rozpatrzmy takie dwa równania:

$$\begin{aligned} f(0) &= 0; \\ f(s(m)) &= s(s(f(m)). \end{aligned}$$

Równania te nie stanowią definicji warunkowej funkcji f, bo symbol f występuje w nich także po prawej stronie. Formalnie, są to więc jakby „postulaty", które żądana funkcja powinna spełniać. Ale te postulaty określają jednoznacznie wartość funkcji dla każdego $x : \mathbb{N}$. (Możemy uzasadnić przez indukcję, że dla dowolnego $x : \mathbb{N}$ wartość $f(x)$ jest dobrze określona.)

Taki sposób definiowania może być zastosowany także do funkcji dwu- i więcej argumentowych. Oto dwa najważniejsze przykłady indukcji *ze względu na pierwszy argument*:

$$\begin{aligned} 0 + n &= n; & 0 \cdot n &= 0; \\ s(m) + n &= s(m + n). & s(m) \cdot n &= m \cdot n + n. \end{aligned}$$

Przykład 6.9 Możemy teraz policzyć ile jest dwa razy dwa: $2 \cdot 2 = 1 \cdot 2 + 2 = (0 \cdot 2 + 2) + 2 = (0 + 2) + 2 = 2 + 2 = s(1 + 2) = s(s(0 + 2)) = s(s(2)) = s(s(s(s(0)))) = 4$.

Następujący lemat formułuje własność znaną jako łączność dodawania. Przemienność dodawania, łączność i przemienność mnożenia pozostawiamy jako ćwiczenie.

Lemat 6.10 *Dla dowolnych liczb $m, k, l \in \mathbb{N}$ zachodzi równość $m + (k + l) = (m + k) + l$.*

Dowód: Udowodnimy tezę przez indukcję ze względu na m. Inaczej mówiąc udowodnimy, ze każda liczba $m : \mathbb{N}$ ma własność

$$\forall k, l : \mathbb{N}.\, m + (k + l) = (m + k) + l.$$

Po pierwsze, $0 + (k + l) = (k + l) = (0 + k) + l$, po drugie z warunku $m + (k + l) = (m + k) + l$ wynika $s(m) + (k + l) = s(m + (k + l)) = s((m + k) + l) = s(m + k) + l = (s(m) + k) + l$. □

Dodawanie i mnożenie są przykładami funkcji definiowanych za pomocą tzw. *rekursji prostej*. Ogólny schemat rekursji prostej wygląda tak:

$$\begin{aligned} f(0, n_1, \ldots, n_k) &= g(n_1, \ldots, n_k); \\ f(s(m), n_1, \ldots, n_k) &= h(m, n_1, \ldots, n_k, f(m, n_1, \ldots, n_k)). \end{aligned}$$

Tutaj definiujemy funkcję $f : \mathbb{N}^{k+1} \to \mathbb{N}$ przez indukcję ze względu na pierwszy argument, z pomocą już określonych funkcji $g : \mathbb{N}^k \to \mathbb{N}$ i $h : \mathbb{N}^{k+2} \to \mathbb{N}$.

Nietrywialnym uogólnieniem rekursji prostej jest definiowanie przez indukcję funkcji o argumentach i wartościach dowolnych typów. Oto schemat definiowania funkcji $f : \mathbb{N} \times \mathcal{D} \to \mathcal{E}$.

$$f(0, d) = g(d); \tag{6.1}$$

$$f(s(m), d) = h(m, d, f(m, d)). \tag{6.2}$$

Przykład 6.11 Istnieją różne schematy definiowania indukcyjnego. Na przykład poniższa definicja *funkcji Ackermanna-Sudana* używa „rekursji podwójnej" a nie rekursji prostej:

$$\begin{aligned} A(0, x) &= s(x); \\ A(s(n), 0) &= A(n, 1); \\ A(s(n), s(x)) &= A(n, A(s(n), x)). \end{aligned}$$

Można pokazać (łatwe ćwiczenie), że wartość $A(n, x)$ jest dobrze określona dla każdych n i x (tj. że proces obliczenia $A(n, x)$ wg powyższych reguł musi się zakończyć). Funkcja Ackermanna nie może być jednak zdefiniowana z pomocą rekursji prostej.

Ćwiczenie 6.12 Jak dużą liczbą jest $A(5, 4)$? [12]

Definicja 6.13 Za pomocą dodawania można także zdefiniować relację nierówności pomiędzy liczbami naturalnymi.

$$m \leq n \quad \text{wtedy i tylko wtedy, gdy} \quad \exists k(k + m = n).$$

[12] Odpowiedź: Bardzo dużą.

Fakt 6.14 *Definicje 6.4 i 6.13 są równoważne, tj.* $m\,\mathbf{s}^*n$ *zachodzi wtedy i tylko wtedy, gdy* $k+m=n$ *dla pewnego* k*.*

Dowód: Implikacja z lewej do prawej zachodzi dlatego, że relacja „$\exists k(k+m=n)$" jest przechodnia (jeśli $k+m=n$ i $l+n=p$ to $(l+k)+m=p$ na mocy łączności dodawania). Implikację z prawej do lewej można udowodnić przez indukcję ze względu na k. Dokładniej, pokażemy, że każda liczba $k:\mathbb{N}$ ma własność $\forall m{:}\mathbb{N}\,(m\ \mathbf{s}^*\ k+m)$. Dla zera własność ta wynika wprost z definicji. W kroku indukcyjnym mamy $m\ \mathbf{s}^*(k+m)\ \mathbf{s}\ s(k+m)=s(k)+m$. □

7 Typy ilorazowe i indukcyjne

Konstrukcja liczb całkowitych

Liczby całkowite wymyślono po to, żeby można było odejmować dowolne liczby naturalne. Liczby całkowite to „formalne różnice" $x - y$ dla $x, y \in \mathbb{N}$, które można „implementować" jako pary uporządkowane. Odejmowanie to operacja odwrotna do dodawania, zatem powinna zachodzić równość $(x - y) + y = x$. Ale wtedy także $(x - y) + y + z = x + z$, i widzimy, że różnice $x - y$ oraz $(x + z) - (y + z)$ powinny być takie same. A więc odpowiednie pary należy utożsamiać. Dlatego typ liczb całkowitych powinien być zdefiniowany jako iloraz.

Rozpatrzmy następującą relację w typie $\mathbb{N} \times \mathbb{N}$:

$$\langle m, n\rangle \sim \langle m', n'\rangle \quad \text{wtedy i tylko wtedy, gdy} \quad m + n' = m' + n.$$

Nietrudno zauważyć, że to jest relacja równoważności.[13] Typ ilorazowy $(\mathbb{N} \times \mathbb{N})/_\sim$ nazywamy *typem liczb całkowitych*. Inaczej mówiąc, liczby całkowite to pary liczb naturalnych, brane z dokładnością do relacji $\sim$. Działania na liczbach całkowitych określamy tak:

$$\begin{array}{rcl} [\langle m, n\rangle]_\sim + [\langle m_1, n_1\rangle]_\sim & = & [\langle m + m_1, n + n_1\rangle]_\sim \\ [\langle m, n\rangle]_\sim \cdot [\langle m_1, n_1\rangle]_\sim & = & [\langle mm_1 + nn_1, mn_1 + nm_1\rangle]_\sim \\ -[\langle m, n\rangle]_\sim & = & [\langle n, m\rangle]_\sim \end{array}$$

Uwaga: Te definicje są poprawne, bo jeśli $\langle m, n\rangle \sim \langle m', n'\rangle$ i $\langle m_1, n_1\rangle \sim \langle m'_1, n'_1\rangle$, to:

- $\langle m + m_1, n + n_1\rangle \sim \langle m' + m'_1, n' + n'_1\rangle$;
- $\langle mm_1 + nn_1, mn_1 + nm_1\rangle \sim \langle m'm'_1 + n'n'_1, m'n'_1 + n'm'_1\rangle$;
- $\langle n, m\rangle \sim \langle n', m'\rangle$.

Przy tej definicji, liczby naturalne *nie są* liczbami całkowitymi. Ale możemy się umówić, że tak jest. Mamy bowiem naturalnie określoną koercję $i : \mathbb{N} \xrightarrow{1-1} \mathbb{Z}$, daną warunkiem

$$i(n) = [\langle n, 0\rangle]_\sim$$

i z dużym powodzeniem możemy utożsamiać każdą liczbę naturalną n z liczbą całkowitą $i(n)$. Zauważmy na przykład, że $i(m + n) = i(m) + i(n)$ oraz $i(m \cdot n) = i(m) \cdot i(n)$, a więc arytmetykę liczb naturalnych (a o nią tu przecież głównie chodzi) możemy uprawiać bez przeszkód w zbiorze $\mathrm{Rg}(i) \subseteq \mathbb{Z}$.

Konstrukcja liczb wymiernych

Typ $\mathbb{Q}$ wszystkich liczb wymiernych także zdefiniujemy jako typ ilorazowy. Tym razem chodzi oczywiście o dzielenie liczb całkowitych. Rozważamy częściową relację równoważności $\approx$ w zbiorze par $\mathbb{Z} \times \mathbb{Z}$, daną warunkiem

$$\langle x, y\rangle \approx \langle u, v\rangle \quad \text{wtedy i tylko wtedy, gdy} \quad y, v \neq 0 \ \text{ oraz } \ x \cdot v = u \cdot y,$$

i przyjmujemy $\mathbb{Q} = (\mathbb{Z} \times \mathbb{Z})/_\approx$. Po sprawdzeniu, że $\langle x, y\rangle \approx \langle x', y'\rangle$ i $\langle u, v\rangle \approx \langle u', v'\rangle$ implikuje $\langle xv + yu, yv\rangle \approx \langle x'v' + y'u', y'v'\rangle$ oraz $\langle xu, yv\rangle \approx \langle x'u', y'v'\rangle$, możemy zdefiniować operacje

[13] Wskazówka: udowodnić przez indukcję prawo skracania: *Jeśli* $n + k = m + k$, *to* $n = m$.

na liczbach wymiernych:

$$[\langle x,y\rangle]_{\approx} + [\langle u,v\rangle]_{\approx} = [\langle xv+yu, yv\rangle]_{\approx}$$
$$[\langle x,y\rangle]_{\approx} \cdot [\langle u,v\rangle]_{\approx} = [\langle xu, yv\rangle]_{\approx}$$

Liczby całkowite interpretujemy jako liczby wymierne za pomocą koercji

$$j(z) = [\langle z,1\rangle]_{\approx}.$$

Oczywiście zamiast $[\langle x,y\rangle]_{\approx}$ piszemy $\frac{x}{y}$.

Konstrukcja liczb rzeczywistych

Zobaczmy teraz pokrótce, jak można zdefiniować liczby rzeczywiste. Funkcję $f : \mathbb{N} \to \mathbb{Q}$ nazwiemy *ciągiem Cauchy'ego*, gdy

$$\forall\varepsilon{:}\mathbb{Q}\,(\varepsilon > 0 \to \exists n{:}\mathbb{N}\,\forall k{:}\mathbb{N}(k \geq n \to (f(n) - \varepsilon < f(k) < f(n) + \varepsilon)))$$

W typie $\mathbb{N} \to \mathbb{Q}$ określimy częściową relację równoważności $\equiv$, której dziedziną jest zbiór $\mathcal{C}$ wszystkich ciągów Cauchy'ego:

$$f \equiv g \quad \Leftrightarrow \quad \forall\varepsilon{:}\mathbb{Q}\,(\varepsilon > 0 \to \exists n{:}\mathbb{N}\,\forall k{:}\mathbb{N}(k \geq n \to (f(k) - \varepsilon < g(k) < f(k) + \varepsilon))).$$

Typ $\mathbb{R}$ liczb rzeczywistych definiujemy jako $(\mathbb{N} \to \mathbb{Q})/_{\equiv}$. Działania na liczbach rzeczywistych definiujemy „po współrzędnych". Wynikiem dodawania $([f]_{\equiv}) + ([g]_{\equiv})$ jest więc abstrakt ciągu h, określonego równaniem $h(n) = f(n) + g(n)$. Przyjmujemy, że $\mathbb{Q} \subseteq \mathbb{R}$ poprzez identyfikację każdej liczby wymiernej $q \in \mathbb{Q}$ z ciągiem stałym o wartości q.

Typy indukcyjne

Powróćmy jeszcze do liczb naturalnych. Każda liczba naturalna powstaje na jeden z dwóch sposobów. Albo jest to po prostu zero, albo następnik liczby wcześniej otrzymanej. Mamy więc dwa *konstruktory* liczb naturalnych: zero i następnik. Następnik $s : \mathbb{N} \to \mathbb{N}$ jest konstruktorem jednoargumentowym, stała $0 : \mathbb{N}$ jest konstruktorem bez argumentów. Istotne jest to, że te dwa sposoby wzajemnie się wykluczają, żadna liczba nie jest jednocześnie zerem i następnikiem. Mamy tu szczególny przypadek ogólniejszego zjawiska, kiedy elementy pewnego typu tworzone są przez kilka niezależnych konstruktorów, a przy tym każdy element można otrzymać tylko na jeden sposób. Takie typy nazywamy *indukcyjnymi*.

Słowa

Jako przykład weźmy typ indukcyjny o jednym konstruktorze zeroargumentowym ε i dwóch jednoargumentowych a i b. Każdy element tego typu, na przykład $a(b(a(a(b(\varepsilon)))))$, jest więc otrzymany przez aplikowanie a i b w różny sposób do ε. Możemy taki element reprezentować w zapisie beznawiasowym jako $abaab\varepsilon$, a jeszcze lepiej odwrotnie – jako $\varepsilon baaba$. Wreszcie możemy opuścić ε na początku i napisać tylko $baaba$.

Elementy naszego typu indukcyjnego można więc utożsamiać z ciągami liter a i b, przy czym stała ε odpowiada ciągowi pustemu. Takie ciągi nazywamy *słowami* nad alfabetem $\{a, b\}$.

Naturalnym uogólnieniem jest pojęcie słowa nad dowolnym alfabetem Σ. Typ słów nad alfabetem Σ oznaczany jest przez Σ^*. Poniżej dla uproszczenia przyjmujemy, że $\Sigma = \{a, b\}$. Podstawowe własności typu $\{a, b\}^*$ są takie:

- Słowo puste ε jest słowem;
- Jeśli $w : \{a, b\}^*$, to $wa, wb : \{a, b\}^*$;
- Jeśli $wx = vy$, gdzie $w, v : \{a, b\}^*$ oraz $x, y \in \{a, b\}$, to $w = v$ i $x = y$;
- Słowa postaci wa, wb nie są puste.

Główną metodą dowodzenia własności słów i definiowania operacji na słowach jest indukcja. Zasada indukcji dla słów ma postać:

$$\textit{Jeśli}\ \ W(\varepsilon)\ \ \textit{oraz}\ \ \forall w{:}\{a,b\}^*(W(w) \to W(wa) \wedge W(wb))\ \ \textit{to}\ \ \forall w{:}\{a,b\}^*.\, W(w).$$

Przykładem definicji przez indukcję jest zaś poniższa definicja *długości słowa*. Określamy tu funkcję, która każdemu słowu $w : \{a, b\}^*$ przypisuje pewną liczbę $|w| : \mathbb{N}$.

$$|\varepsilon| = 0; \qquad |wa| = |w| + 1; \qquad |wb| = |w| + 1.$$

Widoczne jest podobieństwo pomiędzy słowami długości n i n-krotkami. Dlatego często przyjmuje się utożsamienie $\Sigma^* = \bigcup_{n\in\mathbb{N}} \Sigma^n$, ale trzeba się wtedy umówić, że $\Sigma^0 = \{\varepsilon\}$.

Schemat definiowania przez indukcję funkcji $f : \{a, b\}^* \times \mathcal{D} \to \mathcal{E}$ można napisać podobnie do schematu dla liczb naturalnych (6.1–6.2):

$$\begin{aligned} f(\varepsilon, d) &= g(d); && (7.1)\\ f(wa, d) &= h_a(w, d, f(w, d)); && (7.2)\\ f(wb, d) &= h_b(w, d, f(w, d)). && (7.3) \end{aligned}$$

Ważną operacją na słowach jest *składanie słów* czyli *konkatenacja*. *Konkatenacją* (złożeniem) słów w i v nazywamy słowo $w \cdot v$ powstałe przez dopisanie słowa v na końcu słowa w. Definicja indukcyjna (ze względu na drugi argument) jest taka:

$$w \cdot \varepsilon = w; \qquad w \cdot va = (w \cdot v)a; \qquad w \cdot vb = (w \cdot v)b.$$

Operacja konkatenacji jest łączna, co można łatwo wykazać przez indukcję (ćwiczenie), na przykład:

$$ein \cdot (und \cdot zwanzig) = (ein \cdot und) \cdot zwanzig = einundzwanzig.$$

A oto inne przykłady wnioskowania przez indukcję dla słów.

Fakt 7.1 *Dla dowolnego słowa w zachodzi $\varepsilon \cdot w = w$.*

Dowód: Zgodnie z zasadą indukcji dla słów należy sprawdzić trzy hipotezy:

1. $\varepsilon \cdot \varepsilon = \varepsilon$;
2. Jeśli $\varepsilon \cdot w = w$ to $\varepsilon \cdot wa = wa$;
3. Jeśli $\varepsilon \cdot w = w$ to $\varepsilon \cdot wb = wb$.

Pierwsza z nich (krok bazowy) wynika natychmiast z definicji konkatenacji. W części drugiej (krok indukcyjny, część pierwsza) mamy $\varepsilon \cdot wa = (\varepsilon \cdot w)a$, a to z założenia indukcyjnego jest równe wa. Podobnie w drugiej części kroku indukcyjnego (3). □

Fakt 7.2 *Dla dowolnych słów w i v zachodzi $|w \cdot v| = |w| + |v|$.*

Dowód: Nasz dowód jest przez indukcję ze względu na v. Inaczej mówiąc, udowodnimy, że każde słowo v ma własność

$$\forall w{:}\{a,b\}^*.|w \cdot v| = |w| + |v|.$$

Krok bazowy polega na sprawdzeniu równości $|w \cdot \varepsilon| = |w|$, oczywistej, bo $w \cdot \varepsilon = w$.

W kroku indukcyjnym (przypadek pierwszy) zakładamy, że $|w \cdot v| = |w| + |v|$ dla wszystkich w i liczymy $|w \cdot va| = |(w \cdot v)a| = |w \cdot v| + 1 = |w| + |v| + 1 = |w| + |va|$, korzystając z założenia indukcyjnego. Drugi przypadek jest analogiczny. □

Porządek prefiksowy

Piszemy $w \subseteq v$, gdy słowo w jest przedrostkiem (prefiksem) słowa v, tj. gdy istnieje takie słowo u, że $v = w \cdot u$. Relację $\subseteq$ nazywamy *porządkiem prefiksowym*. Łatwo zauważyć analogię pomiędzy porządkiem prefiksowym i porządkiem $\leq$ w $\mathbb{N}$, mamy bowiem (por. fakt 6.14):

Fakt 7.3 *Niech* $\mathbf{s}$ *oznacza relację następnika dla słów, tj. niech $w\,\mathbf{s}\,v$ zachodzi wtedy i tylko wtedy, gdy $v = wa$ lub $v = wb$. Porządek prefiksowy jest domknięciem przechodnio-zwrotnym relacji* $\mathbf{s}$.

Dowód: Ćwiczenie. □

Fakt 7.4 *Porządek prefiksowy jest częściowym porządkiem w zbiorze słów.*

Dowód: Zwrotność i przechodniość są oczywiste. Antysymetria wynika np. z takiej obserwacji: jeśli $w = vx$ i $v = wy$ to $w = wyx$. A skoro $|w| = |wyx| = |w| + |x| + |y|$ (fakt 7.2), to $x = y = \varepsilon$. □

Analogicznie do porządku prefiksowego można zdefiniować porządek *sufiksowy* o podobnych własnościach.

Listy

Lista liczb naturalnych to skończony ciąg liczb. Lista może być pusta (ozn. nil) lub otrzymana przez dopisanie jakiejś liczby $n : \mathbb{N}$ do już istniejącej listy ℓ, co oznaczamy przez $\mathsf{cons}(n, \ell)$ lub przez $n :: \ell$. A więc listy tworzą typ indukcyjny **list** generowany przez dwa konstruktory:

$$\mathsf{nil} : \mathbf{list}, \qquad \text{oraz} \qquad \mathsf{cons} : \mathbb{N} \times \mathbf{list} \to \mathbf{list}.$$

Zasada indukcji dla list jest następująca:

$$\textit{Jeśli} \quad W(\mathsf{nil}) \quad \textit{oraz} \quad \forall \ell{:}\mathbf{list}(W(\ell) \to \forall n{:}\mathbb{N}.\, W(n :: \ell)) \quad \textit{to} \quad \forall \ell{:}\mathbf{list}.\, W(\ell).$$

Schemat definiowania przez indukcję funkcji $f : \mathbf{list} \times \mathcal{D} \to \mathcal{E}$ ma postać:

$$\begin{array}{rcl} f(\mathsf{nil}, d) & = & g(d); \\ f(n :: \ell, d) & = & h(\ell, n, d, f(\ell, d)), \end{array}$$

gdzie $g : \mathcal{D} \to \mathcal{E}$ oraz $h : \mathbf{list} \times \mathbb{N} \times \mathcal{D} \times \mathcal{E} \to \mathcal{E}$.

Wiele obiektów naturalnie pojawiających się w matematyce i informatyce można uważać za elementy pewnych typów indukcyjnych. Na przykład skończone drzewa binarne tworzą typ indukcyjny z jednym konstruktorem dwuargumentowym i jedną stałą. Nieco bardziej wyrafinowany przykład to typ ω-**tree**, który ma dwa konstruktory

$$\mathsf{leaf} : \omega\text{-}\mathbf{tree} \qquad \text{oraz} \qquad \mathsf{node} : (\mathbb{N} \to \omega\text{-}\mathbf{tree}) \to \omega\text{-}\mathbf{tree}.$$

Elementy tego typu możemy interpretować jako drzewa, w których każdy wierzchołek wewnętrzny ma tyle następników ile jest liczb naturalnych. Pierwszy konstruktor, to drzewo złożone z jednego liścia. Argumentem drugiego konstruktora jest funkcja, która każdej liczbie naturalnej przypisuje drzewo typu ω-**tree**. A zatem $\mathsf{node}(f)$ to drzewo, którego korzeń połączony jest z poddrzewami $f(n)$ dla wszystkich $n \in \mathbb{N}$.

Wspólne cechy typów indukcyjnych to

- jednoznaczność sposobu w jaki każdy element typu jest otrzymany przez działanie konstruktorów;
- swoista zasada indukcji;
- swoisty schemat definiowania przez indukcję.

8 Równoliczność

Definicja 8.1 Mówimy, że zbiory A i B są *równoliczne*, albo że są to zbiory *tej samej mocy* (i piszemy $A \sim B$) wtedy i tylko wtedy, gdy istnieje bijekcja $f : A \xrightarrow[\text{na}]{1-1} B$.

Powyższa definicja opiera się na tym samym pomyśle, którego używa dwoje dzieci nie znających arytmetyki do podzielenia się po równo kasztanami, jabłkami itp. Wystarczy dawać każdemu po jednym, aż do wyczerpania zasobów.

Uwaga 8.2 Pojęcie *mocy zbioru*, inaczej zwanej jego *liczbą kardynalną* jest wygodnym skrótem myślowym. Zamiast mówić, że dwa zbiory są lub nie są równoliczne, mówimy że mają (lub nie) taką samą moc. Jeśli użyjemy znaku $\mathfrak{m}$ na oznaczenie mocy jakiegoś zbioru A, to napis $\overline{\overline{B}} = \mathfrak{m}$ czytamy „B jest mocy $\mathfrak{m}$". Taki napis w istocie oznacza tyle samo co $B \sim A$.

Przykład 8.3

- Przedziały otwarte (a, b) i (c, d) są równoliczne, bo funkcja $f : (a, b) \xrightarrow[\text{na}]{1-1} (c, d)$ może być określona wzorem $f(x) = \frac{d-c}{b-a} \cdot x + \frac{bc-ad}{b-a}$.
- Przedział $(-\frac{\pi}{2}, \frac{\pi}{2})$ (a zatem także każdy inny przedział otwarty) jest równoliczny z całym typem $\mathbb{R}$ liczb rzeczywistych. Dla dowodu wystarczy użyć funkcji tangens.
- Przedziały $(0, 1]$ i $(0, 1)$ są równoliczne, bo mamy taką funkcję $f : (0, 1] \xrightarrow[\text{na}]{1-1} (0, 1)$:

$$f(x) = \begin{cases} \frac{1}{n+1}, & \text{jeśli } x = \frac{1}{n}, \text{ dla pewnego } n \in \mathbb{N}; \\ x, & \text{w przeciwnym przypadku.} \end{cases}$$

- Typ $\mathbb{R}$ jest równoliczny ze zbiorem wszystkich liczb rzeczywistych dodatnich, a równoliczność ustala np. funkcja logarytm.

Fakt 8.4 *Dla dowolnych zbiorów A, B, C,*

- $A \sim A$;
- *Jeśli* $A \sim B$ *to* $B \sim A$;
- *Jeśli* $A \sim B$ *i* $B \sim C$ *to* $A \sim C$.

Uwaga: Z powyższego wynika, że równoliczność zbiorów tego samego typu $\mathsf{P}(\mathcal{D})$ jest relacją równoważności w $\mathsf{P}(\mathcal{D})$.

Teraz jeszcze jeden przydatny lemat.

Lemat 8.5 *Niech $a \notin A$ i $b \notin B$. Wówczas:*

- $A \cup \{a\} \sim B \cup \{b\}$ *wtedy i tylko wtedy, gdy* $A \sim B$.

- *Injekcja* $f : A \cup \{a\} \xrightarrow{1-1} B \cup \{b\}$ *istnieje wtedy i tylko wtedy, gdy istnieje injekcja* $g : A \xrightarrow{1-1} B$.

Dowód: Jeśli $g : A \xrightarrow{1-1} B$, to injekcję $f : A \cup \{a\} \xrightarrow{1-1} B \cup \{b\}$ określamy wzorem

$$f(x) = \begin{cases} b, & \text{jeśli } x = a; \\ g(x), & \text{w przeciwnym przypadku.} \end{cases}$$

Jeśli na dodatek funkcja g jest na B, to także f jest „na". To dowodzi implikacji ($\Leftarrow$) w obu częściach lematu. Przypuśćmy więc, że $f : A \cup \{a\} \xrightarrow{1-1} B \cup \{b\}$. Określimy funkcję $g : A \to B$ definicją warunkową:

$$g(x) = \begin{cases} f(a), & \text{jeśli } f(x) = b; \\ f(x), & \text{w przeciwnym przypadku.} \end{cases}$$

Nietrudno zauważyć, że g jest różnowartościowa, a jeśli f jest „na" to także g jest „na". □

Zbiory skończone

Powiemy, że podzbiór $A \subseteq \mathbb{N}$ jest *ograniczony*, gdy istnieje takie $n : \mathbb{N}$, że $a \leq n$ dla wszystkich $a \in A$. Można wtedy napisać $A \leq n$. Jeśli ograniczenie jest „ostre", tj. dla każdego $a \in A$ zachodzi $a < n$ to napiszemy $A < n$. Oczywiście każdy zbiór ograniczony ma ograniczenia ostre. Szczególnym rodzajem zbioru ograniczonego jest zbiór $\mathcal{O}(n) = \{m : \mathbb{N} \mid m < n\}$ zwany *odcinkiem początkowym wyznaczonym przez* n, który dla prostoty będziemy oznaczać przez $\overline{n}$.

Lemat 8.6 *Niech* $A \subseteq \mathbb{N}$. *Jeśli* A *jest ograniczony, to* $A \sim \overline{n}$ *dla pewnego* $n \in \mathbb{N}$. *A jeśli* A *jest nieograniczony, to* $A \sim \mathbb{N}$.

Dowód: Zdefiniujemy funkcję częściową $f : \mathbb{N} \rightharpoonup A$ równaniem indukcyjnym

$$f(n) = \min(A - f(\overline{n})),$$

które należy rozumieć tak: wartość $f(n)$ jest określona wtedy i tylko wtedy, gdy określone są już wszystkie $f(k)$, dla $k < n$, oraz zbiór $A - f(\overline{n})$ jest niepusty. Są więc dwie możliwości: albo f jest określona dla wszystkich $n \in \mathbb{N}$ (jest funkcją całkowitą) albo istnieje najmniejsza liczba n, dla której $f(n)$ jest nieokreślone. Z definicji, $f(k)$ jest określone dla wszystkich $k < n$ i nieokreślone dla wszystkich $k \geq n$, zatem $\mathrm{Dom}(f) = \overline{n}$.

Udowodnimy, że f jest bijekcją z $\mathrm{Dom}(f)$ do A. Funkcja f jest różnowartościowa, bo jeśli $n < m$, to $f(m) \notin f(\overline{m})$, ale $f(n) \in f(\overline{m})$, więc $f(n) < f(m)$, w szczególności $f(n) \neq f(m)$.

Korzystając z tej obserwacji łatwo też udowodnimy przez indukcję, że jeśli $f(n)$ jest określone, to $f(n) \geq n$. W kroku indukcyjnym mamy bowiem $f(n) > f(k) \geq k$, dla wszystkich $k < n$.

Z nierówności $n \leq f(n)$ wynika też, że jeśli $\mathrm{Dom}(f) = \mathbb{N}$, to zbiór wartości $\mathrm{Rg}(f) \subseteq A$ nie jest ograniczony. Bo jeśli $a \geq A$, to $f : \mathbb{N} \xrightarrow{1-1} \overline{a}$, co oznacza, że dla $n > a$ zachodzi $f(n) \leq a < n$.

Teraz pokażemy, że f jest na A. Przypuśćmy przeciwnie i niech $a = \min(A - \mathrm{Rg}(f))$. To znaczy, że dla każdego $n \in \mathrm{Dom}(f)$ zachodzi $a \in A - f(\overline{n})$, czyli $a > f(n)$. Jeśli $\mathrm{Dom}(f) = \overline{n}$, to zbiór $A - f(\overline{n})$ jest niepusty i wartość $f(n)$ powinna być określona, a nie jest. Stąd wynika, że $\mathrm{Dom}(f) = \mathbb{N}$. Ale wtedy $\mathrm{Rg}(f) \subseteq \overline{a}$ byłby ograniczony i też mamy sprzeczność.

A zatem f jest bijekcją o zbiorze wartości A i dziedzinie $\mathbb{N}$ lub $\overline{n}$. Wiemy już z powyższego, że dla ograniczonego A musi zachodzić ten drugi przypadek. Pozostaje zauważyć, że jeśli A nie jest ograniczony, to $\mathrm{Dom}(f) = \mathbb{N}$, tj. że $f(n)$ jest zawsze określone, tj. że zbiory $A - f(\overline{n})$ są niepuste. Wystarczy w tym celu wiedzieć, że zbiory $f(\overline{n})$ są ograniczone. To też udowodnimy przez indukcję: zbiór pusty $f(\overline{0})$ jest ograniczony przez zero, a zbiór $f(\overline{s(n)})$ jest ograniczony przez $f(n)$. □

Definicja 8.7 Zbiór B jest *skończony* wtedy i tylko wtedy, gdy jest równoliczny z pewnym ograniczonym podzbiorem $\mathbb{N}$. W przeciwnym razie mówimy, że B jest *nieskończony*.

Z lematu 8.6 wynika, że zbiory skończone są w istocie równoliczne z odcinkami $\overline{n}$ dla $n : \mathbb{N}$.

Lemat 8.8 *Dla dowolnych $n, m \in \mathbb{N}$:*

1) Jeśli istnieje injekcja $f : \overline{m} \xrightarrow{1-1} \overline{n}$, to $m \leq n$.

2) Jeśli istnieje surjekcja $f : \overline{m} \xrightarrow{\text{na}} \overline{n}$, to $m \geq n$.

3) Jeśli $\overline{m} \sim \overline{n}$ to $m = n$.

Dowód: (1) Indukcja ze względu na m. Jeśli $m = 0$, to teza jest oczywista. W kroku indukcyjnym mamy $f : \overline{s(m)} \xrightarrow{1-1} \overline{s(n)}$, czyli $f : \overline{m} \cup \{m\} \xrightarrow{1-1} \overline{n} \cup \{n\}$. Zatem na mocy lematu 8.5 istnieje funkcja $f' : \overline{m} \xrightarrow{1-1} \overline{n}$ i z założenia indukcyjnego wynika $m \leq n$. A stąd już mamy $s(m) \leq s(n)$.

(2) Ta część łatwo wynika z poprzedniej i z twierdzenia 5.11. Ale można ją też udowodnić przez indukcję (bez pomocy pewnika wyboru) z pomocą lematu 8.5 (ćwiczenie).

(3) Wynika natychmiast z części (1) i (2). □

Jeśli $n \in \mathbb{N}$, to piszemy $\overline{\overline{A}} = n$ gdy $A \sim \overline{n}$. Poprawność tego oznaczenia wynika z lematu 8.8, bo może być tylko jedno takie n. Oczywiście mówimy wtedy, że A ma n elementów. Z tego samego lematu wynika też następujący użyteczny fakt:

Twierdzenie 8.9 *Jeśli A jest zbiorem skończonym, oraz $f : A \to A$ to f jest różnowartościowa wtedy i tylko wtedy, gdy jest na A.*

Dowód: Mamy lemat 8.6, więc wystarczy jeśli rozpatrzymy przypadek $A = \overline{n}$, gdzie $n \in \mathbb{N}$.

($\Rightarrow$) Przypuśćmy, że $f : \overline{n} \to \overline{n}$ jest różnowartościowa, ale nie jest „na", tj. istnieje liczba $a \in \overline{n} - \mathrm{Rg}(f)$. To w szczególności oznacza, że $n \neq 0$, czyli $n = s(m)$ dla pewnego m. Wtedy $\overline{n} = (\overline{n} - \{a\}) \cup \{a\} = \overline{m} \cup \{m\}$, więc z lematu 8.5 wynika, że $\overline{n} - \{a\} \sim \overline{m}$. Zatem istnieje funkcja różnowartościowa z $\overline{s(m)}$ do $\overline{m}$, co jest sprzeczne z lematem 8.8(1).

($\Leftarrow$) Przypuśćmy, że $f : \overline{n} \xrightarrow{\text{na}} \overline{n}$ nie jest różnowartościowa. To znaczy, że istnieją takie dwie różne liczby $a, b < n$, że $f(a) = f(b)$. Jasne, że wtedy $n \neq 0$, czyli $n = s(m)$ dla pewnego m. Wówczas $f\restriction_{\overline{n}-\{a\}} : \overline{n} - \{a\} \xrightarrow{\text{na}} \overline{n}$, a ponieważ $\overline{m} \sim \overline{n} - \{a\}$, więc istnieje surjekcja z $\overline{m}$ na $\overline{s(m)}$. Z lematu 8.8(2) wynika teraz sprzeczność. □

Następujące twierdzenie zbiera kilka ważnych własności zbiorów skończonych.

Fakt 8.10

0) Jeśli A jest skończony, to $A \cup \{a\}$ jest skończony.

1) Każdy podzbiór zbioru skończonego jest skończony.

2) Jeśli A jest nieskończony i B jest skończony, to $A - B \neq \varnothing$.

3) Jeśli A jest skończony i $f : A \xrightarrow{\text{na}} B$, to B jest skończony.

4) Suma i iloczyn kartezjański dwóch zbiorów skończonych są skończone.

Dowód: (0) Jeśli $\overline{\overline{A}} = n$ oraz $a \notin A$ to $\overline{\overline{A \cup \{a\}}} = s(n)$.

(1) Jeśli $B \subseteq A$ i A jest skończony, to mamy bijekcję $f : A \xrightarrow[\text{na}]{1-1} \overline{n}$, dla pewnego $n \in \mathbb{N}$. Zbiór B jest więc równoliczny z podzbiorem $f(B) \subseteq \overline{n}$, który oczywiście jest ograniczony. Zatem B też jest skończony.

(2) W przeciwnym razie $A \subseteq B$ i A byłby skończony na mocy części (1).

(3) Z twierdzenia 5.11 wynika, że istnieje wtedy funkcja $g : B \xrightarrow{1-1} A$, a więc B jest równoliczny ze zbiorem $\mathrm{Rg}(g) \subseteq A$, który musi być skończony, jako podzbiór zbioru skończonego.[14]

(4) Ćwiczenie. Wskazówka: przez indukcję należy wykazać, że suma dwóch *rozłącznych* zbiorów, które mają n i m elementów, jest zbiorem o $n+m$ elementach. Podobnie dla iloczynu kartezjańskiego i mnożenia. □

[14] Tę część można też udowodnić bez pomocy twierdzenia 5.11. Wskazówka: zacząć od przypadku $A \subseteq \mathbb{N}$.

9 Zbiory przeliczalne

Fakt 9.1 *Jeśli $n \in \mathbb{N}$, to nie istnieje funkcja $f : \mathbb{N} \xrightarrow{1-1} \overline{n}$. W szczególności zbiór liczb naturalnych $\mathbb{N}$ jest nieskończony.*

Dowód: Gdyby taka funkcja istniała, to $f\restriction_{\overline{s(n)}} : \overline{s(n)} \xrightarrow{1-1} \overline{n}$, a to być nie może z powodu lematu 8.8(1).[15] □

Definicja 9.2 Liczbę kardynalną (moc) zbioru $\mathbb{N}$ oznaczamy symbolem $\aleph_0$ („alef zero"). Zbiory równoliczne z $\mathbb{N}$ nazywamy więc *zbiorami mocy* $\aleph_0$ (por. uwagę 8.2). Mówimy, że zbiór A jest *przeliczalny* wtedy i tylko wtedy, gdy jest skończony lub jest zbiorem mocy $\aleph_0$. W przeciwnym razie zbiór A jest *nieprzeliczalny*.

Moc $\aleph_0$ jest najmniejszą mocą nieskończoną, w następującym sensie:

Twierdzenie 9.3 *Zbiór A jest nieskończony wtedy i tylko wtedy, gdy ma podzbiór mocy $\aleph_0$.*

Dowód: ($\Rightarrow$) Niech ϑ będzie funkcją wyboru dla rodziny $\mathsf{P}(A) - \{\varnothing\}$. Określimy funkcję całkowitą $f : \mathbb{N} \to A$, za pomocą takiej definicji indukcyjnej:

$$f(n) = \vartheta(A - f(\overline{n})) \qquad (*)$$

Poprawność tej definicji nie jest oczywista i wymaga takiej obserwacji: Skoro $\overline{n}$ jest zbiorem skończonym, to $f(\overline{n})$ też jest skończone (na mocy faktu 8.10(3)) a zatem zbiór $A - f(\overline{n})$ jest niepusty (na mocy faktu 8.10(2)) i dlatego prawa strona równania jest określona.

Zauważmy, że funkcja f jest różnowartościowa. W rzeczy samej, jeśli $m \neq n$, to na przykład $m < n$. Wtedy $f(m) \in f(\overline{n})$, a więc wartość $f(n)$, wybierana z dopełnienia zbioru $f(\overline{n})$, musi być różna od $f(m)$. Zatem $f : \mathbb{N} \xrightarrow[\text{na}]{1-1} \mathrm{Rg}(f)$. Zbiór $\mathrm{Rg}(f)$ ma więc moc $\aleph_0$ i jest podzbiorem A.

($\Leftarrow$) Jeżeli $\mathbb{N} \sim B \subseteq A$ i $\overline{\overline{A}} = n \in \mathbb{N}$, to istnieją funkcje $f : \mathbb{N} \xrightarrow[\text{na}]{1-1} B$ i $g : A \xrightarrow[\text{na}]{1-1} \overline{n}$. Stąd $g \circ f : \mathbb{N} \xrightarrow{1-1} \overline{n}$, co jest sprzeczne z faktem 9.1. □

Wniosek 9.4 *Zbiór jest nieskończony wtedy i tylko wtedy, gdy jest równoliczny z pewnym swoim podzbiorem właściwym.*

Dowód: ($\Leftarrow$) Ta część wynika wprost z twierdzenia 8.9.

($\Rightarrow$) Skorzystamy z poprzedniego twierdzenia. Jeśli zbiór A jest nieskończony, to ma podzbiór B o mocy $\aleph_0$. Mamy więc funkcję $f : \mathbb{N} \xrightarrow[\text{na}]{1-1} B$ i możemy określić $g : A \to A$ warunkiem

$$g(x) = \begin{cases} f(f^{-1}(x) + 1), & \text{jeśli } x \in B; \\ x, & \text{w przeciwnym przypadku.} \end{cases}$$

Łatwo zauważyć, że funkcja g jest różnowartościowa. Ale ta funkcja nie jest na A, bo element $f(0)$ nie należy do $\mathrm{Rg}(g)$. Zatem $A \sim \mathrm{Rg}(g) \subsetneq A$. □

[15] Przypomnijmy, że symbol $\restriction$ oznacza obcięcie funkcji.

Fakt 9.5 *Każdy podzbiór zbioru przeliczalnego jest przeliczalny.*

Dowód: Wynika łatwo z lematu 8.6. □

Następujący fakt uzasadnia nazwę „zbiór przeliczalny". Zbiór jest przeliczalny, gdy jego elementy można *przeliczać* (niekoniecznie *przeliczyć*), tj. ustawić je w ciąg nieskończony.

Wniosek 9.6 *Niepusty zbiór A jest przeliczalny wtedy i tylko wtedy, gdy istnieje surjekcja $f : \mathbb{N} \xrightarrow{\text{na}} A$.*

Dowód: ($\Rightarrow$) Jeśli $\overline{\overline{A}} = \aleph_0$ to taka funkcja istnieje z definicji i nawet jest różnowartościowa. Jeśli $\overline{\overline{A}} = n \in \mathbb{N}$, to $n \neq 0$ i mamy funkcję $g : \overline{n} \xrightarrow[\text{na}]{1-1} A$, którą można poprawić tak:

$$h(m) = \begin{cases} g(m), & \text{jeśli } m < n; \\ g(0), & \text{w przeciwnym przypadku.} \end{cases}$$

($\Leftarrow$) Niech $f : \mathbb{N} \xrightarrow{\text{na}} A$ i niech $g = \lambda a{:}A.\ \min\{i \in \mathbb{N} \mid f(i) = a\}$. Wtedy $g : A \xrightarrow{1-1} \mathbb{N}$, zbiór A jest więc równoliczny z podzbiorem $\mathrm{Rg}(g)$ zbioru $\mathbb{N}$, a zatem przeliczalny. □

Lemat 9.7 *Jeśli zbiór A jest przeliczalny i $f : A \xrightarrow{\text{na}} B$, to B jest przeliczalny.*

Dowód: Na mocy wniosku 9.6 istnieje surjekcja $g : \mathbb{N} \xrightarrow{\text{na}} A$. Wtedy $f \circ g : \mathbb{N} \xrightarrow{\text{na}} B$, więc na mocy tego samego wniosku zbiór B jest przeliczalny. □

Fakt 9.8 *Jeśli zbiory A i B są przeliczalne to $A \cup B$ i $A \times B$ też są przeliczalne.*

Dowód: Jeśli któryś ze zbiorów A i B jest pusty, to teza jest oczywista. Załóżmy więc, że A i B są niepuste. Na mocy wniosku 9.6 istnieją więc funkcje $f : \mathbb{N} \xrightarrow{\text{na}} A$ i $g : \mathbb{N} \xrightarrow{\text{na}} B$. Możemy teraz określić funkcję $\varphi : \mathbb{N} \xrightarrow{\text{na}} A \cup B$ wzorem

$$\varphi(n) = \begin{cases} f(k), & \text{jeśli } n = 2k, \text{ dla pewnego } k; \\ g(k), & \text{jeśli } n = 2k+1, \text{ dla pewnego } k \end{cases}$$

A zatem $A \cup B$ jest zbiorem przeliczalnym, co też wynika z wniosku 9.6. Aby określić funkcję $\psi : \mathbb{N} \xrightarrow{\text{na}} A \times B$, skorzystamy z jednoznaczności rozkładu liczb naturalnych na czynniki pierwsze. Każdą liczbę $n \neq 0$ możemy jednoznacznie zapisać w postaci

$$n = 2^i 3^j q,$$

gdzie q nie jest podzielne ani przez 2 ani przez 3. Przyjmujemy

$$\psi(n) = \begin{cases} \langle f(0), g(0) \rangle, & \text{jeśli } n = 0; \\ \langle f(i), g(j) \rangle, & \text{jeśli } n = 2^i 3^j q \text{ oraz } q \text{ nie dzieli się przez 2 ani 3} \end{cases}$$

Funkcja ψ jest „na", bo dla dowolnych $a \in A$, $b \in B$ istnieją takie liczby i, j, że $f(i) = a$ i $g(j) = b$. A więc $\langle a, b \rangle = \psi(2^i 3^j)$. □

Przykład 9.9 Funkcja $t : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ dana wzorem $t(n, m) = 2^n 3^m$ jest różnowartościowa. Natomiast następujące funkcje $u, v : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ są nawet bijekcjami.[16]

$$u(m, n) = 2^m(2n + 1) - 1;$$
$$v(m, n) = \frac{(m + n)(m + n + 1)}{2} + m.$$

Sprawdzenie, że tak jest w istocie, pozostawiamy jako ćwiczenie. Wskazówka: pierwszy składnik w definicji $v(m, n)$ przedstawia sumę liczb naturalnych od zera do $m + n$.

Przykłady zbiorów przeliczalnych

- Zbiór $\mathbb{N} \times \mathbb{N}$ jest przeliczalny (jak widać z powyższego).
- Zbiór $\mathbb{Z}$ wszystkich liczb całkowitych jest przeliczalny. Skoro bowiem $\mathbb{Z} = \mathbb{N} \times \mathbb{N}/_\sim$ to mamy funkcję $\kappa : \mathbb{N} \times \mathbb{N} \xrightarrow{\text{na}} \mathbb{Z}$ określoną warunkiem $\kappa(m, n) = [\langle m, n\rangle]_\sim$. (Mówiąc po ludzku, chodzi o funkcję $\kappa(m, n) = m - n$. Każda liczba całkowita jest różnicą dwóch liczb naturalnych.)
- Zbiór $\mathbb{Q}$ wszystkich liczb wymiernych jest przeliczalny z analogicznych powodów.
- A więc przeliczalny jest też np. zbiór wszystkich punktów płaszczyzny o współrzędnych wymiernych. Utożsamiamy go przecież ze zbiorem $\mathbb{Q} \times \mathbb{Q}$.

Twierdzenie 9.10 *Suma przeliczalnej rodziny zbiorów przeliczalnych jest przeliczalna.*

Dowód: Niech $\mathcal{A}$ będzie przeliczalną rodziną zbiorów przeliczalnych. Bez straty ogólności możemy założyć, że:

- $\mathcal{A} \neq \varnothing$, bo inaczej $\bigcup \mathcal{A} = \varnothing$, czyli teza jest oczywista;
- $\varnothing \notin \mathcal{A}$, bo $\bigcup \mathcal{A} = \bigcup(\mathcal{A} - \{\varnothing\})$, więc zamiast $\mathcal{A}$ możemy wziąć $\mathcal{A} - \{\varnothing\}$.

A więc, na mocy wniosku 9.6 mamy funkcję:

$$F : \mathbb{N} \xrightarrow{\text{na}} \mathcal{A},$$

a ponieważ elementy $\mathcal{A}$ są też przeliczalne, więc dla dowolnego $m \in \mathbb{N}$ jest też funkcja

$$f_m : \mathbb{N} \xrightarrow{\text{na}} F(m).$$

Wtedy $G : \mathbb{N} \times \mathbb{N} \xrightarrow{\text{na}} \bigcup \mathcal{A}$, gdzie $G(m, n) = f_m(n)$. Sprawdźmy, że funkcja G jest faktycznie „na". Ponieważ F jest „na", więc każdy element $a \in \bigcup \mathcal{A}$ należy do pewnego $F(m)$. Zatem a jest postaci $f_m(n)$, bo f_m też jest „na". Wnioskujemy, że $\bigcup \mathcal{A}$ jest zbiorem przeliczalnym, jako obraz zbioru przeliczalnego (lemat 9.7). □

Uwaga: * Choć nie widać tego na pierwszy rzut oka, dowód powyższego twierdzenia w istotny sposób opiera się na pewniku wyboru. Przypisujemy bowiem każdej liczbie m pewną funkcję $f_m : \mathbb{N} \xrightarrow{\text{na}} F(m)$, a więc *implicite* stosujemy funkcję wyboru dla rodziny $\mathcal{A}$. Ściślej, powołujemy się tu na twierdzenie 5.10 o niepustości produktu zbiorów niepustych.

[16] Czasami o bijekcji z $\mathbb{N} \times \mathbb{N}$ na $\mathbb{N}$ mówimy *funkcja pary*. Taka funkcja pozwala na zakodowanie dwóch liczb naturalnych za pomocą jednej.

Wniosek 9.11 *Jeśli alfabet A jest przeliczalny to zbiór wszystkich słów A^* też jest przeliczalny.*

Dowód: Niech A^n oznacza zbiór wszystkich słów nad A długości n. Nietrudno pokazać przez indukcję, że każdy ze zbiorów A^n jest przeliczalny. Istotnie, zbiór $A^0 = \{\varepsilon\}$ jest jednoelementowy, a krok indukcyjny wynika z łatwej równoliczności $A^{n+1} \sim A^n \times A$. Skoro A^* jest sumą wszystkich A^n, dla $n \in \mathbb{N}$, to teza wynika z twierdzenia 9.10. □

Uwaga: Analogiczny fakt jest prawdziwy dla takich typów indukcyjnych jak skończone drzewa binarne, listy itp.

Liczby algebraiczne to pierwiastki rzeczywiste wielomianów o współczynnikach wymiernych. Pozostałe liczby (takie jak e i π) nazywamy *przestępnymi*.

Wniosek 9.12 *Zbiór wszystkich liczb algebraicznych jest przeliczalny.*

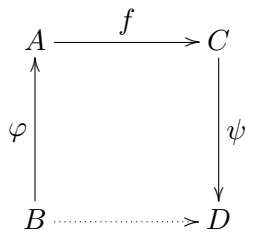
Dowód: Wielomian jest wyznaczony przez skończony ciąg swoich współczynników. Zbiór wielomianów $\mathbb{Q}[x]$ jest więc równoliczny z $\mathbb{Q}^*$ i też przeliczalny.

Wielomian ma skończenie wiele pierwiastków, więc zbiór liczb algebraicznych to przeliczalna suma zbiorów skończonych. □

10 Teoria mocy

Lemat 10.1 *Jeżeli $A \sim B$ i $C \sim D$ oraz istnieje injekcja $f : A \xrightarrow{1-1} C$, to istnieje też injekcja $g : B \xrightarrow{1-1} D$.*

Dowód: Istnieją bijekcje $\varphi : B \xrightarrow[\text{na}]{1-1} A$ oraz $\psi : C \xrightarrow[\text{na}]{1-1} D$. Zatem $\psi \circ f \circ \varphi : B \xrightarrow{1-1} D$. Tę konstrukcję przedstawia poniższy diagram:



□

Definicja 10.2 Mówimy, że moc zbioru A jest *mniejsza lub równa* mocy zbioru B (i piszemy $\overline{\overline{A}} \leq \overline{\overline{B}}$), wtedy i tylko wtedy, gdy istnieje injekcja $f : A \xrightarrow{1-1} B$. Jeżeli $\overline{\overline{A}} \leq \overline{\overline{B}}$ ale zbiory A i B nie są równoliczne, to piszemy $\overline{\overline{A}} < \overline{\overline{B}}$ i mówimy, że zbiór A jest mocy *mniejszej* niż zbiór B.

Uwaga:

- Poprawność powyższej definicji wynika z lematu 10.1.
- Jeśli $\mathfrak{m}, \mathfrak{n}$ są liczbami kardynalnymi to $\mathfrak{m} \leq \mathfrak{n}$ oznacza, że $\overline{\overline{A}} \leq \overline{\overline{B}}$, dla $\overline{\overline{A}} = \mathfrak{m}$, $\overline{\overline{B}} = \mathfrak{n}$.
- Jeśli $f : A \xrightarrow{1-1} B$, ale f nie jest bijekcją, to nie znaczy, że $\overline{\overline{A}} < \overline{\overline{B}}$. Na przykład funkcja następnika jest injekcją z $\mathbb{N}$ w $\mathbb{N}$ i nie jest „na”, ale przecież $\overline{\overline{\mathbb{N}}} \not< \overline{\overline{\mathbb{N}}}$.

Przykład 10.3

- Jeśli $A \subseteq B$, to $\overline{\overline{A}} \leq \overline{\overline{B}}$.
- Dla dowolnej liczby naturalnej n zachodzi $n < \aleph_0$.
- Dla dowolnego zbioru A zachodzi $\overline{\overline{A}} \leq \overline{\overline{\mathsf{P}(A)}}$. Istotnie, mamy $\zeta : A \xrightarrow{1-1} \mathsf{P}(A)$, gdzie $\zeta(a) = \{a\}$ dla $a \in A$.
- Zbiór A jest nieskończony wtedy i tylko wtedy, gdy $\aleph_0 \leq \overline{\overline{A}}$ (twierdzenie 9.3).

Fakt 10.4 *Dla dowolnych niepustych zbiorów A, B następujące warunki są równoważne:*

1) $\overline{\overline{A}} \leq \overline{\overline{B}}$;
2) Istnieje $g : B \xrightarrow{\text{na}} A$;
3) Zbiór A jest równoliczny z pewnym podzbiorem zbioru B.

Dowód: Równoważność warunków (1) i (2) to dokładnie treść twierdzenia 5.12. Równoważność (1) i (3) wynika z następujących obserwacji:

- Jeśli $f : A \xrightarrow{1-1} B$ to $A \sim \mathrm{Rg}(f)$.
- Jeśli $f : A \xrightarrow[\text{na}]{1-1} C \subseteq B$ to $f : A \xrightarrow{1-1} B$. □

Fakt 10.5 *Dla dowolnych zbiorów A, B, C:*

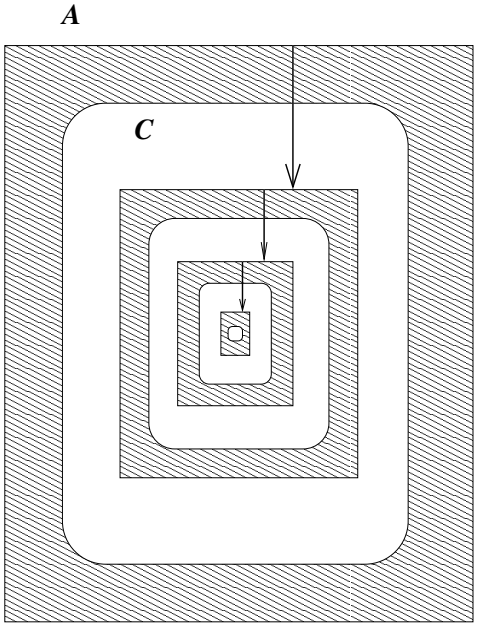
- $\overline{\overline{A}} \leq \overline{\overline{A}}$;
- *Jeśli* $\overline{\overline{A}} \leq \overline{\overline{B}}$ *i* $\overline{\overline{B}} \leq \overline{\overline{C}}$ *to* $\overline{\overline{A}} \leq \overline{\overline{C}}$.

O ile powyższy fakt jest całkiem oczywisty, to antysymetria nierówności

$$\textit{Jeśli } \overline{\overline{A}} \leq \overline{\overline{B}} \quad i \quad \overline{\overline{B}} \leq \overline{\overline{A}} \quad to \quad \overline{\overline{A}} = \overline{\overline{B}}$$

(zwana twierdzeniem Cantora-Bernsteina) nie jest już oczywista. Udowodnimy ją najpierw w takiej wersji:

Lemat 10.6 *Jeśli $\varphi : A \xrightarrow{1-1} C \subseteq A$ to $C \sim A$.*



Rysunek 1: Dowód lematu 10.6

Dowód: Zaczniemy od określenia ciągu zbiorów X_n, dla $n \in \mathbb{N}$.

$$\begin{aligned} X_0 &= A - C; \\ X_{n+1} &= \varphi(X_n). \end{aligned}$$

Niech $X = \bigcup\{X_n \mid n \in \mathbb{N}\}$ i niech $Y = A - X$. Zauważmy, że $C = A - X_0 = (X \cup Y) - X_0 = (X - X_0) \cup Y$, bo $Y \cap X_0 = \varnothing$. Określimy bijekcję $\psi : A \xrightarrow[\text{na}]{1-1} C$ jak następuje:

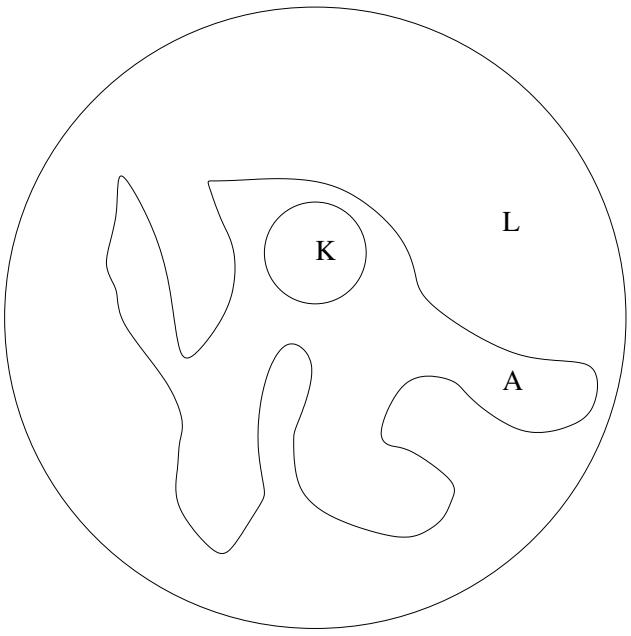
$$\psi(x) = \begin{cases} x, & \text{jeśli } x \in Y; \\ \varphi(x), & \text{jeśli } x \in X \end{cases}$$

Inaczej, $\psi = \varphi\restriction_X \cup \mathrm{id}_Y$. Na rysunku 1 zbiór X odpowiada obszarowi zakreskowanemu, a zbiór Y to cała reszta. Poszczególne zakreskowane składowe to zbiory X_n. A zatem funkcja ψ jest identycznością na obszarze białym, a każdą z zakreskowanych składowych przekształca w następną. Funkcja ψ jest różnowartościowa, ponieważ $\mathrm{id}_Y : Y \xrightarrow{1-1} Y$ oraz $\varphi\restriction_X : X \xrightarrow{1-1} X$ są są funkcjami różnowartościowymi, a przy tym X i Y są rozłączne. Ponadto ψ jest na C. Jeśli bowiem $c \in C$, to są dwie możliwości. Albo $c \in Y$ i wtedy $c = \psi(c)$, albo $c \in X - X_0$ i mamy $c \in X_{n+1}$ dla pewnego n. A wtedy $c = \varphi(x) = \psi(x)$ dla pewnego $x \in X_n$. □

Twierdzenie 10.7 (Cantora-Bernsteina) *Jeśli* $\overline{\overline{A}} \leq \overline{\overline{B}}$ *i* $\overline{\overline{B}} \leq \overline{\overline{A}}$ *to* $\overline{\overline{A}} = \overline{\overline{B}}$.

Dowód: Z założenia istnieją funkcje $f : A \xrightarrow{1-1} B$ i $g : B \xrightarrow{1-1} A$. Zbiór $C = \mathrm{Rg}(g)$ jest oczywiście równoliczny z B. Jeśli teraz $\varphi = g \circ f$ to $\varphi : A \xrightarrow{1-1} C$. Na mocy lematu 10.6, zbiór A jest równoliczny z C, a więc także z B. □

Twierdzenie Cantora-Bernsteina jest niezwykle użytecznym narzędziem do badania mocy zbiorów. Zwykle znacznie łatwiej jest wskazać dwie funkcje różnowartościowe, jedną z A do B i drugą z B do A, niż bijekcję pomiędzy A i B. Na przykład moc dziwnej figury A na rysunku 2 jest taka sama jak moc każdego z dwóch kół (otwartych). Mamy bowiem $\overline{\overline{\mathrm{K}}} \leq \overline{\overline{\mathrm{A}}} \leq \overline{\overline{\mathrm{L}}}$, bo $\mathrm{K} \subseteq \mathrm{A} \subseteq \mathrm{L}$. Ponieważ łatwo zauważyć, że dowolne dwa koła otwarte są równoliczne, więc $\overline{\overline{\mathrm{K}}} = \overline{\overline{\mathrm{L}}}$, i możemy użyć twierdzenia Cantora-Bernsteina.



Rysunek 2: Zastosowanie twierdzenia Cantora-Bernsteina

Twierdzenie 10.8 (Cantora) *Dla dowolnego zbioru* A *zachodzi* $\overline{\overline{A}} < \overline{\overline{\mathsf{P}(A)}}$.

Dowód: Już poprzednio zauważyliśmy, że $\overline{\overline{A}} \leq \overline{\overline{\mathsf{P}(A)}}$, należy więc pokazać, że nie istnieje bijekcja $F : A \xrightarrow[\text{na}]{1-1} \mathsf{P}(A)$. Przypuśćmy, że taka jest, i niech $B = \{x \in A \mid x \notin F(x)\}$. Skoro F jest „na", to istnieje takie $b \in A$, że $F(b) = B$. Pytamy, czy $b \in B$. Jeśli $b \in B$, to z definicji zbioru B mamy $b \notin F(b) = B$. Ale jeśli $b \notin B$, to też źle, bo wtedy warunek $b \notin F(b)$ nie powinien zachodzić, czyli mielibyśmy właśnie $b \in B$. Otrzymana sprzeczność wynikła z założenia, że $F : A \xrightarrow[\text{na}]{1-1} \mathsf{P}(A)$, a więc takiej funkcji nie ma. □

Rozumowanie użyte w dowodzie twierdzenia Cantora stosuje tzw. *metodę przekątniową* (rozważamy dwuargumentowy predykat „$x \notin F(y)$" dla $x = y$, czyli wybieramy z niego „przekątną"). Do sprzeczności doprowadziło nas zjawisko podobne do *paradoksu kłamcy*,[17] znane też z anegdoty o wojskowym fryzjerze, któremu polecono golić tych i tylko tych żołnierzy, co sami się nie golą. Porównajmy dwa niemożliwe do spełnienia warunki:

$$\forall x \in A(x \in F(b) \Leftrightarrow x \notin F(x));$$

$$\forall x(b \text{ goli } x \Leftrightarrow x \text{ nie goli } x).$$

a zobaczymy, że chodzi tu o to samo zjawisko.

Wniosek 10.9

1. *Istnieją zbiory nieprzeliczalne, na przykład* $\mathsf{P}(\mathbb{N})$.
2. *Istnieje nieskończenie wiele nieskończonych liczb kardynalnych.*

Dowód: (1) Natychmiastowy wniosek z twierdzenia Cantora.

(2) Łatwo widzieć, że $\overline{\overline{\mathbb{N}}} < \overline{\overline{\mathsf{P}(\mathbb{N})}} < \overline{\overline{\mathsf{P}(\mathsf{P}(\mathbb{N}))}} < \overline{\overline{\mathsf{P}(\mathsf{P}(\mathsf{P}(\mathbb{N})))}} < \cdots$ □

Uwaga 10.10 Przypisując inne typy zbiorom A i $\mathsf{P}(A)$ wykluczamy tę możliwość, że jakiś podzbiór B zbioru A jest jednocześnie jego elementem. W „beztypowej" teorii mnogości takie zjawiska są dopuszczalne, ale i tam:

Nie istnieje zbiór wszystkich zbiorów, tj. zbiór Ω *spełniający warunek* $\forall x(x \in \Omega)$.

Istotnie, gdyby Ω był zbiorem wszystkich zbiorów, to także każdy podzbiór Ω byłby jego elementem, mielibyśmy więc $\mathsf{P}(\Omega) \subseteq \Omega$, skąd $\overline{\overline{\mathsf{P}(\Omega)}} \leq \overline{\overline{\Omega}}$.

Liczby rzeczywiste

Definicja 10.11 Moc zbioru wszystkich liczb rzeczywistych nazywamy *continuum* i oznaczamy przez $\mathfrak{c}$.

Przypomnijmy, że $\overline{2}^{\mathbb{N}}$ to zbiór wszystkich funkcji z $\mathbb{N}$ do $\overline{2} = \{0, 1\}$, inaczej – zbiór wszystkich nieskończonych ciągów zerojedynkowych. Zamiast $\overline{2}^{\mathbb{N}}$ zwykle piszemy po prostu $2^{\mathbb{N}}$.

[17] Stwierdzenie *„To zdanie jest fałszywe"* nie może być ani prawdziwe ani fałszywe.

Fakt 10.12 $\mathfrak{C} = \overline{\overline{\mathsf{P}(\mathbb{N})}} = \overline{\overline{2^{\mathbb{N}}}}$.

Dowód: Najpierw zauważmy, że $F : 2^{\mathbb{N}} \xrightarrow[\text{na}]{1-1} \mathsf{P}(\mathbb{N})$, gdzie $F(f) = f^{-1}(\{1\})$. Istotnie, dla $f \neq g$ istnieje jakieś n, dla którego $f(n) = 1$ i $g(n) = 0$ albo na odwrót. Zatem $n \in F(f)$ i $n \notin F(g)$ albo na odwrót, funkcja F jest więc różnowartościowa. Jest też na $\mathsf{P}(\mathbb{N})$, bo jeśli $B \subseteq \mathbb{N}$ to $B = F(\chi_B)$, gdzie χ_B to *funkcja charakterystyczna* zbioru B, czyli:

$$\chi_B(n) = \begin{cases} 1, & \text{jeśli } n \in B; \\ 0, & \text{w przeciwnym przypadku.} \end{cases}$$

A zatem zbiory $\mathsf{P}(\mathbb{N})$ i $2^{\mathbb{N}}$ są tej samej mocy. Aby pokazać, że jest to moc continuum, skorzystamy z twierdzenia 10.7, tj. udowodnimy dwie nierówności: $\overline{\overline{2^{\mathbb{N}}}} \leq \mathfrak{C}$ i $\mathfrak{C} \leq \overline{\overline{\mathsf{P}(\mathbb{N})}}$.

$[\overline{\overline{2^{\mathbb{N}}}} \leq \mathfrak{C}]$ Niech $H : 2^{\mathbb{N}} \to \mathbb{R}$ przyporządkowuje każdemu ciągowi zerojedynkowemu liczbę rzeczywistą z przedziału $(0, 1)$, której zapis *dziesiętny* po przecinku odpowiada temu ciągowi. A więc na przykład $H(0110001110\ldots) = 0,0110001110\ldots$ Dokładniej, dla dowolnego $f \in 2^{\mathbb{N}}$

$$H(f) = \sum_{i=0}^{\infty} \frac{f(i)}{10^{i+1}}$$

Aby sprawdzić, że funkcja H jest różnowartościowa, przypuśćmy, że $f \neq g$. Niech teraz $n = \min\{i \mid f(i) \neq g(i)\}$. Wtedy $\sum_{i<n} \frac{f(i)}{10^{i+1}} = \sum_{i<n} \frac{g(i)}{10^{i+1}}$. Oznaczmy tę sumę przez b i przypuśćmy na przykład, że $f(n) = 0$ i $g(n) = 1$. Wtedy

$$H(f) = b + \sum_{i=n+1}^{\infty} \frac{f(i)}{10^{i+1}} < b + \frac{1}{10^{n+1}} \leq H(g).$$

$[\mathfrak{C} \leq \overline{\overline{\mathsf{P}(\mathbb{N})}}]$ Niech $\alpha : \mathbb{N} \xrightarrow[\text{na}]{1-1} \mathbb{Q}$ będzie dowolną ustaloną bijekcją, i niech

$$G(x) = \{n \in \mathbb{N} \mid \alpha(n) < x\},$$

dla dowolnego $x \in \mathbb{R}$. W ten sposób określiliśmy funkcję $G : \mathbb{R} \to \mathsf{P}(\mathbb{N})$. Ta funkcja jest różnowartościowa, bo jeśli $x \neq y$ to na przykład $x < y$, a wtedy istnieje liczba wymierna q spełniająca nierówności $x < q < y$. Mamy więc $\alpha^{-1}(q) \in G(y) - G(x)$. □

11 Arytmetyka liczb kardynalnych

Lemat 11.1 *Niech* $\overline{\overline{A}} \leq \overline{\overline{B}}$ *i* $\overline{\overline{C}} \leq \overline{\overline{D}}$. *Wtedy:*

1) $\overline{\overline{A \oplus C}} \leq \overline{\overline{B \oplus D}}$;

2) $\overline{\overline{A \times C}} \leq \overline{\overline{B \times D}}$;

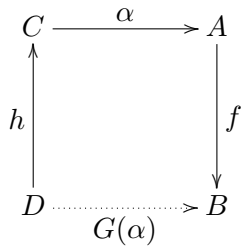
3) Jeśli $C \neq \varnothing$, *to* $\overline{\overline{A^C}} \leq \overline{\overline{B^D}}$.

Dowód: Istnieją funkcje $f : A \xrightarrow{1-1} B$ i $g : C \xrightarrow{1-1} D$.

(1) Wówczas oczywiście $f \oplus g : A \oplus C \xrightarrow{1-1} B \oplus D$.

(2) Funkcja $F : A \times C \xrightarrow{1-1} B \times D$ może być określona warunkiem $F(a, c) = \langle f(a), g(c)\rangle$. Różnowartościowość F łatwo wynika z różnowartościowości f i g.

(3) Ponieważ $C \neq \varnothing$, więc istnieje funkcja $h : D \xrightarrow{\text{na}} C$. Funkcję $G : A^C \to B^D$ określimy równaniem $G(\alpha) = f \circ \alpha \circ h$. Rysunek poniżej objaśnia tę definicję:



Sprawdźmy, że funkcja G jest różnowartościowa. Jeśli $\alpha, \beta \in A^C$ oraz $\alpha \neq \beta$, to $\alpha(c) \neq \beta(c)$, dla pewnego $c \in C$. Funkcja h jest „na", więc istnieje takie $d \in D$, że $h(d) = c$. Z różnowartościowości funkcji f wnioskujemy, że

$$G(\alpha)(d) = f(\alpha(h(d))) = f(\alpha(c)) \neq f(\beta(c)) = f(\beta(h(d))) = G(\beta)(d),$$

czyli, że $G(\alpha) \neq G(\beta)$. □

Wniosek 11.2 *Jeśli* $\overline{\overline{A}} = \overline{\overline{B}}$ *i* $\overline{\overline{C}} = \overline{\overline{D}}$ *to*

- $\overline{\overline{A \oplus C}} = \overline{\overline{B \oplus D}}$;
- $\overline{\overline{A \times C}} = \overline{\overline{B \times D}}$;
- $\overline{\overline{A^C}} = \overline{\overline{B^D}}$.

Dowód: Łatwa konsekwencja lematu 11.1. Uwaga: należy zauważyć, że $\overline{\overline{A^\varnothing}} = 1$ dla dowolnego zbioru A. □

Z wniosku 11.2 wynika poprawność następującej definicji:

Definicja 11.3

Sumą $\mathfrak{m}+\mathfrak{n}$ liczb kardynalnych $\mathfrak{m}$ i $\mathfrak{n}$ nazywamy moc dowolnego zbioru postaci $A \oplus C$, gdzie $\overline{\overline{A}} = \mathfrak{m}$ i $\overline{\overline{C}} = \mathfrak{n}$. (Inaczej, $\mathfrak{m}+\mathfrak{n}$ to moc zbioru $A \cup C$, gdzie $\overline{\overline{A}} = \mathfrak{m}$, $\overline{\overline{C}} = \mathfrak{n}$, oraz $A \cap C = \varnothing$.)

Iloczynem $\mathfrak{m} \cdot \mathfrak{n}$ liczb kardynalnych $\mathfrak{m}$ i $\mathfrak{n}$ nazywamy moc dowolnego zbioru postaci $A \times C$, gdzie $\overline{\overline{A}} = \mathfrak{m}$, $\overline{\overline{C}} = \mathfrak{n}$.

Potęgą $\mathfrak{m}^{\mathfrak{n}}$ o podstawie $\mathfrak{m}$ i wykładniku $\mathfrak{n}$ nazywamy moc dowolnego zbioru postaci A^C, gdzie $\overline{\overline{A}} = \mathfrak{m}$, $\overline{\overline{C}} = \mathfrak{n}$.

Uwaga: Zwykłe działania na liczbach naturalnych pokrywają się z działaniami określonymi powyżej.

Przykład 11.4

- $\aleph_0 + \aleph_0 = \aleph_0$, bo $\mathbb{Z} \sim \mathbb{N}$.
- $\aleph_0 \cdot \aleph_0 = \aleph_0$, bo $\mathbb{N} \times \mathbb{N} \sim \mathbb{N}$.
- $2^{\aleph_0} = \mathfrak{c}$, na mocy faktu 10.12.
- Przyjmijmy $\beth_0 = \aleph_0$ i dalej $\beth_{n+1} = 2^{\beth_n}$. (Hebrajską literę $\beth$ czytamy „bet”.) Wtedy $\overline{\overline{\mathsf{P}(\mathbb{N})}} = \overline{\overline{\mathbb{R}}} = \beth_1$, $\overline{\overline{\mathsf{P}(\mathsf{P}(\mathbb{N}))}} = \beth_2$, itd.

Fakt 11.5 *Jeśli* $\mathfrak{m} \geq \aleph_0$ *to* $\mathfrak{m} + \aleph_0 = \mathfrak{m}$.

Dowód: Niech $\overline{\overline{A}} = \mathfrak{m}$ i $\overline{\overline{C}} = \aleph_0$, a przy tym $A \cap C = \varnothing$. Na mocy twierdzenia 9.3, istnieje podzbiór $B \subseteq A$, o mocy $\aleph_0$. Wtedy $A \cup C = (A - B) \cup (B \cup C) \sim (A - B) \cup B = A$, ponieważ $B \cup C$ też jest mocy $\aleph_0$. A zatem $\mathfrak{m} + \aleph_0 = \overline{\overline{A \cup C}} = \overline{\overline{A}} = \mathfrak{m}$. □

Wiele praw arytmetyki liczb naturalnych można uogólnić na dowolne liczby kardynalne. W szczególności, dla dowolnych liczb kardynalnych $\mathfrak{m}$, $\mathfrak{n}$ i $\mathfrak{p}$, zachodzą następujące równości:

- $\mathfrak{m} + 0 = \mathfrak{m}$ (bo $A \oplus \varnothing \sim A$).
- $\mathfrak{m} + \mathfrak{n} = \mathfrak{n} + \mathfrak{m}$ (bo $A \oplus B \sim B \oplus A$).
- $(\mathfrak{m} + \mathfrak{n}) + \mathfrak{p} = \mathfrak{m} + (\mathfrak{n} + \mathfrak{p})$ (bo $(A \oplus B) \oplus C \sim A \oplus (B \oplus C)$).
- $\mathfrak{m} \cdot 1 = \mathfrak{m}$ (bo $A \times 1 \sim A$).
- $\mathfrak{m} \cdot 0 = 0$ (bo $A \times \varnothing = \varnothing$).
- $\mathfrak{m} \cdot \mathfrak{n} = \mathfrak{n} \cdot \mathfrak{m}$ (bo $A \times B \sim B \times A$).
- $(\mathfrak{m} \cdot \mathfrak{n}) \cdot \mathfrak{p} = \mathfrak{m} \cdot (\mathfrak{n} \cdot \mathfrak{p})$ (bo $(A \times B) \times C \sim A \times (B \times C)$).
- $\mathfrak{m} \cdot (\mathfrak{n} + \mathfrak{p}) = \mathfrak{m} \cdot \mathfrak{n} + \mathfrak{m} \cdot \mathfrak{p}$ (bo $A \times (B \oplus C) \sim (A \times B) \oplus (A \times C)$).
- $\mathfrak{m}^0 = 1$ (bo tylko $\varnothing$ należy do $A^{\varnothing}$).

- $\mathfrak{m}^1 = \mathfrak{m}$ (bo elementy A^1 to funkcje stałe).
- $1^{\mathfrak{m}} = 1$ (bo funkcja należąca do $\{0\}^A$ musi być stale równa zero).
- $0^{\mathfrak{m}} = 0$, o ile $\mathfrak{m} \neq 0$ (bo nie ma funkcji ze zbioru niepustego w pusty).

Mniej oczywiste są trzy prawa potęgowania.

Fakt 11.6 *Dla dowolnych liczb kardynalnych* $\mathfrak{m}$, $\mathfrak{n}$ *i* $\mathfrak{p}$, *zachodzą następujące równości:*

1) $\mathfrak{m}^{\mathfrak{n}} \cdot \mathfrak{m}^{\mathfrak{p}} = \mathfrak{m}^{(\mathfrak{n}+\mathfrak{p})}$;

2) $\mathfrak{m}^{\mathfrak{n}} \cdot \mathfrak{p}^{\mathfrak{n}} = (\mathfrak{m} \cdot \mathfrak{p})^{\mathfrak{n}}$;

3) $(\mathfrak{m}^{\mathfrak{n}})^{\mathfrak{p}} = \mathfrak{m}^{\mathfrak{n}\cdot\mathfrak{p}}$.

Dowód: W części (1) należy pokazać, że $A^B \times A^C \sim A^{B\oplus C}$. W tym celu można określić bijekcję $F : A^B \times A^C \xrightarrow[\text{na}]{1-1} A^{B\oplus C}$ wzorem $F(f,g) = f \oplus g$, dla $f : B \to A$ i $g : C \to A$.

Dla dowodu części (2) potrzebna jest bijekcja $G : A^B \times C^B \xrightarrow[\text{na}]{1-1} (A \times C)^B$, którą zdefiniujemy tak: $G(f,g)(b) = \langle f(b), g(b)\rangle$, dla $f : B \to A$, $g : B \to C$ i $b \in B$.

W części (3) posłużymy się bijekcją $H : (A^B)^C \xrightarrow[\text{na}]{1-1} A^{B\times C}$, która jest określona wzorem $H(\varphi)(b,c) = \varphi(c)(b)$, dla $\varphi : C \to A^B$ i dla $c \in C$, $b \in B$. Dowód, że jest to istotnie bijekcja, podobnie jak funkcje określone w (1) i (2) pozostawiamy jako ćwiczenie. □

Lemat 11.1 stwierdza, że działania na liczbach kardynalnych są operacjami monotonicznymi w następującym sensie. Jeśli $\mathfrak{m} \leq \mathfrak{n}$ i $\mathfrak{p} \leq \mathfrak{q}$ (na przykład $\mathfrak{p} = \mathfrak{q}$), to:

- $\mathfrak{m} + \mathfrak{p} \leq \mathfrak{n} + \mathfrak{q}$;
- $\mathfrak{m} \cdot \mathfrak{p} \leq \mathfrak{n} \cdot \mathfrak{q}$;
- $\mathfrak{m}^{\mathfrak{p}} \leq \mathfrak{n}^{\mathfrak{q}}$, pod warunkiem, że $\mathfrak{p} \neq 0$.

Wniosek 11.7

1) $\aleph_0 \cdot \mathfrak{c} = \mathfrak{c} \cdot \mathfrak{c} = \mathfrak{c}$.

2) $\aleph_0^{\aleph_0} = \mathfrak{c}^{\aleph_0} = \mathfrak{c}$.

3) $2^{\mathfrak{c}} = \aleph_0^{\mathfrak{c}} = \mathfrak{c}^{\mathfrak{c}}$.

Dowód:

1) Bo $\mathfrak{c} \leq \aleph_0 \cdot \mathfrak{c} \leq \mathfrak{c} \cdot \mathfrak{c} = 2^{\aleph_0} \cdot 2^{\aleph_0} = 2^{\aleph_0+\aleph_0} = 2^{\aleph_0} = \mathfrak{c}$.

2) Bo $\mathfrak{c} = 2^{\aleph_0} \leq \aleph_0^{\aleph_0} \leq \mathfrak{c}^{\aleph_0} = (2^{\aleph_0})^{\aleph_0} = 2^{\aleph_0\cdot\aleph_0} = 2^{\aleph_0} = \mathfrak{c}$.

3) Bo $2^{\mathfrak{c}} \leq \aleph_0^{\mathfrak{c}} \leq \mathfrak{c}^{\mathfrak{c}} = (2^{\aleph_0})^{\mathfrak{c}} = 2^{\aleph_0 \cdot \mathfrak{c}} = 2^{\mathfrak{c}}$. □

Uwaga 1: Jak już stwierdziliśmy, działania na liczbach kardynalnych są monotoniczne. Zatem $\mathfrak{m} \leq \mathfrak{n}$ implikuje $\mathfrak{m}+\mathfrak{p} \leq \mathfrak{n}+\mathfrak{p}$, $\mathfrak{m} \cdot \mathfrak{p} \leq \mathfrak{n} \cdot \mathfrak{p}$, $\mathfrak{m}^{\mathfrak{p}} \leq \mathfrak{n}^{\mathfrak{p}}$ i $p^{\mathfrak{m}} \leq \mathfrak{n}^{\mathfrak{p}}$. Takie wynikania nie zachodzą jednak dla nierówności ostrej $<$. Istotnie: mamy wprawdzie $5 < \aleph_0$, ale:

- $5 + \aleph_0 = \aleph_0 + \aleph_0 = \aleph_0$;
- $5 \cdot \aleph_0 = \aleph_0 \cdot \aleph_0 = \aleph_0$;
- $2^{\aleph_0} = \aleph_0^{\aleph_0} = \mathfrak{c}$;
- $\mathfrak{c}^5 = \mathfrak{c}^{\aleph_0} = \mathfrak{c}$.

Uwaga 2: Nie można w sensowny sposób określić odejmowania liczb kardynalnych. Odejmowanie jest działaniem odwrotnym do dodawania. Aby można było je zdefiniować, z warunku $\mathfrak{m} + \mathfrak{p} = \mathfrak{m} + \mathfrak{q} = \mathfrak{n}$ musiałoby wynikać $\mathfrak{p} = \mathfrak{q}$. Wtedy przyjęlibyśmy, że $\mathfrak{n} - \mathfrak{m} = \mathfrak{p}$. Ale skoro na przykład $\aleph_0 + 5 = \aleph_0 + \aleph_0 = \aleph_0$, to różnica $\aleph_0 - \aleph_0$ nie ma sensu. Podobnie nie można zdefiniować dzielenia, pierwiastkowania ani logarytmowania liczb kardynalnych.

Hipoteza continuum *

Nie znamy żadnej liczby $\mathfrak{m}$ spełniającej nierówności $\aleph_0 < \mathfrak{m} < \mathfrak{c}$. Przypuszczenie, że takiej liczby nie ma nazywane jest *hipotezą continuum*. Hipoteza continuum okazała się zdaniem niezależnym od teorii zbiorów. Oznacza to, że na gruncie tej teorii nie można udowodnić ani hipotezy (P.J. Cohen, 1964) ani jej zaprzeczenia (K. Gödel, 1939).

12 Relacje porządkujące

Definicja 12.1 Relację $r \subseteq A \times A$ nazywamy relacją *częściowego porządku* w A, gdy jest

$$\begin{array}{rll} \textit{zwrotna w } A, & \text{czyli} & \forall x(x \in A \to x\, r\, x); \\ \textit{przechodnia}, & \text{czyli} & \forall x \forall y \forall z(x\, r\, y \wedge y\, r\, z \to x\, r\, z); \\ \textit{antysymetryczna}, & \text{czyli} & \forall x \forall y(x\, r\, y \wedge y\, r\, x \to x = y). \end{array}$$

Parę $\langle A, r\rangle$, a czasami sam zbiór A, nazywamy *zbiorem częściowo uporządkowanym*, lub po prostu *częściowym porządkiem*. Określenie „częściowy porządek" jest też używane w stosunku do samej relacji. Jeśli dodatkowo relacja r jest spójna w A, tj.

$$\forall x \forall y(x, y \in A \to (x\, r\, y \vee y\, r\, x))$$

to mówimy, że jest to relacja *liniowego porządku*. Określenia *liniowy porządek*, *zbiór liniowo uporządkowany*, stosuje się odpowiednio.

Przykład 12.2

- Relacja $\leq$ w zbiorze liczb naturalnych jest liniowym porządkiem.
- Typ $\mathbb{N}$ jest częściowo uporządkowany przez relację podzielności:[18]

  $$m|n \text{ wtedy i tylko wtedy, gdy } \exists k{:}\mathbb{N}\,(k \cdot m = n).$$

- Każda rodzina zbiorów jest częściowo uporządkowana przez inkluzję.
- Słowa są częściowo uporządkowane przez relację $\subseteq$ porządku prefiksowego.

Relacje częściowo porządkujące najczęściej oznaczamy symbolami $\leq$, $\preceq$, $\sqsubseteq$ i podobnymi. Jeśli relacja $\leq$ jest częściowym porządkiem w A , to relacja $<$ jest domyślnie określona tak:

$$x < y \quad \text{wtedy i tylko wtedy, gdy } \; x \leq y \text{ i } x \neq y.$$

Dla $A \neq \varnothing$, ta relacja *nie jest* częściowym porządkiem, bo nie jest zwrotna. Notację $\prec$, $\sqsubset$ itp. stosujemy odpowiednio.

Jeśli $\langle A, r\rangle$ jest częściowym (liniowym) porządkiem, oraz $B \subseteq A$, to łatwo zauważyć, że $\langle B, r \cap (B \times B)\rangle$ jest też częściowym (liniowym) porządkiem. Dla prostoty oznaczamy go przez $\langle B, r\rangle$. Na przykład każdy podzbiór $\mathbb{N}$ jest częściowo uporządkowany przez relację podzielności.

Definicja 12.3 Niech $\langle A, \leq\rangle$ będzie częściowym porządkiem.

1. Elementy $a, b \in A$ są *porównywalne*, gdy $a \leq b$ lub $b \leq a$. W przeciwnym razie a, b są *nieporównywalne*.
2. Jeśli $B \subseteq A$ i każde dwa elementy zbioru B są porównywalne (tj. $\langle B, \leq\rangle$ jest liniowo uporządkowany) to mówimy, że B jest *łańcuchem* w A.

[18] Uwaga: w myśl tej definicji, zero jest podzielne przez każdą liczbę, w tym przez siebie.

3. Jeśli $B \subseteq A$ i każde dwa różne elementy zbioru B są nieporównywalne, to mówimy, że B jest *antyłańcuchem* w A.

Ostrzeżenie: W zbiorze częściowo uporządkowanym z warunku $x \not\leq y$ nie wynika $x > y$! Elementy x, y mogą być nieporównywalne.

Porządkowanie słów

Niech A będzie ustalonym alfabetem. Przypomnijmy, że typ słów nad A oznaczamy przez A^*, i że symbol $\subseteq$ oznacza także porządek prefiksowy na słowach. Jeśli alfabet A jest uporządkowany przez jakąś relację $\leq$, to w A^* możemy określić *porządek leksykograficzny* $\preceq$. Przyjmujemy, że $w \preceq v$, gdy zachodzi jedna z możliwości

- $w \subseteq v$;
- Istnieje takie słowo u, że $ua \subseteq w$ i $ub \subseteq v$, dla pewnych $a, b \in A$ takich, że $a < b$.

Na przykład, jeśli $a < b$, to $\varepsilon \preceq ab \preceq aba \preceq baba \preceq bba$. Porządek leksykograficzny jest wyznaczony przez „pierwszą różnicę" pomiędzy słowami. Żeby to wyrazić ściślej, oznaczmy przez $w(0)$, $w(1)$, $w(2), \ldots$ kolejne litery słowa w, zdefiniowane przez taką indukcję: $\varepsilon(n)$ jest zawsze nieokreślone, $aw(0) = a$, $bw(0) = b$, $aw(n+1) = bw(n+1) = w(n)$.

Lemat 12.4 *Niech $w \not\subseteq v$, $v \not\subseteq w$ i niech k będzie najmniejszą taką liczbą, że $w(k) \neq v(k)$. Wówczas $w \preceq v$ wtedy i tylko wtedy, gdy $w(k) < v(k)$.*

Dowód: Ćwiczenie. □

Lemat 12.5 *Jeśli $w \subseteq xay$ to albo $w \subseteq x$ albo $xa \subseteq w$.*

Dowód: Ćwiczenie. *Wskazówka:* Pokazać najpierw, że $w \subseteq v$ zachodzi wtedy i tylko wtedy, gdy $w(n) = v(n)$ dla $n < |w|$. □

Fakt 12.6 *Porządek leksykograficzny jest relacją częściowego porządku w zbiorze A^*. Jeśli alfabet jest liniowo uporządkowany, to porządek leksykograficzny też jest liniowy.*

Dowód: Zwrotność relacji $\preceq$ wynika ze zwrotności relacji $\subseteq$. Aby udowodnić przechodniość załóżmy, że $w \preceq v$ i $v \preceq x$. Mamy do rozpatrzenia 4 przypadki.

Przypadek 1: $w \subseteq v$ i $v \subseteq x$. Wtedy oczywiście $w \subseteq x$.

Przypadek 2: $w \subseteq v = uav'$, oraz $x = ubx'$, gdzie $a < b$. Mamy tu dwie możliwości (ćwiczenie 12.5): albo $w \subseteq u$ albo $ua \subseteq w$. Wtedy odpowiednio, albo $w \subseteq x$,albo $w = uaw'$, a wtedy $w \preceq x$ na mocy drugiej części definicji.

Przypadek 3: $w = uaw'$ oraz $v = ubv' \subseteq x$ i $a < b$. Wtedy $x = ubv'x'$ i mamy $w \preceq x$ na mocy drugiej części definicji.

Przypadek 4: $w = uaw'$ i $v = ubv'$ oraz jednocześnie $v = u'a'v''$ i $x = u'b'x'$ gdzie $a < b$ i $a' < b'$. Skoro $v = ubv' = u'a'v''$, to $ub \subseteq u'$ lub $u'a' \subseteq u$ (ćwiczenie 12.5). W pierwszym przypadku $u' = ubu''$, więc $w = uaw' \preceq ubu''b'x' = x$. W drugim przypadku $u = u'a'u''$, więc $w = u'au''w' \preceq u'bx' = x$.

Pozostaje wykazać antysymetrię. Niech więc $w \preceq v$ i $v \preceq w$. Tu też mamy cztery przypadki, analogiczne do rozpatrzonych powyżej. Zauważmy jednak, że powtarzając poprzednie rozumowanie dla $x = w$, w przypadkach 2, 3 i 4 otrzymamy sprzeczność. Okaże się bowiem, że $w = uaw' = ubw''$, gdzie $a < b$. Zostaje więc tylko przypadek 1, a wtedy $w = v$.

Jeśli alfabet jest liniowo uporządkowany, to spójność relacji $\preceq$ wynika z lematu 12.4. □

Elementy wyróżnione

Definicja 12.7 Niech $\langle A, \leq \rangle$ będzie częściowym porządkiem i niech $a \in A$. Mówimy, że element a jest w zbiorze A:

największy,	gdy	$\forall x \in A\,(x \leq a)$;
maksymalny,	gdy	$\forall x \in A\,(a \leq x \to a = x)$;
najmniejszy,	gdy	$\forall x \in A\,(a \leq x)$;
minimalny,	gdy	$\forall x \in A\,(x \leq a \to a = x)$.

Fakt 12.8 *Jeśli a jest elementem największym (najmniejszym) w $\langle A, \leq \rangle$, to jest też elementem maksymalnym (minimalnym). Innych elementów maksymalnych (minimalnych) wtedy nie ma.*

Dowód: Załóżmy, że a jest największy w A. Aby pokazać, że jest maksymalny, przypuśćmy, że $a \leq x$. Ale skoro a jest największy, to $x \leq a$ więc $a = x$. Niech teraz $b \in A$ będzie też elementem maksymalnym. Skoro a jest największy, to $b \leq a$ więc $b = a$ bo b jest maksymalny. A więc a jest jedynym elementem maksymalnym w A. □

Przykład 12.9

- W zbiorze uporządkowanym $\langle \mathbb{N}, \mid \rangle$ gdzie $\mid$ oznacza relację podzielności (przykład 12.2), zero jest elementem największym, a 1 najmniejszym.
- W porządku $\langle \mathbb{N} - \{0, 1\}, \mid \rangle$ nie ma elementu najmniejszego ani żadnych elementów maksymalnych. Natomiast liczby pierwsze są elementami minimalnymi.
- W zbiorze $\mathbb{Z}$ liczb całkowitych, uporządkowanym przez zwykłą relację $\leq$, nie ma żadnych elementów minimalnych ani maksymalnych.
- Rozpatrzmy częściowy porządek $\langle \mathbb{Z} \oplus \{\omega\}, \preceq \rangle$ gdzie:

  $$x \preceq y \quad \Leftrightarrow \quad [(x, y \in \mathbb{Z}) \wedge (x \leq y)] \vee [x = y = \omega]$$

  Ten porządek ma tylko jeden element minimalny ω, ale nie ma elementu najmniejszego.

Uwaga: Relacja odwrotna do relacji częściowo porządkującej r też jest relacją częściowo porządkującą. Elementy minimalne ze względu na r są elementami maksymalnymi ze względu na r^{-1} i na odwrót. Podobny dualizm dotyczy elementów największych i najmniejszych. Dlatego wszystkie fakty dotyczące elementów maksymalnych i największych stosują się też odpowiednio do elementów minimalnych i najmniejszych.

Fakt 12.10

1) Każdy skończony i niepusty częściowy porządek ma element maksymalny.

2) Jeśli $\langle A, \leq \rangle$ jest porządkiem liniowym i $a \in A$ jest jego elementem maksymalnym to a jest elementem największym.

3) A zatem każdy skończony i niepusty liniowy porządek ma element największy.

4) Analogiczne fakty mają miejsce w odniesieniu do elementów najmniejszych i minimalnych.

Dowód: (1) Przez indukcję ze względu na $n \geq 1$, pokażemy, że każdy częściowy porządek mocy n ma element maksymalny. Jeśli zbiór ma tylko jeden element, to ten element jest oczywiście maksymalny. Załóżmy więc, że teza zachodzi dla zbiorów n-elementowych i niech $\langle A, \leq \rangle$ będzie zbiorem częściowo uporządkowanym o $n+1$ elementach. Wtedy możemy przedstawić zbiór A jako sumę $A = B \cup \{a\}$, gdzie B jest zbiorem n-elementowym oraz $a \notin B$. Z założenia indukcyjnego B ma element maksymalny b. Jeśli teraz $b \not\leq a$, to b jest elementem maksymalnym w A. W przeciwnym razie elementem maksymalnym jest a. Istotnie, przypuśćmy, że $a \leq c$. Wtedy $c = a$ (i dobrze) lub $c \in B$. W tym drugim przypadku łatwo zauważyć, że $a = b = c$, bo b jest maksymalny w B.

(2) Załóżmy, że $\langle A, \leq \rangle$ jest porządkiem liniowym i $a \in A$ jest maksymalny. Niech $b \in A$. Gdyby $b \not\leq a$ to $a \leq b$, więc $a = b$ z maksymalności.

(3) Oczywista konsekwencja (1) i (2).

(4) Należy zastosować (1), (2) i (3) do porządku odwrotnego. □

Definicja 12.11 Niech $\langle A, \leq \rangle$ będzie porządkiem częściowym i niech $B \subseteq A$ i $a \in A$. Mówimy, że a jest *ograniczeniem górnym* zbioru B (oznaczenie $a \geq B$), gdy $b \leq a$ dla wszystkich $b \in B$.

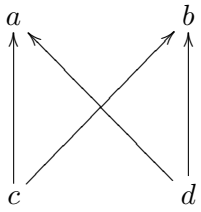
Element a jest *kresem górnym* zbioru B (oznaczenie $a = \sup B$), gdy jest najmniejszym ograniczeniem górnym B, czyli:

- $a \geq B$;
- jeśli $c \geq B$ to $c \geq a$, dla dowolnego $c \in A$.

Analogicznie definiujemy ograniczenia dolne (oznaczenie $a \leq B$) i kresy dolne (oznaczenie $a = \inf B$).

Przykład 12.12

- W rodzinie wszystkich podzbiorów zbioru A (uporządkowanej przez inkluzję) kresem górnym dowolnej podrodziny $X \subseteq \mathsf{P}(A)$ jest suma $\bigcup X$.
- W rodzinie wszystkich wypukłych[19] podzbiorów płaszczyzny, każdy podzbiór X ma kres górny. Kresem tym jest iloczyn wszystkich zbiorów wypukłych zawierających wszystkie zbiory z X. Zwykle nie jest to $\bigcup X$, bo suma nie musi być wypukła.
- W zbiorze liczb wymiernych $\mathbb{Q}$ ze zwykłym uporządkowaniem zbiór $\{q \in \mathbb{Q} \mid q^2 < 2\}$ ma ograniczenia górne ale nie ma kresu górnego.
- W zbiorze liczb rzeczywistych $\mathbb{R}$ każdy niepusty podzbiór ograniczony z góry ma kres górny (i analogicznie z dołu). Własność tę nazywamy *ciągłością*.
- W zbiorze $\{a, b, c, d\}$ uporządkowanym jak na rysunku, podzbiór $\{c, d\}$ ma dwa ograniczenia górne, ale nie ma kresu górnego.



Następujący fakt podamy na razie bez dowodu.

Twierdzenie 12.13 (Lemat Kuratowskiego-Zorna) *Niech $\langle A, \leq \rangle$ będzie zbiorem częściowo uporządkowanym, spełniającym następujący warunek:*

(*) *Każdy łańcuch ma w A ograniczenie górne.*

Wtedy w A istnieje element maksymalny.

Następujące twierdzenie stanowi ważny przykład zastosowania Lematu Kuratowskiego-Zorna. Przypomnijmy, że podzbiór A przestrzeni liniowej V jest *liniowo niezależny*, jeśli z warunku $k_1 v_1 + \cdots + k_n v_n = 0$, gdzie $v_1, \ldots, v_n \in A$, wynika $k_1 = \cdots = k_n = 0$. Zbiór A jest *bazą* przestrzeni V, wtedy i tylko wtedy, gdy jest liniowo niezależny, oraz każdy element przestrzeni jest kombinacją liniową elementów zbioru A.

Twierdzenie 12.14 *Każda przestrzeń liniowa ma bazę.*

Dowód: Nietrudno zauważyć, że liniowo niezależny zbiór A jest bazą przestrzeni V wtedy i tylko wtedy, gdy dodanie do zbioru A dowolnego nowego elementu powoduje utratę liniowej niezależności. A zatem baza to element maksymalny rodziny

$$Z = \{A \subseteq V \mid A \text{ jest liniowo niezależny}\},$$

[19] Zbiór jest *wypukły* wtedy i tylko wtedy, gdy wraz z dowolnymi dwoma punktami zawiera odcinek łączący te punkty.

uporządkowanej przez inkluzję. Użyjemy więc Lematu Kuratowskiego-Zorna, aby wykazać istnienie elementu maksymalnego zbioru Z. W tym celu wystarczy stwierdzić, że każdy łańcuch jest w Z ograniczony z góry. Niech więc $Ł$ będzie łańcuchem w Z i niech $B = \bigcup Ł$. Pokażemy, że zbiór B jest liniowo niezależny.

Istotnie, przypuśćmy, że $k_1v_1 + \cdots + k_nv_n = 0$, gdzie $v_1, \ldots, v_n \in B$. Skoro wektory $v_1, \ldots, v_n$ należą do sumy łańcucha $Ł$, to każdy z nich należy do pewnego składnika. Stąd wynika, że $v_1 \in A_1, \ldots, v_n \in A_n$ dla pewnych $A_1, \ldots, A_n \in Ł$. Rodzina zbiorów $\{A_1, \ldots, A_n\}$ jest skończona i liniowo uporządkowana przez inkluzję, ma więc element największy na mocy faktu 12.10(3). To znaczy, że dla pewnego i mamy $v_1, \ldots, v_n \in A_i$, a przecież zbiór A_i jest liniowo niezależny. Stąd kombinacja liniowa $k_1v_1 + \cdots + k_nv_n = 0$ musi byc trywialna i mamy $k_1 = \cdots = k_n = 0$.

Ponieważ B jest liniowo niezależny, więc $B \in Z$, a przy tym oczywiście B zawiera wszystkie elementy $Ł$, jest więc ograniczeniem górnym naszego łańcucha w zbiorze Z. Spełnione jest więc założenie twierdzenia 12.13 i musi istnieć element maksymalny. □

Strukturę powyższego dowodu zilustrujemy schematycznie z pomocą pudełek.

Cel 1: Każdy łańcuch jest ograniczony z góry.

Załóżmy, że $Ł$ jest łańcuchem. *Cel 2: $B = \bigcup Ł$ ogranicza $Ł$ w Z.*
(Uwaga: Cel 2 oznacza, że $B \in Z$ oraz $\forall A\,.A \in Ł \to A \subseteq B$)
Cel 3: $B \in Z$
(Uwaga: Cel 3 oznacza, że B jest liniowo niezależny)

> Niech $v_1, \ldots, v_n \in B$ będą takie, że $k_1v_1 + \cdots + k_nv_n = 0 \ldots$ *Cel 4: $k_1 = \cdots = k_n = 0$.*
> ⋮
> Zatem $k_1 = \cdots = k_n = 0$ *(Cel 4 osiągnięty)*

Zatem B jest liniowo niezależny, tj. $B \in Z$ *(Cel 3 osiągnięty)*
Łatwo widzieć, że $\forall A\,.A \in Ł \to A \subseteq B$
Zatem B jest ograniczeniem $Ł$ w Z *(Cel 2 osiągnięty)*
Zatem każdy łańcuch ma ograniczenie górne *(Cel 1 osiągnięty)*
Z lematu Kuratowskiego-Zorna istnieje element maksymalny.

13 Punkty stałe

Definicja 13.1 Niech $\langle A, \leq \rangle$ będzie porządkiem częściowym.

- Podzbiór B zbioru A jest *skierowany* wtedy i tylko wtedy, gdy dla dowolnych $a, b \in B$ istnieje takie $c \in B$, że $a, b \leq c$.
- Zbiór A jest *zupełnym* porządkiem częściowym (cpo) wtedy i tylko wtedy, gdy każdy jego skierowany podzbiór ma kres górny.
- Zbiór A jest *kratą zupełną* wtedy i tylko wtedy, gdy każdy podzbiór A ma kres górny.

Oczywiście każdy łańcuch jest zbiorem skierowanym. W szczególności elementy dowolnego ciągu wstępującego $a_0 \leq a_1 \leq a_2 \leq \ldots$ tworzą zbiór skierowany. Także zbiór pusty jest zbiorem skierowanym. Z definicji porządku zupełnego wynika więc istnienie elementu najmniejszego $\sup \varnothing$, tradycyjnie oznaczanego przez $\bot$.

Fakt 13.2 *W kracie zupełnej każdy podzbiór ma kres dolny.*

Dowód: Niech $\langle A, \leq \rangle$ będzie kratą zupełną i niech $B \subseteq A$. Przez C oznaczmy zbiór wszystkich ograniczeń dolnych zbioru B:

$$C = \{x \in A \mid x \leq B\}.$$

Teraz jeśli $b \in B$ to $b \geq C$, więc dla $c = \sup C$ mamy $b \geq c$. To znaczy, że c jest ograniczeniem dolnym zbioru B. Co więcej, c jest kresem dolnym, bo $x \leq B$ oczywiście implikuje $x \leq c$. □

Definicja 13.3 Niech $\langle A, \leq \rangle$ i $\langle B, \leq \rangle$ będą porządkami częściowymi.

- Funkcja $f : A \to B$ jest *monotoniczna* wtedy i tylko wtedy, gdy dla dowolnych $x, y \in A$ nierówność $x \leq y$ implikuje $f(x) \leq f(y)$.
- Jeśli $\langle A, \leq \rangle$ i $\langle B, \leq \rangle$ są zupełnymi porządkami częściowymi to funkcja $f : A \to B$ jest *ciągła* wtedy i tylko wtedy, gdy f zachowuje kresy górne niepustych zbiorów skierowanych, tj. dla dowolnego skierowanego i niepustego podzbioru $X \subseteq A$ istnieje $\sup f(X)$ i zachodzi równość $f(\sup X) = \sup f(X)$.
- Jeśli $f : A \to A$ oraz $f(a) = a$, to mówimy, że a jest *punktem stałym* funkcji f. *Najmniejszy* punkt stały danej funkcji to najmniejszy element zbioru wszystkich jej punktów stałych (o ile taki istnieje).

Fakt 13.4 *Każda funkcja ciągła jest monotoniczna.*

Dowód: Niech $x \leq y$. Wtedy zbiór $\{x, y\}$ jest skierowany, a jego kresem górnym jest y. Zatem $f(y)$ jest kresem górnym zbioru $\{f(x), f(y)\}$, czyli $f(x) \leq f(y)$. □

Twierdzenie 13.5 *Jeśli zbiór częściowo uporządkowany $\langle A, \leq \rangle$ jest kratą zupełną, to każda funkcja monotoniczna $f : A \to A$ ma najmniejszy punkt stały.*

Dowód: Rozpatrzmy zbiór $B = \{x \in A \mid f(x) \leq x\}$. Niech $a = \inf B$. Pokażemy, że a jest najmniejszym punktem stałym funkcji f.

Dla dowolnego $x \in B$ mamy $a \leq x$, więc $f(a) \leq f(x) \leq x$. Zatem $f(a)$ jest ograniczeniem dolnym zbioru B, skąd $f(a) \leq a$, bo a jest kresem dolnym.

Ale skoro $f(a) \leq a$, to także $f(f(a)) \leq f(a)$, więc $f(a) \in B$. Zatem $a \leq f(a)$ i mamy równość.

Ponieważ wszystkie punkty stałe funkcji f muszą należeć do B, więc a jest najmniejszym punktem stałym. □

Nie zawsze mamy do czynienia z kratami zupełnymi. Ale jeśli funkcja jest ciągła, to można to założenie osłabić. Przypomnijmy, że dla dowolnej funkcji $f : A \to A$ notacja f^n oznacza n-krotne złożenie funkcji f, tj. $f^0 = \mathrm{id}_A$ oraz $f^{n+1} = f \circ f^n$.

Twierdzenie 13.6 *Jeśli $\langle A, \leq \rangle$ jest zupełnym porządkiem częściowym, to każda funkcja ciągła $f : A \to A$ ma najmniejszy punkt stały, którym jest* $\sup\{f^n(\bot) \mid n \in \mathbb{N}\}$.

Dowód: Oczywiście $\bot \leq f(\bot)$. Ponieważ f jest monotoniczna (fakt 13.4), więc przez łatwą indukcję wnioskujemy, że ciąg $f^n(\bot)$ jest wstępujący: $f^n(\bot) \leq f^m(\bot)$ dla $n \leq m$. A zatem zbiór $\{f^n(\bot) \mid n \in \mathbb{N}\}$ jest skierowany i ma kres górny. Z ciągłości funkcji dostajemy

$$f(\sup\{f^n(\bot) \mid n \in \mathbb{N}\}) = \sup\{f^{n+1}(\bot) \mid n \in \mathbb{N}\} = \sup\{f^n(\bot) \mid n \in \mathbb{N}\},$$

czyli $a = \sup\{f^n(\bot) \mid n \in \mathbb{N}\}$ jest punktem stałym. Pozostaje sprawdzić, że jest najmniejszy.

Jeśli b jest innym punktem stałym, to przez indukcję wnioskujemy, że $f^n(\bot) \leq b$ dla dowolnego $n \in \mathbb{N}$. (Zaczynamy od oczywistej nierówności $\bot \leq b$, a krok indukcyjny wynika z monotoniczności: $f^{n+1}(\bot) \leq f(b) = b$.) A zatem $b \geq \{f^n(\bot) \mid n \in \mathbb{N}\}$ skąd $b \geq a$. □

Uwaga: W przypadku funkcji monotonicznej, która nie jest ciągła, kres górny ciągu $f^n(\bot)$ może nie być punktem stałym. Ale jeśli jest punktem stałym, to zawsze najmniejszym.

Omówimy teraz kilka przykładów, w których występują punkty stałe przekształceń monotonicznych. Pierwszy przykład dotyczy dosyć typowej sytuacji gdy pewien zbiór rozszerzamy o nowe elementy, tak aby otrzymać nowy zbiór zamknięty ze względu na pewne operacje. Drugi dotyczy definiowania języka formalnego.

Przykład 13.7 Niech r będzie relacją w zbiorze A. Przypomnijmy, że $s \cdot s'$ oznacza złożenie relacji s i s'. Rozpatrzmy zbiór $\mathsf{P}(A \times A)$ uporządkowany przez inkluzję, oraz funkcję $f : \mathsf{P}(A \times A) \to \mathsf{P}(A \times A)$ określoną tak:

$$f(s) = \mathbf{1}_A \cup r \cup s \cup (s \cdot s).$$

Funkcja f jest ciągła (ćwiczenie), więc ma najmniejszy punkt stały. Jest to relacja r^*, czyli przechodnio-zwrotne domknięcie relacji r. Żeby się o tym przekonać, należy zauważyć, że warunek $f(s) = s$ zachodzi wtedy i tylko wtedy, gdy relacja s jest przechodnia (tj. $s \cdot s \subseteq s$) i zwrotna (tj. $\mathbf{1}_A \subseteq s$) oraz $r \subseteq s$.

Przykład 13.8 *Palindrom* to słowo, które czyta się w obie strony tak samo, np.

adapannapocałowanawołacopannapada

Definicję (niepustego) palindromu nad $\{a, b\}^*$ można podać w postaci *gramatyki bezkontekstowej:*

$$X ::= \varepsilon \mid a \mid b \mid a\,X\,a \mid b\,X\,b$$

Rozumiemy ją tak:

- Słowo puste i słowo jednoliterowe jest palindromem.
- Jeśli X jest palindromem, to $a\,X\,a$ jest palindromem.
- Jeśli X jest palindromem, to $b\,X\,b$ jest palindromem.
- Nie ma innych palindromów.

Zbiór $\mathcal{P}$ wszystkich palindromów spełnia warunek

$$\mathcal{P} = \{\varepsilon, a, b\} \;\cup\; \{aXa \mid X \in \mathcal{P}\} \;\cup\; \{bXb \mid X \in \mathcal{P}\}.$$

Jest to najmniejszy zbiór o tej własności, tj. najmniejszy punkt stały przekształcenia

$$\lambda P{:}\mathsf{P}(\{a, b\}^*).\, \{\varepsilon, a, b\} \;\cup\; \{aXa \mid X \in P\} \;\cup\; \{bXb \mid X \in P\}.$$

Przedsmak semantyki denotacyjnej

Kolej na nieco bardziej rozbudowany przykład. Typ $A \multimap B$ wszystkich funkcji częściowych z A do B jest częściowo uporządkowany przez inkluzję. Co więcej, jest to porządek zupełny (chociaż nie krata zupełna). Funkcje częściowe z A do B wygodnie jest utożsamiać z funkcjami całkowitymi o dziedzinie A i wartościach w $B_\perp = B \oplus \{\perp\}$, gdzie $\perp$ reprezentuje „wartość nieokreśloną". Jeśli umówimy się, że zbiór $B_\perp$ jest uporządkowany tak:

$$b \le b' \quad \text{wtedy i tylko wtedy, gdy} \quad b = \perp \text{ lub } b = b',$$

to możemy zauważyć, że $A \multimap B$ ma uporządkowanie „po współrzędnych":

$$f \le g \quad \text{wtedy i tylko wtedy, gdy} \quad \forall a \in A\,(f(a) \le g(a)).$$

Rozpatrzmy teraz następujący „program" definiujący funkcję częściową $f : \mathbb{Z} \times \mathbb{Z} \multimap \mathbb{Z}$.

$$f(m, n) = \textbf{if } m = n \textbf{ then } 0 \textbf{ else } f(m + 3, n) + 3. \qquad (*)$$

Ta definicja, rozumiana jako równanie na funkcjach częściowych, nie wyznacza jednoznacznie funkcji f. Równanie ma więcej niż jedno rozwiązanie. Inaczej mówiąc, operator na funkcjach częściowych

$$\Phi : (\mathbb{Z} \times \mathbb{Z} \multimap \mathbb{Z}) \to (\mathbb{Z} \times \mathbb{Z} \multimap \mathbb{Z}),$$

określony warunkiem

$$\Phi(f)(m, n) = \textbf{if } m = n \textbf{ then } 0 \textbf{ else } f(m + 3, n) + 3,$$

ma więcej niż jeden punkt stały. Na przykład

- $f_1(m, n) = n - m$;
- $f_2(m, n) = \textbf{if } 3|(n - m) \textbf{ then } n - m \textbf{ else } 7 - m$;
- $f_0(m, n) = \textbf{if } m \le n \wedge 3|(n - m) \textbf{ then } n - m \textbf{ else } \perp$.

Ale tylko jeden z tych punktów stałych odpowiada obliczeniowemu rozumieniu definicji rekurencyjnej (*). Jest to *najmniejszy* punkt stały:

$$f(m,n) = \begin{cases} n-m, & \text{jeśli } m \leq n \text{ oraz } 3|(n-m); \\ \text{nieokreślone}, & \text{w przeciwnym przypadku.} \end{cases}$$

Funkcja f obliczana przez *program* zadany równaniem (*) jest sumą ciągu funkcji częściowych $f_k = \Phi^k(\bot)$, gdzie $\bot$ to funkcja nigdzie nie określona. Łatwo widzieć, że f_k określone jest dla tych par $\langle m, n\rangle$ dla których obliczenie wymaga nie więcej niż $k-1$ odwołań rekurencyjnych.

Ćwiczenie 13.9 *Wyznaczyć kilka początkowych wartości ciągu $\Phi^k(\bot)$, gdzie Φ jest zadane definicją rekurencyjną*

$$f(m) = \textbf{if } m \leq 1 \textbf{ then } 1 \textbf{ else if } \mathit{parzyste}(m) \textbf{ then } f(m/2) \textbf{ else } f(3m+1).$$

Bisymulacje

Najmniejsze punkty stałe występują wszędzie tam, gdzie mamy do czynienia z indukcją, rekursją itp. Ale czasami przydatne jest też pojęcie największego punktu stałego. Rozpatrzmy następujący przykład.

Przypuśćmy, że dany jest pewien zbiór A, w którym określona jest rodzina P relacji dwuargumentowych. O elementach A myślimy jako o możliwych stanach pewnego procesu, a relacje ze zbioru P reprezentują różne rodzaje możliwych przejść pomiędzy stanami. Aby to podkreślić, zamiast $\langle a, b\rangle \in \alpha$ (dla $\alpha \in P$) piszemy $a \rightsquigarrow_\alpha b$. Załóżmy dodatkowo, nasz proces ma własność „skończonego niedeterminizmu", tj. że

$$\textit{dla dowolnych } a \textit{ i } \alpha \textit{ zbiór } \{b \mid a \rightsquigarrow_\alpha b\} \textit{ jest skończony.} \qquad (*)$$

Relacja $\sim$ w zbiorze A jest *(częściową) bisymulacją*[20], gdy dla dowolnych $a_1, a_2 \in A$ takich, że $a_1 \sim a_2$, i dowolnego $\alpha \in P$, zachodzą następujące warunki:

- Jeśli $a_1 \rightsquigarrow_\alpha b_1$ dla pewnego b_1, to istnieje takie b_2, że $a_2 \rightsquigarrow_\alpha b_2$ i $b_1 \sim b_2$.
- Jeśli $a_2 \rightsquigarrow_\alpha b_2$ dla pewnego b_2, to istnieje takie b_1, że $a_1 \rightsquigarrow_\alpha b_1$ i $b_1 \sim b_2$.

Sens tej definicji jest taki: zachowanie procesu uruchomionego w stanie a_1 może być „symulowane" przez proces uruchomiony w stanie a_2, i na odwrót.

Zauważmy, że suma wszystkich bisymulacji częściowych jest bisymulacją. Jest to największa możliwa bisymulacja. Oznaczymy ją przez $\approx$ i nazwiemy *pełną bisymulacją*[21]. A więc warunek $a_1 \approx a_2$ to najsłabszy warunek gwarantujący analogiczne zachowanie procesu w obu stanach.

Rozpatrzmy teraz następujący operator $\mathcal{F} : \mathsf{P}(A \times A) \to \mathsf{P}(A \times A)$:

$$\begin{aligned} \mathcal{F}(r) = &\{\langle a_1, a_2\rangle \mid \forall\alpha\forall b_1\,(a_1 \rightsquigarrow_\alpha b_1 \to \exists b_2\,(b_1\ r\ b_2 \wedge a_2 \rightsquigarrow_\alpha b_2))\} \\ &\cap \{\langle a_1, a_2\rangle \mid \forall\alpha\forall b_2\,(a_2 \rightsquigarrow_\alpha b_2 \to \exists b_1\,(b_1\ r\ b_2 \wedge a_1 \rightsquigarrow_\alpha b_1))\}. \end{aligned} \qquad (**)$$

[20] Ang.: bisimulation.
[21] Ang.: bisimilarity.

Nietrudno zauważyć, że $\mathcal{F}$ jest operatorem ciągłym[22] (ćwiczenie). Częściowe bisymulacje to dokładnie te relacje, które spełniają warunek $r \subseteq \mathcal{F}(r)$. Pełna bisymulacja jest największym punktem stałym operatora $\mathcal{F}$ (zauważmy tu analogię z konstrukcją w dowodzie twierdzenia 13.5). Co więcej, relacja $\approx$ jest iloczynem (kresem dolnym) zstępującego ciągu relacji $\top, \mathcal{F}(\top), \mathcal{F}^2(\top), \ldots$ Symbol $\top$ oznacza oczywiście relację pełną $A \times A$, czyli największy element kraty zupełnej $\mathsf{P}(A \times A)$. Zauważmy jeszcze, że k-te przybliżenie $\mathcal{F}^k(\top)$ relacji $\approx$ można interpretować jako najsłabszą relację gwarantującą takie same zachowanie procesu przez pierwsze k kroków.

Ćwiczenie 13.10 Czy bisymulacja musi być przechodnia? A zwrotna? Czy relacja pusta jest bisymulacją? Czy pełna bisymulacja jest relacją równoważności?

[22] Nie będzie to jednak prawdą, jeśli zrezygnujemy z założenia (*). Mimo to, największy punkt stały istnieje, chociaż nie jest kresem dolnym ciągu $\mathcal{F}^n(\top)$.

14 Izomorfizmy porządków

Często mamy do czynienia z dwoma zbiorami, które są różne, ale „tak samo" uporządkowane. Takie porządki nazywamy izomorficznymi.

Definicja 14.1 Mówimy, że zbiory częściowo uporządkowane $\langle A, \leq\rangle$ i $\langle B, \leq\rangle$ są *izomorficzne*, gdy istnieje bijekcja $f : A \xrightarrow[\text{na}]{1-1} B$ spełniająca warunek

$$a \leq a' \quad \Leftrightarrow \quad f(a) \leq f(a'),$$

dla dowolnych $a, a' \in A$. Piszemy wtedy $\langle A, \leq\rangle \approx \langle B, \leq\rangle$ albo po prostu $A \approx B$, a funkcję f nazywamy *izomorfizmem*.

Jeśli dwa zbiory częściowo uporządkowane są izomorficzne i jeden z nich

- ma element najmniejszy, największy, maksymalny, minimalny;[23]
- jest liniowo uporządkowany;
- jest cpo, jest kratą zupełną;
- i tak dalej,

to ten drugi też ma odpowiednią własność. Zamiast „i tak dalej" można wstawić dowolny warunek dotyczący tylko relacji porządkującej.

Przykład 14.2 Rozpatrzmy następujące podzbiory $\mathbb{R}$, uporządkowane jak zwykle:

- $A = \{1 - \frac{1}{n} \mid n \in \mathbb{N} - \{0\}\}$;
- $A' = \{1 - \frac{1}{n} \mid n \in \mathbb{N} - \{0\}\} \cup \{1\}$;
- $A'' = \{1 - \frac{1}{n} \mid n \in \mathbb{N} - \{0\}\} \cup \{1, 2\}$;
- $B = \{m - \frac{1}{n} \mid m, n \in \mathbb{N} - \{0\}\}$.

Zbiór wszystkich liczb naturalnych $\mathbb{N}$ jest izomorficzny ze zbiorem A, ale żadne dwa spośród zbiorów: A, A', A'', B nie są izomorficzne. (Na przykład $A \not\approx A'$, bo A nie ma elementu największego.)

Mniej oczywisty jest następny fakt. Mówimy, że zbiór liniowo uporządkowany A jest *gęsty*, gdy dla dowolnych $a, b \in A$, jeśli $a < b$ to $a < c < b$ dla pewnego c.

Twierdzenie 14.3

1. *Każdy przeliczalny zbiór liniowo uporządkowany jest izomorficzny z pewnym podzbiorem zbioru $\mathbb{Q}$ wszystkich liczb wymiernych.*

[23] Niepotrzebne skreślić.

2. *Każdy przeliczalny zbiór gęsty bez końców (tj. bez elementu największego i najmniejszego) jest izomorficzny z* $\mathbb{Q}$.

Dowód: Załóżmy, że $\langle A, \leq\rangle$ jest przeliczalnym zbiorem liniowo uporządkowanym. Bez straty ogólności zakładamy, że A jest nieskończony, tj. $A = \{a_n \mid n \in \mathbb{N}\}$, gdzie wszystkie a_n są różne. Podobnie, zbiór liczb wymiernych przedstawimy w postaci $\mathbb{Q} = \{q_n \mid n \in \mathbb{N}\}$, gdzie wszystkie q_n są różne.

Określamy funkcję $f : A \xrightarrow{1-1} \mathbb{Q}$, definiując $f(a_n)$ przez indukcję ze względu na n, w ten sposób, aby dla dowolnych $i, j \leq n$ zachodziła równoważność:

$$a_i < a_j \qquad \text{wtedy i tylko wtedy, gdy} \qquad f(a_i) < f(a_j). \tag{$*$}$$

Przypuśćmy więc, że $f(a_i)$ są już określone dla $i < n$, i że założenie indukcyjne $(*)$ zachodzi dla $i, j < n$. Ustawmy w ciąg rosnący $a_{i_1} < a_{i_1} < \cdots < a_{i_n}$ elementy $a_0, \ldots, a_{n-1}$. Wtedy liczby $f(a_{i_1}) < f(a_{i_1}) < \cdots < f(a_{i_n})$ także tworzą ciąg rosnący.

Jeśli $n = 0$, to przyjmijmy $X_0 = A$ i $Y_0 = \mathbb{Q}$. Jeśli zaś $n > 0$, to niech

- $X_0 = \{a \in A \mid a < a_{i_1}\}$ oraz $Y_0 = (-\infty, f(a_{i_1})) \cap \mathbb{Q}$;
- $X_j = \{a \in A \mid a_{i_j} < a < a_{i_{j+1}}\}$ oraz $Y_j = (f(a_{i_j}), f(a_{i_{j+1}})) \cap \mathbb{Q}$, dla $j \in \{1, \ldots, n-1\}$;
- $X_n = \{a \in A \mid a_{i_n} < a\}$ oraz $Y_n = (f(a_{i_n}), \infty) \cap \mathbb{Q}$.

Element a_n, dla którego chcemy określić wartość $f(a_n)$, musi należeć do jednego ze zbiorów $X_0, X_1, \ldots, X_n$, powiedzmy do X_ℓ. Nazwijmy go *przedziałem krytycznym dla* n. Elementy zbioru Y_ℓ nazwiemy zaś liczbami *dozwolonymi dla* n. Aby zachodził warunek $(*)$, wystarczy, aby $f(a_n)$ było dozwolone dla n. Niech więc $f(a_n) = q_m$, gdzie $m = \min\{k \in \mathbb{N} \mid q_k \in Y_\ell\}$.

Załóżmy teraz, że A jest gęsty i nie ma końców. Wtedy określona wyżej funkcja f jest izomorfizmem porządków. Wystarczy w tym celu sprawdzić, że f jest surjekcją.

Przypuśćmy więc, że tak nie jest i niech $m = \min\{k \mid q_k \notin \mathrm{Rg}(f)\}$. Liczby q_j dla $j < m$, dzielą zbiór $\mathbb{Q}$ na $m + 1$ przedziałów, a do jednego z nich należy q_m. Przypuśćmy, że jest to przedział (q_l, q_r). (W przypadku, gdy jest to przedział niewłaściwy, dowód jest podobny.) Mamy więc $l, r < m$ oraz $q_j \notin (q_l, q_r)$ dla $j < m$. Ponadto $q_l, q_r \in \mathrm{Rg}(f)$, czyli $q_l = f(a_p)$ i $q_r = f(a_s)$ dla pewnych p, s. Niech $d = \min\{k \mid a_p < a_k < a_s\}$ i niech $f(a_d) = q_x$. Ponieważ funkcja f zachowuje porządek i jest injekcją, więc na pewno $q_x \in (q_l, q_r)$, skąd mamy $x > m$.

Przedział krytyczny dla d jest wyznaczony przez jedną lub dwie spośród liczb $a_0, a_1, \ldots, a_{d-1}$, z których żadna nie należy do zbioru $C = \{a \in A \mid a_p < a < a_s\}$. Zatem zbiór C jest zawarty w przedziale krytycznym, a wszystkie liczby z przedziału (q_l, q_r), w tym q_m, są dozwolone dla d. Tu otrzymujemy sprzeczność, bo liczbą dozwoloną dla d o najmniejszym numerze jest q_x, a przecież $x > m$. □

Ćwiczenie 14.4 *Udowodnić, że zbiory* $\{\frac{m}{2^n} \mid m \in \mathbb{Z}, n \in \mathbb{N}\}$ *i* $\{m + n\sqrt{2} \mid m, n \in \mathbb{Z}\}$ *uporządkowane przez zwykłą relację* $\leq$*, są izomorficzne.*

Definicja 14.5

- Powiemy, że zbiór liniowo uporządkowany jest *ciągły*, gdy każdy jego niepusty podzbiór ograniczony z góry ma kres górny. (Taką własność ma na przykład zbiór $\mathbb{R}$ liczb rzeczywistych.)
- Podzbiór A zbioru liniowo uporządkowanego B jest *gęsty w* B, gdy zachodzi warunek $\forall a, b \in B(a < b \to \exists c \in A(a < c < b))$.

Uporządkowanie liczb rzeczywistych jest równie unikalne jak uporządkowanie liczb wymiernych.

Twierdzenie 14.6 *Jeśli liniowo uporządkowany zbiór B (bez końców) jest ciągły i ma przeliczalny podzbiór gęsty, to $B \approx \mathbb{R}$.*

Dowód: Niech W będzie przeliczalnym podzbiorem gęstym w B. Zauważmy, że W nie może mieć elementu pierwszego ani ostatniego, a więc na mocy twierdzenia 14.3(2) jest izomorficzny ze zbiorem liczb wymiernych $\mathbb{Q}$ (uporządkowanym tak jak zwykle). Niech $f : W \xrightarrow[\text{na}]{1-1} \mathbb{Q}$ będzie odpowiednim izomorfizmem.

Dla $b \in B$ przez $b{\downarrow}$ oznaczmy zbiór $\{w \in W \mid w < b\}$. Rozpatrzmy przekształcenie $F : B \to \mathbb{R}$, dane wzorem $F(b) = \sup f(b{\downarrow})$. Przekształcenie to zachowuje porządek i jest różnowartościowe, bo jeśli $x < y$ to $x < b_1 < b_2 < y$ dla pewnych $b_1, b_2 \in W$. Wtedy dla dowolnego $d \in x{\downarrow}$ mamy $f(d) < f(b_1)$, skąd $F(x) \leq f(b_1) < f(b_2) \leq F(y)$.

Pozostaje wykazać, że F jest na $\mathbb{R}$. Dla $r \in \mathbb{R}$, niech $B_r = \{w \in W \mid f(w) < r\}$ (zauważmy, że $B_r = \{f^{-1}(q) \mid q < r\}$) i niech $b_r = \sup B_r$. Wówczas

$$w < b_r \text{ wtedy i tylko wtedy, gdy } f(w) < r,$$

dla $w \in W$. Istotnie, gdyby $w < b_r$ i $f(w) \geq r$ to w byłoby ograniczeniem górnym zbioru B_r a wtedy $w \geq b_r$. Na odwrót, jeśli $f(w) < r$, to $f(w) < q < r$, dla pewnego $q \in \mathbb{Q}$, skąd wynika $w < f^{-1}(q) \leq b_r$.

A więc $F(b_r) = \sup f(b_r{\downarrow}) = \sup\{f(w) \mid w \in W \wedge w < b_r\} = \sup\{q \in \mathbb{Q} \mid q < r\} = r$. □

15 Dobre ufundowanie

Definicja 15.1 Niech $\langle A, \leq \rangle$ będzie zbiorem częściowo uporządkowanym. Jeśli każdy niepusty podzbiór zbioru A ma element minimalny, to mówimy, że $\langle A, \leq \rangle$ jest *częściowym dobrym porządkiem*, lub, że A jest *dobrze ufundowany*. Jeśli ponadto porządek $\langle A, \leq \rangle$ jest liniowy, to jest to *dobry porządek*. (Wtedy każdy niepusty podzbiór A ma element najmniejszy.)

Fakt 15.2 *Zbiór $\langle A, \leq \rangle$ jest dobrze ufundowany wtedy i tylko wtedy, gdy nie istnieje w nim ciąg malejący, tj. taki podzbiór $\{a_i \mid i \in \mathbb{N}\}$, że $a_{i+1} < a_i$ dla dowolnego i.*

Dowód: ($\Rightarrow$) Gdyby taki istniał, to by nie miał elementu minimalnego.

($\Leftarrow$) Weźmy niepusty podzbiór $B \subseteq A$ i przypuśćmy, że B nie ma elementu minimalnego. Skoro B jest niepusty to ma jakiś element b_0. On oczywiście nie jest minimalny, więc jest takie $b_1 \in B$, że $b_1 < b_0$. I tak dalej: przez indukcję określamy ciąg malejący $b_0 > b_1 > b_2 > \cdots$ □

115

Przykład 15.3

- Zbiór $\mathbb{N}$ jest dobrze uporządkowany.
- Zbiory z przykładu 14.2 są dobrze uporządkowane przez zwykły porządek w $\mathbb{R}$.
- Zbiory $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ nie są dobrze uporządkowane.
- Relacja porządku prefiksowego jest dobrym ufundowaniem zbioru A^*. Relacja porządku sufiksowego też.
- Niech $r = \{\langle \ell,\ n :: \ell \rangle \mid n \in \mathbb{N} \wedge \ell \in \mathbf{list}\}$ i niech $\sqsubseteq$ będzie domknięciem przechodnim relacji $r \cup \mathrm{id}_{\mathbf{list}}$. Wtedy $\sqsubseteq$ jest dobrym ufundowaniem typu **list**.
- Jeśli w A są dwa elementy a, b, takie że $a < b$, to porządek leksykograficzny $\preceq$, wyznaczony przez $\leq$, nie jest dobrym ufundowaniem zbioru A^*. (Zbiór $\{a^n b \mid n \in \mathbb{N}\}$ nie ma elementu minimalnego.)

Zauważmy, że definicja dobrego ufundowania dla list oparta jest na takim samym schemacie jak porządek prefiksowy dla słów (por. ćwiczenie 7.3). Jest to przechodnio-zwrotne domknięcie relacji „bezpośredniego następnika". W podobny sposób można dobrze ufundować inne typy indukcyjne.

Drzewa

Definicja 15.4 Podzbiór B zbioru częściowo uporządkowanego A nazywamy *odcinkiem początkowym* w A, gdy

$$\forall x, y \in A\, (x \in B \wedge y \leq x \to y \in B).$$

Szczególny przypadek odcinka początkowego to odcinek wyznaczony przez element $x \in A$:

$$\mathcal{O}_A(x) = \{y \in A \mid y < x\}.$$

Uwaga: nierówność w definicji $\mathcal{O}_A(x)$ jest ostra, tj. $x \notin \mathcal{O}_A(x)$. Jeśli wiadomo o jaki zbiór chodzi, to zamiast $\mathcal{O}_A(x)$ piszemy po prostu $\mathcal{O}(x)$.

Definicja 15.5 Jeśli w zbiorze częściowo uporządkowanym mamy $a < b$, ale dla żadnego c nie zachodzi $a < c < b$, to mówimy, że a jest *bezpośrednim poprzednikiem* b, i że b jest *bezpośrednim następnikiem* a.

Następująca definicja uogólnia pojęcie drzewa binarnego, o którym była mowa w rozdziale 7.

Definicja 15.6 Zbiór częściowo uporządkowany $\langle T, \leq \rangle$ nazywamy *drzewem*, gdy spełnia on następujące warunki:

1) Istnieje element najmniejszy.

2) Każdy odcinek postaci $\mathcal{O}_T(x)$ jest skończonym[24] łańcuchem.

Jeśli łańcuch $\mathcal{O}_T(x)$ ma n elementów, to powiemy, że x jest wierzchołkiem o *wysokości* n. Element najmniejszy, nazywany *korzeniem*, ma wysokość zerową.

Niech A będzie jakimś alfabetem (niekoniecznie skończonym). Niepusty podzbiór T zbioru A^* nazywamy *drzewem słów* (nad A), gdy jest on odcinkiem początkowym w $\langle A^*, \subseteq \rangle$, czyli gdy spełniony jest warunek

$$\forall w, u \in A^* \, (w \cdot u \in T \to w \in T).$$

Na przykład następujący zbiór jest drzewem słów nad alfabetem $\{0, 1\}$:

$$\{\varepsilon, 0, 1, 00, 01, 10, 11, 000, 001, 011, 101, 110, 111, 0010, 0011, 1010, 1101\}.$$

Przedstawiamy go tak jak na rysunku 3.[25]

Twierdzenie 15.7 *Każde drzewo jest izomorficzne z pewnym drzewem słów.*

Dowód: Niech $\langle T, \leq \rangle$ będzie drzewem i niech $\perp$ będzie korzeniem tego drzewa. Dla $a \in T$, przez S_a oznaczymy zbiór wszystkich bezpośrednich następników a. Weźmy dowolny alfabet A spełniający warunek $\overline{\overline{A}} \geq \overline{\overline{S_a}}$, dla dowolnego $a \in T$. Istnieją wtedy funkcje $\xi_a : S_a \xrightarrow{1-1} A$.

Funkcję $f : T \xrightarrow{1-1} A^*$ określimy przez indukcję ze względu na wysokość argumentu, w ten sposób, aby spełniony był warunek:

$$a \leq b \quad \Leftrightarrow \quad f(a) \subseteq f(b), \tag{*}$$

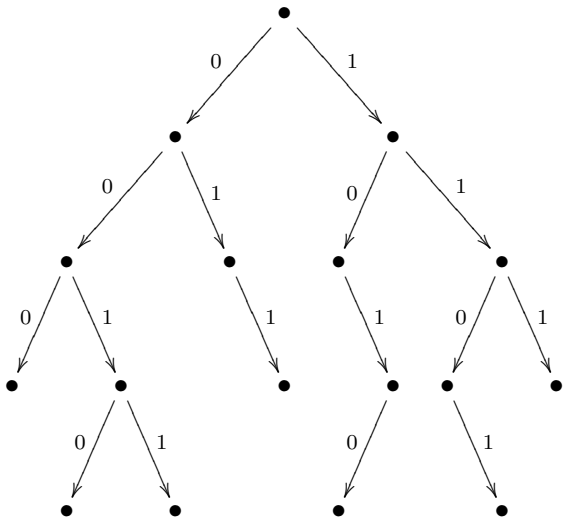
Zaczynamy od $f(\perp) = \varepsilon$. Jeśli $f(a)$ jest określone dla jakiegoś $a \in T$, oraz b jest bezpośrednim następnikiem a, to przyjmujemy

$$f(b) = f(a) \cdot \xi_a(b).$$

Dowód warunku (*) jest przez łatwą indukcję ze względu na wysokość b. □

[24] Czasami drzewem nazywa się każdy porządek, który ma element najmniejszy i w którym wszystkie zbiory $\mathcal{O}(x)$ są dobrze uporządkowane (ale niekoniecznie skończone).

[25] Jak wiadomo, drzewa rosną zwykle z góry na dół.



Rysunek 3: Drzewo

Definicja 15.8

1. *Gałęzią* w drzewie T nazywamy dowolny ciąg postaci $\varepsilon = a_0, a_1, a_2, \ldots$ (skończony lub nieskończony) gdzie każde a_{i+1} jest bezpośrednim następnikiem a_i.
2. Mówimy, że T jest drzewem o skończonym rozgałęzieniu, jeśli każdy element T ma skończenie wiele bezpośrednich następników.

Twierdzenie 15.9 (Lemat Königa) *Jeśli T jest nieskończonym drzewem o skończonym rozgałęzieniu to w T jest gałąź nieskończona.*

Dowód: Dla $a \in T$ niech $T_a = \{b \in T \mid a \leq b\}$. Przez indukcję konstruujemy nieskończoną gałąź $\varepsilon = a_0, a_1, a_2, \ldots$ w ten sposób, aby dla każdego i zbiór T_{a_i} był nieskończony. Krok bazowy jest poprawny, bo $T_\varepsilon = T$. Jeśli teraz T_{a_n} jest zbiorem nieskończonym, oraz wierzchołek a_n ma tylko skończenie wiele bezpośrednich następników $b_1, \ldots, b_k$, to zauważmy, że $T_{a_n} = \{a_n\} \cup T_{b_1} \cup \cdots \cup T_{b_k}$, więc któryś ze zbiorów $T_{b_j}, \ldots, T_{b_k}$ musi być nieskończony, powiedzmy T_{b_j}. Jako a_{n+1} możemy więc przyjąć b_j. □

Lemat Königa ma rozmaite zastosowania. Często używamy go, aby pokazać, że pewne obliczenia muszą się zakończyć w ograniczonym czasie. Oto przykład:

Definicja 15.10 Relacja $\to$ w zbiorze A ma własność *silnej normalizacji* (SN) wtedy i tylko wtedy, gdy nie istnieje nieskończony ciąg postaci $a_0 \to a_1 \to a_2 \to \cdots$

Fakt 15.11 *Relacja $\to$ w zbiorze A ma własność* SN *wtedy i tylko wtedy, gdy $(\to \cap \mathbf{1}_A) = \varnothing$ oraz relacja $\twoheadleftarrow$ jest dobrym ufundowaniem.*[26]

[26]Patrz def. 3.9. Symbol $\leftarrow$ (odp. $\twoheadleftarrow$) oznacza oczywiście relację odwrotną do $\to$ (odp. $\twoheadrightarrow$).

Dowód: Ćwiczenie. (Tylko antysymetria nie jest oczywista.) □

Fakt 15.12 *Załóżmy, że relacja $\to$ w zbiorze A ma własność* SN *oraz dla dowolnego $a \in A$, zbiór $S_a = \{b \in A \mid a \to b\}$ jest skończony. Wówczas dla dowolnego $a \in A$ istnieje taka liczba n, że każdy ciąg postaci $a = a_0 \to a_1 \to a_2 \to \cdots \to a_k$ spełnia warunek $k \leq n$.*

Dowód: Ustalmy $a \in A$ i niech $T \subseteq A^*$ będzie zbiorem wszystkich słów postaci $a_0 a_1 \ldots a_k$, gdzie $a_0 = a$ oraz $a_i \to a_{i+1}$ dla $i < k$. Zbiór T z porządkiem prefiksowym jest drzewem o skończonym rozgałęzieniu, a zatem teza wynika z lematu Königa. □

Następny przykład dotyczy problemu znanego stąd, że jego algorytmiczne rozwiązanie jest w ogólnym przypadku niemożliwe. Przypuśćmy, że dany jest skończony zbiór K (dozwolonych rodzajów kafelków). Na zbiorze K mamy określone relacje zgodności poziomej r i pionowej s. Jeśli $M \subseteq \mathbb{Z} \times \mathbb{Z}$, to mówimy, że funkcja $f : M \to K$ jest *pokryciem* zbioru M, gdy

$$\langle f(x,y), f(x+1,y)\rangle \in r \qquad \langle f(x,y), f(x,y+1)\rangle \in s$$

dla wszystkich x, y dla których odpowiednie punkty leżą w zbiorze M. Mówiąc o pokryciu zbioru $M \subseteq \mathbb{R} \times \mathbb{R}$ mamy na myśli pokrycie dla $M \cap (\mathbb{Z} \times \mathbb{Z})$.

Fakt 15.13 *Jeśli istnieje pokrycie każdego kwadratu to istnieje pokrycie całej płaszczyzny.*

Dowód: Niech $W_n = \{p \in \mathbb{Z} \mid -n < p < n\}$, gdzie $n \in \mathbb{N}$ i niech

$$T = \{f \mid f \text{ jest pokryciem } W_n^2 \text{ dla pewnego } n \in \mathbb{N}\}.$$

Zbiór T uporządkowany przez inkluzję jest drzewem o skończonym rozgałęzieniu. Istotnie, każde pokrycie kwadratu W_n o boku $2n-1$ ma co najwyżej $(\overline{\overline{K}})^{8n}$ rozszerzeń do pokrycia kwadratu W_{n+1}. Drzewo T jest nieskończone, bo istnieją pokrycia dowolnie wielkich kwadratów, a zatem ma nieskończoną gałąź $\varnothing \subseteq f_1 \subseteq f_2 \subseteq f_3 \subseteq \ldots$, gdzie każde f_n jest pokryciem W_n^2. Suma wszystkich funkcji f_n stanowi pokrycie całej płaszczyzny. □

Indukcja

Zasada indukcji, którą znamy dla liczb naturalnych i innych typów indukcyjnych, uogólnia się łatwo na dowolne zbiory dobrze ufundowane. Tę uogólnioną zasadę indukcji nazywamy czasem *indukcją strukturalną* lub *noetherowską* (od nazwiska Emmy Noether).

Fakt 15.14 (Zasada indukcji) *Niech $\langle A, \leq\rangle$ będzie dobrze ufundowany i niech $P \subseteq A$. Załóżmy, że dla dowolnego $a \in A$ zachodzi implikacja:*

$$\mathcal{O}_A(a) \subseteq P \;\Rightarrow\; a \in P.$$

Wtedy $P = A$.

Dowód: Przypuśćmy, że $P \neq A$. Zbiór $A - P$ jest wtedy niepusty i ma element minimalny a. Z minimalności mamy jednak $\mathcal{O}_A(a) \subseteq P$, więc $a \in P$. □

Analogicznie uogólniamy schemat definiowania przez indukcję. Jeśli $\langle A, \leq \rangle$ jest dobrze ufundowany, to możemy definiować funkcję $f : A \to B$, korzystając z dowolnych wartości $f(b)$ dla $b < a$ przy określaniu $f(a)$. Na przykład ta definicja funkcji $f : \mathbb{N} - \{0, 1\} \to \mathbb{N}$

$$f(n) = \begin{cases} n, & \text{jeśli } n \text{ jest pierwsze;} \\ f(m) + f(k), & \text{jeśli } n = mk, \end{cases}$$

jest przez indukcję ze względu na dobrze ufundowaną relację podzielności.

Następująca definicja jest nam potrzebna do przykładu dowodu przez indukcję noetherowską.

Definicja 15.15 Niech $\to$ będzie relacją binarną w zbiorze A. Piszemy $a \downarrow b$ gdy istnieje takie c, że $a \twoheadrightarrow c \twoheadleftarrow b$. Mówimy, że $\to$ ma *własność Churcha-Rossera* (CR), gdy dla dowolnych $a, b, c \in A$:

$$\textit{jeśli} \quad b \twoheadleftarrow a \twoheadrightarrow c \quad \textit{to} \quad b \downarrow c.$$

Relacja $\to$ ma *słabą własność Churcha-Rossera* (WCR), gdy dla dowolnych $a, b, c \in A$:

$$\textit{jeśli} \quad b \leftarrow a \rightarrow c \quad \textit{to} \quad b \downarrow c.$$

Zauważmy, że własność CR nie wynika z WCR. Najprostszy przykład jest chyba taki:

$$\bullet \longleftarrow \bullet \longleftrightarrow \bullet \longrightarrow \bullet$$

Fakt 15.16 (Lemat Newmana) *Relacja o własnościach* WCR *i* SN *ma też własność* CR.

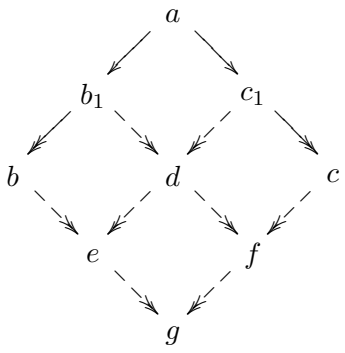
Dowód: Załóżmy, że relacja $\to$ w zbiorze A ma własności WCR i SN. Ponieważ zbiór $\langle A, \twoheadleftarrow \rangle$ jest dobrze ufundowany, więc możemy zastosować indukcję ze względu na porządek $\twoheadleftarrow$. Udowodnimy, że każdy element a ma własność:

„Dla dowolnych b, c, jeśli $b \twoheadleftarrow a \twoheadrightarrow c$, to $b \downarrow c$."

Jeśli $a = b$ lub $a = c$ to teza jest oczywista. Załóżmy więc (zob. rysunek 4), że

$$b \twoheadleftarrow b_1 \leftarrow a \rightarrow c_1 \twoheadrightarrow c.$$

Na mocy WCR jest takie d, że $b_1 \twoheadrightarrow d \twoheadleftarrow c_1$. Z założenia indukcyjnego, zastosowanego do b_1 i c_1 mamy więc $b \twoheadrightarrow e \twoheadleftarrow d \twoheadrightarrow f \twoheadleftarrow c$, dla pewnych e, f. Teraz możemy zastosować założenie indukcyjne dla d. Dostaniemy takie g, że $e \twoheadrightarrow g \twoheadleftarrow f$. Ale wtedy także $b \twoheadrightarrow g \twoheadleftarrow c$. □



Rysunek 4: Dowód lematu Newmana

16 Porządki dobre

Zaczynamy od dwóch nietrywialnych przykładów dobrych porządków.

Fakt 16.1 *Dla dowolnego k, zbiór $\mathbb{N}^k$, złożony z k-krotek liczb naturalnych (słów długości k) jest dobrze uporządkowany przez porządek leksykograficzny (wyznaczony przez zwykłe uporządkowanie zbioru $\mathbb{N}$).*

Dowód: Indukcja ze względu na k. Dla $k = 0, 1$, teza jest oczywista. Załóżmy więc, że zbiór $\mathbb{N}^k$ jest dobrze uporządkowany i niech $B \subseteq \mathbb{N}^{k+1}$ będzie niepusty. Przyjmijmy:

- $b = \min\{n \mid \exists w(n \cdot w \in B)\}$;
- $B' = \{w \in \mathbb{N}^k \mid b \cdot w \in B\}$.

Zbiór B' jest niepustym podzbiorem $\mathbb{N}^k$, ma więc element najmniejszy w. Słowo bw jest wtedy najmniejszym elementem B. □

Czasami wygodne jest pojęcie „zbioru z powtórzeniami", czyli *multizbioru*. Formalnie multizbiory definiujemy jako funkcje. Na przykład multizbiór $\{1, 2, 2, 3, 4, 4, 4\}$ to taka funkcja M, że $M(1) = M(3) = 1$, $M(2) = 2$ i $M(4) = 3$. Dla $x \neq 1, 2, 3, 4$ przyjmujemy $M(x) = 0$.

Definicja 16.2 *Multizbiorem* nad A nazywamy dowolną funkcję $M : A \to \mathbb{N}$.

W stosunku do multizbiorów używamy notacji teoriomnogościowej, pamiętając, że nie należy jej w tym przypadku rozumieć dosłownie. Na przykład piszemy $a \in M$ gdy $M(a) > 0$ oraz $M \subseteq N$ gdy $M(a) \leq N(a)$ dla wszystkich $a \in A$. Możemy też określić działania na multizbiorach, przyjmując

$$(M \cup N)(a) = M(a) + N(a), \text{ oraz } (M - N)(a) = \max\{0, M(a) - N(a)\},$$

dla dowolnego $a \in A$. Powiemy, że multizbiór jest *skończony*, gdy skończony jest zbiór $\mathrm{Rg}(M) = \{a \in A \mid a \in M\}$.

Niech teraz M, N będą skończonymi multizbiorami nad $\mathbb{N}$. Piszemy $M \to N$, gdy istnieją takie a, N', że $N = (M - \{a\}) \cup N'$, i przy tym $a \in M$ oraz $a > b$ dla wszystkich $b \in N'$. Jeśli $\mathrm{Rg}(M) \subseteq \{0, 1, \ldots, n-1\}$, to M można przedstawić krotką $c_M = \langle M(n-1), \ldots, M(0)\rangle \in \mathbb{N}^n$.

Fakt 16.3 *Relacja $\to$ w zbiorze $\mathcal{M}$ wszystkich skończonych multizbiorów nad $\mathbb{N}$ ma własność silnej normalizacji.*

Dowód: Przypuśćmy, że mamy nieskończony ciąg skończonych multizbiorów

$$M_0 \to M_1 \to M_2 \to \cdots$$

i niech $k = 1 + \max\{n \mid n \in M_0\}$. Nietrudno zauważyć, że we wszystkich multizbiorach M_i występują tylko liczby mniejsze od k. Informację o M_i możemy więc przedstawić w postaci k-krotki $w_i = \langle M_i(k-1), \ldots, M_i(0)\rangle \in \mathbb{N}^k$. Na przykład, jeśli $k = 4$, oraz $M_i = \{0, 0, 2, 3, 3, 3\}$,

to $w_i = \langle 3, 1, 0, 2\rangle$. Zauważmy, że krotki c_i tworzą ciąg malejący ze względu na porządek leksykograficzny w zbiorze $\mathbb{N}^k$:

$$c_0 \succ c_1 \succ c_2 \succ \cdots$$

A więc z faktu 16.1 otrzymujemy sprzeczność. □

Wniosek 16.4 *Zbiór $\mathcal{M}$ wszystkich skończonych multizbiorów nad $\mathbb{N}$ jest dobrze uporządkowany przez relację $\twoheadleftarrow$.*

Dowód: Z faktu 16.3 łatwo wynika dobre ufundowanie. Sprawdzenie, że porządek jest liniowy, pozostawiamy jako ćwiczenie. □

Własności dobrych porządków

Lemat 16.5 *Jeśli A jest zbiorem dobrze uporządkowanym, to każdy właściwy odcinek początkowy w A jest postaci $\mathcal{O}_A(x)$.*

Dowód: Niech B będzie właściwym odcinkiem początkowym w A i niech x będzie najmniejszym elementem zbioru $A - B$. Wówczas $B = \mathcal{O}_A(x)$. Rzeczywiście:

- Jeżeli $b \in B$ to $b < x$, bo inaczej $x \leq b$ i byłoby $x \in B$. Zatem $b \in \mathcal{O}_A(x)$.
- Jeżeli $b \in \mathcal{O}_A(x)$, to $b < x$, więc $b \notin A - B$, czyli $b \in B$. □

Lemat 16.6 *Jeśli A jest zbiorem dobrze uporządkowanym, to A nie jest izomorficzny z żadnym swoim właściwym odcinkiem początkowym.*

Dowód: Przypuśćmy, że to nieprawda i niech $x = \min\{y \in A \mid A \approx \mathcal{O}_A(y)\}$. Załóżmy, że $f : A \to \mathcal{O}_A(x)$ jest izomorfizmem. Wtedy f obcięte do odcinka $\mathcal{O}_A(x)$ też jest izomorfizmem, a mianowicie izomorfizmem odcinków $\mathcal{O}_A(x)$ i $\mathcal{O}_A(f(x))$. Stąd $A \approx \mathcal{O}_A(f(x))$, a przy tym $f(x) < x$, co jest sprzeczne z minimalnością x. □

Morał: Różne odcinki początkowe zbioru dobrze uporządkowanego nie są izomorficzne.

Lemat 16.7 *Niech A i B będą dobrymi porządkami i niech*

$$\forall x{\in}A\, \exists y{\in}B\, (\mathcal{O}_A(x) \approx \mathcal{O}_B(y)).$$

Wtedy A jest izomorficzny z pewnym odcinkiem początkowym w B (być może niewłaściwym).

Dowód:* Niech $\Phi = \{\langle x, y\rangle \in A \times B \mid \mathcal{O}_A(x) \approx \mathcal{O}_B(y)\}$. Udowodnimy, że dla dowolnych $\langle x, y\rangle, \langle x', y'\rangle \in \Phi$ zachodzi równoważność:

$$x < x' \quad \Leftrightarrow \quad y < y' \tag{*}$$

($\Rightarrow$) Przypuśćmy, że $x < x'$ ale $y \geq y'$. Niech $f : \mathcal{O}_A(x) \xrightarrow[\text{na}]{1-1} \mathcal{O}_B(y)$ będzie izomorfizmem. Ponieważ $\mathcal{O}_B(y') \subseteq \mathcal{O}_B(y)$ więc odcinek $\mathcal{O}_B(y')$ jest izomorficzny z odcinkiem $\mathcal{O}_A(f^{-1}(y'))$.

Oznacza to jednak, że $\mathcal{O}_A(x') \approx \mathcal{O}_A(f^{-1}(y'))$. Ale $f^{-1}(y') < x'$, bo $f^{-1}(y') \in \mathcal{O}_A(x)$, więc mamy sprzeczność z Lematem 16.6: zbiór $\mathcal{O}_A(x')$ jest izomorficzny ze swoim właściwym odcinkiem początkowym.

Część ($\Leftarrow$) warunku (*) można udowodnić podobnie.

Z warunku (*) wynika, że jeśli dla $x \in A$ przyjmiemy

$$f(x) = \imath y{\in}B.\, \mathcal{O}_A(x) \approx \mathcal{O}_B(y),$$

to $f : A \xrightarrow{1-1} B$ oraz $A \approx f(A)$. Pozostaje zauważyć, że $f(A)$ jest odcinkiem początkowym w B. Ale jeśli $y \in f(A)$ oraz $y' \leq y$, to odcinek $\mathcal{O}_B(y')$ jest izomorficzny z przeciwobrazem $f^{-1}(\mathcal{O}_B(y'))$, który jest odcinkiem początkowym w A. Stąd $y' \in f(A)$ (por. Lemat 16.5). □

Twierdzenie 16.8 *Jeśli A i B są dobrze uporządkowane, to jeden z nich jest izomorficzny z odcinkiem początkowym drugiego.*

Dowód:* Przypuśćmy, że B nie jest izomorficzny z żadnym właściwym odcinkiem początkowym zbioru A. Przez indukcję ze względu na uporządkowanie zbioru A pokażemy:

$$\forall x{\in}A\, \exists y{\in}B\, (\mathcal{O}_A(x) \approx \mathcal{O}_B(y)) \qquad (**)$$

Niech $x \in A$ i przypuśćmy, że każdy odcinek $\mathcal{O}_A(x')$, gdzie $x' < x$ jest izomorficzny z pewnym $\mathcal{O}_B(y')$. Z lematu 16.7 wnioskujemy, że $\mathcal{O}_A(x)$ jest izomorficzne z pewnym odcinkiem początkowym zbioru B. Nie może to być cały zbiór B, bo założyliśmy, że B nie jest izomorficzny z odcinkami właściwymi w A. A zatem $\mathcal{O}_A(x) \approx \mathcal{O}_B(y)$ dla pewnego y.

Stosując jeszcze raz Lemat 16.7 otrzymujemy, że A jest izomorficzny z jakimś odcinkiem początkowym zbioru B (możliwe, że z całym B). □

Z powyższego wynika, że uporządkowanie dobre jest pojęciem bardzo jednoznacznym. Dwa dobre porządki albo są izomorficzne, albo jeden z nich jest „dłuższy". Innych różnic między dobrymi porządkami nie ma.

Teraz jeszcze definicja, która za chwilę będzie przydatna.

Definicja 16.9 Mówimy, że element a zbioru dobrze uporządkowanego jest *graniczny*, gdy nie jest bezpośrednim następnikiem innego elementu. W przeciwnym razie element a nazywamy *niegranicznym*.

Liczby porządkowe

Zbiorom dobrze uporządkowanym przypisujemy *liczby porządkowe*, podobnie jak dowolnym zbiorom przypisujemy liczby kardynalne. Umawiamy się mianowicie, że liczby porządkowe dwóch zbiorów są równe wtedy i tylko wtedy, gdy są to zbiory izomorficzne. Ponadto, przyjmujemy, że liczba porządkowa odcinka jest zawsze *mniejsza lub równa* liczbie porządkowej całego zbioru. (Na mocy twierdzenia 16.8 dwie liczby porządkowe są więc zawsze porównywalne.) Jeśli α jest liczbą porządkową zbioru A to często mówi się, że A jest *typu porządkowego* α. Liczba porządkowa skończonego dobrego porządku to po prostu liczba jego elementów, a liczbę porządkową zbioru $\langle \mathbb{N}, \leq \rangle$ oznaczamy przez ω. Dalsze liczby porządkowe otrzymujemy przez działania arytmetyczne:

Definicja 16.10 Niech α, β będą odpowiednio liczbami porządkowymi zbiorów A i B. Wówczas $\alpha + \beta$ oznacza liczbę porządkową zbioru $A \oplus B$ uporządkowanego tak:

$$\langle x\rangle_i \leq \langle y\rangle_j \qquad \text{wtedy i tylko wtedy, gdy} \qquad i < j, \text{ lub } i = j \text{ oraz } x \leq y.$$

Natomiast $\alpha \cdot \beta$ to liczba porządkowa zbioru $A \times B$ uporządkowanego „antyleksykograficznie":

$$\langle a, b\rangle \leq \langle a', b'\rangle \qquad \text{wtedy i tylko wtedy, gdy} \qquad b < b', \text{ lub } b = b' \text{ oraz } a \leq a'.$$

Porządek typu $\alpha + \beta$ jest utworzony z dwóch łańcuchów: pierwszy typu α, drugi typu β. Zauważmy, że $n + \omega = \omega$ dla dowolnego $n \in \mathbb{N}$, ale $\omega + n \neq \omega$.

Przykład 16.11 W przykładzie 15.3 zbiór A jest typu ω, zbiór A' typu $\omega + 1$, a zbiór A'' typu $\omega + 2$. Natomiast zbiór $\{1 - \frac{1}{n} \mid n \in \mathbb{N} - \{0\}\} \cup \{2 - \frac{1}{n} \mid n \in \mathbb{N} - \{0\}\}$ jest typu $\omega + \omega$.

Porządek typu $\alpha \cdot \beta$ można sobie wyobrażać tak: weźmy porządek typu β i w miejsce każdego elementu wstawmy nowy egzemplarz porządku typu α. Widzimy, że tak określone działanie nie jest przemienne: mamy bowiem $2 \cdot \omega = \omega$ ale $\omega \cdot 2 = \omega + \omega$.

Przykład 16.12 Zbiór B z przykładu 15.3 jest typu $\omega \cdot \omega$.

Oczywiście zamiast $\omega \cdot \omega$ napiszemy ω^2 i ogólnie przyjmiemy, że $\omega^n = \omega \cdots \omega$ (n razy). Dalej można określić liczbę ω^ω jako najmniejszą liczbę porządkową, większą od wszystkich ω^n, liczbę $\omega^{\omega+1}$ jako $\omega^\omega \cdot \omega$, itd.

Przykład 16.13 Zbiór $\mathcal{M}$ wszystkich skończonych multizbiorów nad $\mathbb{N}$ jest typu ω^ω.

Uwaga: Liczba kardynalna $\aleph_0^{\aleph_0}$ jest nieprzeliczalna. Ale liczba porządkowa ω^ω jest typem zbioru przeliczalnego.

Twierdzenie o dobrym uporządkowaniu

Poniższe twierdzenie znacznie ułatwia dowody wielu faktów, pozwala bowiem na postępowanie przez indukcję. Trzeba jednak pamiętać o jego niekonstruktywnym charakterze. Wynika z niego np. że istnieje relacja dobrze porządkująca zbiór liczb rzeczywistych, ale nie wynika, jaka ta relacja naprawdę jest.

Twierdzenie 16.14 (Zermelo) *Każdy zbiór można dobrze uporządkować.*

Dowód:* Niech A będzie dowolnym zbiorem i niech Φ będzie funkcją wyboru dla rodziny $\mathsf{P}(A) - \{\varnothing\}$. Powiemy, że zbiór uporządkowany $\langle D, \leq\rangle$ jest *fajny*, gdy $D \subseteq A$, oraz

$$\forall x \in D\,(x = \Phi(A - \mathcal{O}_D(x))).$$

Część 1: Pokażemy najpierw, że jeśli $\langle D_1, \leq_1\rangle$ i $\langle D_2, \leq_2\rangle$ są fajne, to jeden z nich jest (dosłownie) odcinkiem początkowym drugiego.

Dla ustalenia uwagi, załóżmy, że D_2 nie jest właściwym podzbiorem D_1. Przez indukcję ze względu na porządek $\leq_1$ dowodzimy, że dla dowolnego $x \in D_1$:

1) $x \in D_2$;

2) $\mathcal{O}_{D_1}(x) = \mathcal{O}_{D_2}(x)$.

Przypuśćmy, że wszystkie elementy odcinka $\mathcal{O}_{D_1}(x)$ spełniają powyższe warunki. Jeśli x jest graniczny, to mamy $\mathcal{O}_{D_1}(x) = \bigcup\{\mathcal{O}_{D_1}(y) \mid y <_1 x\}$. Z założenia indukcyjnego jest to suma odcinków początkowych w D_2, a więc $\mathcal{O}_{D_1}(x)$ też jest odcinkiem początkowym w D_2. Jeśli x jest niegraniczny, to mamy natomiast $\mathcal{O}_{D_1}(x) = \mathcal{O}_{D_1}(x') \cup \{x'\} = \mathcal{O}_{D_2}(x') \cup \{x'\}$, dla odpowiedniego x'. (Skorzystaliśmy tu z założenia indukcyjnego o x'.) Zbiór $\mathcal{O}_{D_2}(x') \cup \{x'\}$ jest oczywiście odcinkiem początkowym w D_2.

A więc $\mathcal{O}_{D_1}(x)$ w każdym przypadku jest odcinkiem początkowym w D_2. Jest to odcinek właściwy (bo inaczej $D_2 = \mathcal{O}_{D_1}(x) \subsetneq D_1$) czyli mamy $\mathcal{O}_{D_1}(x) = \mathcal{O}_{D_2}(y)$, dla pewnego y. Ale oba zbiory D_1 i D_2 są fajne, więc $x = \Phi(A - \mathcal{O}_{D_1}(x)) = \Phi(A - \mathcal{O}_{D_2}(y)) = y$, a stąd od razu wynika (1) i (2).

Ponieważ warunki (1) i (2) zachodzą dla wszystkich elementów zbioru D_1, więc $D_1 \subseteq D_2$. Musimy jeszcze sprawdzić, że D_1 jest odcinkiem początkowym w D_2. Niech więc $x \in D_1$ oraz $y <_2 x$ i $y \in D_2$. Wtedy $y \in \mathcal{O}_{D_2}(x) = \mathcal{O}_{D_1}(x)$, w szczególności $y \in D_1$. To kończy część 1 naszego dowodu.

Morał: Jeśli $\langle D_1, \leq_1\rangle$ i $\langle D_2, \leq_2\rangle$ są fajne, to warunki $a \leq_1 b$ i $a \leq_2 b$ są równoważne, jesli tylko $a, b \in D_1 \cap D_2$.

Część 2: Następna obserwacja jest taka: suma wszystkich zbiorów fajnych jest fajna. Niech F oznacza tę sumę. Uporządkowanie $\leq_F$ zbioru F można okreslić jako (dosłownie) sumę uporządkowań wszystkich zbiorów fajnych. Sprawdźmy, czy to jest dobre uporządkowanie.

- Zwrotność: Jeśli $a \in F$ to $a \in D$ dla pewnego fajnego $\langle D, \leq_D\rangle$. Wtedy $a \leq_D a$, więc także $a \leq_F a$.

- Antysymetria: Niech $a \leq_F b$ i $b \leq_F a$. To znaczy, że $a \leq_{D_1} b$ i $b \leq_{D_2} a$, dla pewnych fajnych $\langle D_1, \leq_{D_1}\rangle$ i $\langle D_2, \leq_{D_2}\rangle$. Ale jeden z tych zbiorów jest odcinkiem początkowym drugiego, co oznacza, że tak naprawdę zachodzi też np. $a \leq_{D_2} b$. A więc $a = b$.

- Przechodniość: Niech $a \leq_F b$ i $b \leq_F c$. Są więc takie fajne porządki $\langle D_1, \leq_{D_1}\rangle$ i $\langle D_2, \leq_{D_2}\rangle$, że $a \leq_{D_1} b$ i $b \leq_{D_2} c$. Jeden z nich (niech będzie to np. D_1) jest odcinkiem początkowym drugiego, mamy więc $a \leq_{D_2} b$ i $b \leq_{D_2} c$, skąd wnioskujemy $a \leq_{D_2} c$ i wreszcie $a \leq_F c$.

- Spójność: Niech $a, b \in F$. Wtedy $a \in D_1$ i $b \in D_2$ dla pewnych fajnych D_1 i D_2. Jeśli na przykład $D_1 \subseteq D_2$ to elementy a i b są porównywalne w D_2, a więc i w F.

- Dobroć: Rozpatrzmy dowolny niepusty podzbiór $B \subseteq F$. Niech a będzie dowolnym jego elementem i niech D będzie takim fajnym zbiorem, że $a \in D$. Podzbiór $B \cap D$ zbioru D jest niepusty, ma więc element najmniejszy b. Jest to także najmniejszy element zbioru B ze względu na porządek $\leq_F$. Istotnie, niech $c \in B$. Wtedy albo $c \geq_F a \geq_F b$, albo $c \leq_F a$. W tym drugim przypadku $c \in D$, więc także $c \geq_F b$.

Przyjemność sprawdzenia, że porządek $\langle F, \leq_F\rangle$ jest fajny, pozostawiamy czytelnikowi.

Część 3: Zbiór F jest identyczny ze zbiorem A. Rzeczywiście, przypuśćmy, że $F \neq A$, i niech $a = \Phi(A - F)$. Uporządkowanie zbioru F można teraz rozszerzyć do uporządkowania zbioru $F_1 = F \cup \{a\}$, przyjmując, że a jest elementem największym. Tak uporządkowany zbiór F_1 jest fajny, ale nie jest zawarty w sumie wszystkich zbiorów fajnych i mamy sprzeczność.

Ostatecznie otrzymujemy, że $\langle A, \leq_F \rangle$ jest zbiorem fajnym, w szczególności jest dobrze uporządkowany. □

Z twierdzenia 16.14 wynika istotna własność liczb kardynalnych:

Wniosek 16.15 *Dla dowolnych zbiorów A i B zachodzi $\overline{\overline{A}} \leq \overline{\overline{B}}$ lub $\overline{\overline{B}} \leq \overline{\overline{A}}$.*

Dowód: Zbiory A i B można dobrze uporządkować, a wtedy jeden z nich jest izomorficzny z odcinkiem początkowym drugiego. □

Możemy teraz udowodnić twierdzenie 12.13.

Wniosek 16.16 (Lemat Kuratowskiego-Zorna) *Niech $\langle A, \leq \rangle$ będzie porządkiem częściowym, spełniającym następujący warunek:*

Każdy łańcuch ma w A ograniczenie górne

Wtedy w A istnieje element maksymalny.

Dowód:* Niech $\preceq$ będzie relacją dobrze porządkującą zbiór A. Bez straty ogólności można założyć, że $\langle A, \preceq \rangle$ nie ma elementu ostatniego (ćwiczenie).

Dla dowolnego $a \in A$ określimy przez indukcję pewien zbiór $Ł_a$, w ten sposób, że:

a) $Ł_a \subseteq \{x \in A \mid x \prec a\}$;

b) $Ł_a$ jest łańcuchem ze względu na porządek $\leq$.

Zakładając, że $Ł_b$ jest już określone dla wszystkich $b \prec a$, definiujemy $Ł_a = \bigcup\{Ł_b \mid b \prec a\}$, gdy a jest elementem granicznym. Jeśli natomiast a jest bezpośrednim następnikiem pewnego b, to przyjmujemy:

$$Ł_a = \begin{cases} Ł_b \cup \{b\}, & \text{jeśli } Ł_b \cup \{b\} \text{ jest łańcuchem;} \\ Ł_b, & \text{w przeciwnym przypadku.} \end{cases}$$

Nietrudno sprawdzić, że warunki (a) i (b) są spełnione, i że suma $Ł = \bigcup\{Ł_a \mid a \in A\}$ jest też łańcuchem ze względu na $\leq$. Niech c będzie ograniczeniem górnym łańcucha $Ł$. Twierdzimy, że c jest elementem maksymalnym ze względu na $\leq$.

Istotnie, jeśli $c \leq a$, to a jest porównywalne z każdym elementem zbioru $Ł$, tym bardziej z każdym elementem zbioru $Ł_a$. Wtedy jednak $a \in Ł_b$, gdzie b jest bezpośrednim następnikiem a ze względu na $\preceq$. (Taki bezpośredni następnik istnieje, bo założyliśmy, że elementu ostatniego nie ma.) Ostatecznie wnioskujemy, że $a \in Ł$, czyli $a \leq c$. □

Uwaga*: Dowody twierdzenia Zermeli i lematu Kuratowskiego-Zorna mają charakter niekonstruktywny, bo opierają się istotnie na pewniku wyboru. Ten pierwszy nie podaje definicji dobrego porządku, a jedynie uzasadnia jego istnienie. A ten drugi też nie wskazuje elementu maksymalnego. W istocie, oba te twierdzenia są równoważne pewnikowi wyboru.

17 Klasyczny rachunek zdań

Języka logiki symbolicznej używaliśmy dotąd jako sposobu ścisłego zapisywania faktów i definicji matematycznych wyrażanych w języku polskim. Teraz zobaczymy jak systemy *logiki formalnej* modelują sposoby wnioskowania matematycznego metodami samej matematyki. Najwięcej uwagi poświęcimy *rachunkowi zdań*. Rachunek zdań to formalny model języka, w którym występują spójniki zdaniowe. Wyrażenia tego języka nazywamy *formułami zdaniowymi*. Definiujemy je przez indukcję:

Definicja 17.1

- *Symbole zdaniowe* (zwykle $p, q, r, \ldots$), nazywane też *zmiennymi zdaniowymi*, oraz znaki $\bot$ i $\top$ są formułami zdaniowymi.
- Jeśli napis α jest formułą zdaniową, to także napis $\neg\alpha$ jest formułą zdaniową.
- Jeśli α i β są formułami zdaniowymi to napisy $(\alpha \to \beta)$, $(\alpha \leftrightarrow \beta)$, $(\alpha \vee \beta)$ i $(\alpha \wedge \beta)$ też są formułami zdaniowymi.

Uwaga: Dla pełnej jednoznaczności składni nasze formuły są w pełni nawiasowane. W praktyce wiele nawiasów pomijamy, stosując zwykłe priorytety.

Znaczenie formuł

Wartość logiczna formuły zdaniowej to zawsze 0 (fałsz) lub 1 (prawda). Aby ustalić tę wartość trzeba jednak najpierw znać wartości symboli zdaniowych.

Definicja 17.2 *Interpretacja zdaniowa* (także: *wartościowanie zdaniowe*) to funkcja ϱ, która każdemu symbolowi zdaniowemu p przypisuje wartość logiczną $\varrho(p)$ równą 0 lub 1. *Wartość* formuły zdaniowej α przy interpretacji ϱ oznaczamy przez $[\![\alpha]\!]_\varrho$ i określamy przez indukcję:

- $[\![\bot]\!]_\varrho = 0$ oraz $[\![\top]\!]_\varrho = 1$;
- $[\![p]\!]_\varrho = \varrho(p)$, gdy p jest symbolem zdaniowym;
- $[\![\neg\alpha]\!]_\varrho = 1 - [\![\alpha]\!]_\varrho$;
- $[\![\alpha \vee \beta]\!]_\varrho = \max\{[\![\alpha]\!]_\varrho, [\![\beta]\!]_\varrho\}$;
- $[\![\alpha \wedge \beta]\!]_\varrho = \min\{[\![\alpha]\!]_\varrho, [\![\beta]\!]_\varrho\}$;
- $[\![\alpha \to \beta]\!]_\varrho = 0$, gdy $[\![\alpha]\!]_\varrho = 1$ i $[\![\beta]\!]_\varrho = 0$;
- $[\![\alpha \to \beta]\!]_\varrho = 1$, w przeciwnym przypadku;
- $[\![\alpha \leftrightarrow \beta]\!]_\varrho = 1$, gdy $[\![\alpha]\!]_\varrho = [\![\beta]\!]_\varrho$;
- $[\![\alpha \leftrightarrow \beta]\!]_\varrho = 0$, w przeciwnym przypadku.

Łatwo można zauważyć, że $[\![\alpha \to \beta]\!]_\varrho = \max\{[\![\beta]\!]_\varrho, 1 - [\![\alpha]\!]_\varrho\}$, czyli $[\![\alpha \to \beta]\!]_\varrho = [\![\neg\alpha \vee \beta]\!]_\varrho$, dla dowolnego ϱ. A zatem zamiast formuły $\alpha \to \beta$ moglibyśmy z równym powodzeniem używać wyrażenia $\neg\alpha \vee \beta$, lub też odwrotnie: zamiast alternatywy $\alpha \vee \beta$ pisać $\neg\alpha \to \beta$. Podobnie $\alpha \leftrightarrow \beta$ znaczy to samo co $(\alpha \to \beta) \wedge (\beta \to \alpha)$. Nasz wybór spójników nie jest więc „oszczędny", w istocie w logice klasycznej wystarczy używać np. alternatywy i negacji.

Notacja i terminologia: Jeśli $[\![\varphi]\!]_\varrho = 1$ to piszemy też $\varrho \models \varphi$ lub $\models \varphi[\varrho]$ i mówimy, że φ jest *spełniona* przez interpretację ϱ. Napis $\Gamma \models \varphi$ oznacza, że każda interpretacja spełniająca wszystkie formuły z Γ spełnia także formułę φ. Mówimy wtedy, że φ jest *semantyczną konsekwencją* zbioru Γ. Jeśli Γ jest puste, to zamiast $\Gamma \models \varphi$ piszemy po prostu $\models \varphi$. Oznacza to, że formuła φ jest spełniona przez każdą interpretację. Na koniec powiedzmy jeszcze, że formułami *równoważnymi* nazywamy takie formuły φ i ψ, których wartości przy każdej interpretacji są takie same (tj. takie, że $\models \varphi \leftrightarrow \psi$).

Definicja 17.3 Formuła φ jest *spełnialna*, gdy $\varrho \models \varphi$ zachodzi dla pewnej interpretacji ϱ. Jeśli zaś $\models \varphi$ to mówimy, że φ jest *tautologią* (klasycznego) rachunku zdań. Oczywiście $\neg\varphi$ jest spełnialna wtedy i tylko wtedy, gdy φ nie jest tautologią.

Tautologie rachunku zdań

Przypuśćmy, że każdemu symbolowi zdaniowemu p występującemu w formule φ przyporządkowaliśmy pewną formułę $S(p)$. Przez $S(\varphi)$ oznaczymy wtedy formułę otrzymaną z φ na skutek jednoczesnej zamiany wszystkich symboli zdaniowych p na odpowiadające im formuły $S(p)$. Mówimy, że formuła $S(\varphi)$ jest *instancją* schematu zdaniowego φ. Na przykład, jeśli $S(p) = p \to q$ oraz $S(q) = q \to r$, to $S(p \to q)$ oznacza formułę $(p \to q) \to (q \to r)$.

Fakt 17.4 *Jeżeli φ jest tautologią, to $S(\varphi)$ jest też tautologią.*

Dowód: Dla dowolnej interpretacji zdaniowej ϱ zdefiniujemy inną interpretację $\overline{\varrho}$, przyjmując $\overline{\varrho}(p) = [\![S(p)]\!]_\varrho$, dla każdego symbolu zdaniowego p. Teraz, przez indukcję ze względu na długość formuły ξ udowodnimy, że $[\![\xi]\!]_{\overline{\varrho}} = [\![S(\xi)]\!]_\varrho$. Zaczynamy od przypadku bazowego, kiedy ξ jest atomem (stałą $\bot$, stałą $\top$, lub symbolem zdaniowym). Wtedy teza wynika wprost z definicji. Załóżmy teraz, że ξ jest formułą złożoną, na przykład $\xi = \alpha \vee \beta$. Formuły α i β są krótsze od ξ, spełniają więc założenie indukcyjne. Mamy stąd

$$[\![\xi]\!]_{\overline{\varrho}} = [\![\alpha \vee \beta]\!]_{\overline{\varrho}} = \max\{[\![\alpha]\!]_{\overline{\varrho}}, [\![\beta]\!]_{\overline{\varrho}}\} = \max\{[\![S(\alpha)]\!]_\varrho, [\![S(\beta)]\!]_\varrho\} = [\![S(\alpha) \vee S(\beta)]\!]_\varrho = [\![S(\xi)]\!]_\varrho.$$

Podobnie postępujemy, gdy ξ jest implikacją, koniunkcją czy też negacją. Jeśli teraz φ jest tautologią, to $[\![S(\varphi)]\!]_\varrho = [\![\varphi]\!]_{\overline{\varrho}} = 1$, dla każdego ϱ. □

A zatem każda instancja tautologii jest także tautologią. Nietrudno uogólnić tę obserwację na przypadek, w którym wyrażenia $S(p)$ mogą być dowolnymi stwierdzeniami o określonych (jakkolwiek) wartościach logicznych. A zatem tautologie rachunku zdań (i ogólniej związki postaci $\Gamma \models \varphi$) opisują wzorce poprawnego wnioskowania, w których symbole zdaniowe reprezentują dowolne zdania.

Przyjrzymy się teraz niektórym tautologiom.

Przykład 17.5 Następujące formuły (i wszystkie ich instancje) są tautologiami:

1. $p \to p$;
2. $(p \to q) \to ((q \to r) \to (p \to r))$;
3. $p \to (q \to p)$;
4. $(p \to (q \to r)) \to ((p \to q) \to (p \to r))$;
5. $\bot \to p$ i $p \to \top$ oraz $\top$;
6. $p \to (p \vee q)$, $q \to (p \vee q)$ oraz $(p \to r) \to ((q \to r) \to (p \vee q \to r))$;
7. $(p \wedge q) \to p$, $(p \wedge q) \to q$ oraz $(r \to p) \to ((r \to q) \to (r \to p \wedge q))$;
8. $p \wedge p \leftrightarrow p$ oraz $p \vee p \leftrightarrow p$;
9. $p \wedge (q \wedge r) \leftrightarrow (p \wedge q) \wedge r$ oraz $p \vee (q \vee r) \leftrightarrow (p \vee q) \vee r$;
10. $p \wedge q \leftrightarrow q \wedge p$ oraz $p \vee q \leftrightarrow q \vee p$;
11. $p \wedge (q \vee r) \leftrightarrow (p \wedge q) \vee (p \wedge r)$ oraz $p \vee (q \wedge r) \leftrightarrow (p \vee q) \wedge (p \vee r)$;
12. $p \vee \bot \leftrightarrow p$ oraz $p \wedge \top \leftrightarrow p$.
13. $p \vee \neg p$;
14. $p \to \neg\neg p$ oraz $\neg\neg p \to p$;
15. $(p \to q) \to (\neg q \to \neg p)$ oraz $(\neg q \to \neg p) \to (p \to q)$;
16. $((p \to q) \to p) \to p$;
17. $(\neg p \to p) \to p$;
18. $\neg(p \vee q) \leftrightarrow (\neg p \wedge \neg q)$ oraz $\neg(p \wedge q) \leftrightarrow (\neg p \vee \neg q)$;
19. $(p \to q) \leftrightarrow (\neg p \vee q)$;
20. $((p \leftrightarrow q) \leftrightarrow r) \leftrightarrow (p \leftrightarrow (q \leftrightarrow r))$;
21. $(p \to q) \vee (q \to p)$ oraz $p \vee (p \to q)$.

Nasza lista tautologii zaczyna się od czterech formuł, w których występuje wyłącznie najważniejszy spójnik logiczny – implikacja. Pierwsze dwie wyrażają własności, które można nazwać *zwrotnością* (1) i *przechodniością* (2) implikacji. Formuła (3) mówi, że dodatkowe założenie można zawsze zignorować. Formuła (4) (prawo Frege) wyraża „dystrybutywność" implikacji względem siebie samej i może być odczytywana tak: jeśli r wynika z q w kontekście p, to ten kontekst może być włączony do założenia i konkluzji. Chociaż mniej intuicyjny niż poprzednie trzy, schemat (4) okazuje się często bardzo użyteczny.

Implikację $A \to B$ można interpretować jako stwierdzenie, że warunek A jest „silniejszy"[27] niż B. Tautologie (5) stwierdzają więc, że najsilniejszym możliwym stwierdzeniem jest fałsz, a najsłabszym prawda, która sama jest najprostszą tautologią.

Formuły (6) mówią, że alternatywa $p \vee q$ jest najsilniejszym warunkiem, który wynika zarówno z p jak i z q. Te trzy formuły składają się więc na pewną definicję alternatywy. Dualne formuły (7) wyrażają najważniejszą własność koniunkcji: jest to najsłabszy warunek implikujący oba argumenty. Inne własności koniunkcji i alternatywy, jak idempotentność (8),

[27] Ściślej, silniejszy lub tak samo silny.

łączność (9), przemienność (10) i wzajemna dystrybutywność (11) można wywnioskować ze schematów (3,4,6,7).

Następne na liście są dwie równoważności (12) wyrażające taką myśl: fałsz jest „elementem neutralnym" dla alternatywy (tak jak zero dla dodawania) a prawda dla koniunkcji (tak jak jedynka dla mnożenia). Dlatego $\bot$ możemy uważać za „pustą alternatywę" a $\top$ za „pustą koniunkcję".

Numerem 13 oznaczone jest *prawo wyłączonego środka* stanowiące fundament logiki dwuwartościowej. Na tym fundamencie oparta jest symetria praw podwójnej negacji (14) i kontrapozycji (15), pozwalających na wnioskowanie przez zaprzeczenie.

Formuła (16) (prawo Peirce'a) wyraża podobną myśl przy pomocy samej implikacji. Sens prawa Peirce'a widać najlepiej gdy q jest fałszem; otrzymujemy wtedy prawo Claviusa (17).

Formuła (19) wyraża implikację z pomocą negacji i alternatywy i jest często bardzo przydatna, gdy np. chcemy przekształcić jakąś formułę do wygodniejszej postaci. Podobnie użyteczne są prawa De Morgana (18).

Własnością, która często uchodzi naszej uwagi, jest łączność równoważności (20). W zwiazku z tym, wyrażenie $A \leftrightarrow B \leftrightarrow C$ można z czystym sumieniem pisać bez nawiasów. Zwróćmy jednak uwagę na to, że oznacza ono zupełnie co innego niż stwierdzenie że A, B i C są sobie nawzajem równoważne! Nieco paradoksalny charakter prawa (20) bierze się z tego, że logika klasyczna jest dwuwartościowa. Jeszcze bardziej paradoksalne wydawać się mogą prawa (21), których uzasadnienie także wymaga odwołania się do rachunków zerojedynkowych.

Postać normalna formuł

Definicja 17.6 Każdy symbol zdaniowy i negację symbolu zdaniowego nazwijmy *literałem*. Mówimy, że formuła zdaniowa φ jest w *koniunkcyjnej postaci normalnej*, gdy φ jest koniunkcją alternatyw literałów, tj. wygląda tak:

$$(p_1^1 \vee \cdots \vee p_1^{k_1}) \wedge \cdots \wedge (p_r^1 \vee \cdots \vee p_r^{k_r}),$$

gdzie $r \geq 0$, $k_i \geq 0$, dla $i = 0, \ldots r$, a wszystkie p_j^i są literałami. Przy tym, w myśl przykładu 17.5(12), pustą koniunkcję ($r = 0$) utożsamiamy ze stałą $\top$, a stała $\bot$ to koniunkcja z jednym pustym składnikiem ($r = 1, k_1 = 0$).

Fakt 17.7 *Dla każdej formuły zdaniowej istnieje równoważna jej formuła w koniunkcyjnej postaci normalnej.*

Dowód: Aby przekształcić daną formułę do postaci normalnej najpierw eliminujemy z niej równoważności, zastępując każde podwyrażenie postaci $\alpha \leftrightarrow \beta$ przez $(\alpha \to \beta) \wedge (\beta \to \alpha)$, a następnie usuwamy też implikacje stosując schemat z przykładu 17.5(19).

Teraz „przesuwamy w dół" wszystkie negacje za pomocą praw De Morgana oraz usuwamy zbędne negacje stosując równoważności $\neg\neg\alpha \leftrightarrow \alpha$, $\neg\bot \leftrightarrow \top$ i $\neg\top \leftrightarrow \bot$. Robimy to tak długo, aż otrzymamy formułę, w której negacja stosowana jest tylko do symboli zdaniowych. Mamy więc tylko koniunkcje, alternatywy, stałe logiczne i literały. Nadmiar stałych logicznych eliminujemy korzystając z równoważności $\top \vee \alpha \leftrightarrow \top$, $\top \wedge \alpha \leftrightarrow \alpha$, $\bot \wedge \alpha \leftrightarrow \bot$, $\bot \vee \alpha \leftrightarrow \alpha$.

Pozostaje „uporządkować" wystąpienia spójników $\wedge$ i $\vee$, zastępując podwyrażenia postaci $\varphi \vee (\psi \wedge \vartheta)$ przez $(\varphi \vee \psi) \wedge (\varphi \vee \vartheta)$.

Aby się przekonać, że ta procedura musi się zakończyć, każdej formule (już bez $\rightarrow$ i $\leftrightarrow$) przypiszemy liczbową *wagę*:

- $waga(p) = 2$, gdy p jest symbolem zdaniowym, lub stałą logiczną.
- $waga(\varphi \wedge \psi) = waga(\varphi) + waga(\psi) + 2$;
- $waga(\varphi \vee \psi) = 2 \cdot waga(\varphi) \cdot waga(\psi)$;
- $waga(\neg\varphi) = 2^{waga(\varphi)}$.

Zauważmy teraz, że opisane wyżej przekształcenia zawsze zmniejszają wagę formuły. Na przykład, jeśli $waga(\alpha) = a$ i $waga(\beta) = b$ to podformuła postaci $\neg(\alpha \vee \beta)$ ma wagę 2^{2ab}. Po przesunięciu negacji „w głąb" otrzymujemy formułę $\neg\alpha \wedge \neg\beta$ o wadze $2^a + 2^b + 2 < 2^{2ab}$. Ponieważ przy każdej „poprawce" waga formuły się zmniejsza, nie możemy jej poprawiać w nieskończoność i wreszcie uzyskamy formułę, do której żadnego przekształcenia nie da się już zastosować.

Taka formuła musi być w postaci normalnej. Mozna to uzasadnić przez indukcję ze względu na jej długość. Jeśli jest to zmienna lub stała logiczna (długość 1), to teza jest oczywista. Jeśli jest to koniunkcja $\alpha \wedge \beta$, to jej człony α i β są krótsze, a więc zgodnie z założeniem indukcyjnym muszą być w postaci normalnej. Cała formuła $\alpha \wedge \beta$ oczywiście też jest wtedy w postaci normalnej.

Niech więc nasza formuła będzie alternatywą $\alpha \vee \beta$. Z założenia indukcyjnego, α i β są w postaci normalnej. Gdyby któraś z tych formuł była koniunkcją, np. $\alpha = \gamma \wedge \delta$ to moglibyśmy przekształcić $\alpha \vee \beta$ w formułę $(\gamma \vee \beta) \wedge (\delta \vee \beta)$, a założyliśmy, że to już niemożliwe. Zatem zarówno α jak i β muszą być alternatywami literałów, skąd $\alpha \vee \beta$ jest w postaci normalnej.

Zostaje przypadek negacji $\neg\alpha$. Wtedy α nie może być niczym innym niż zmienną, bo z założenia indukcyjnego musi być w postaci normalnej, a w każdym innym przypadku $\neg\alpha$ podlega dalszym przekształceniom. □

Logika pierwszego rzędu

Logika pierwszego rzędu to taka logika predykatów, w której stosowanie kwantyfikatorów ograniczone jest do obiektów *indywiduowych* tj. do elementów jakiejś dziedziny. Funkcje i relacje określone w takiej dziedzinie uważa się za ustalone. Formalna definicja systemu pierwszego rzędu wymaga więc określenia listy dozwolonych operacji i relacji, a ściślej, listy dozwolonych *symboli* operacji i relacji.

Struktury relacyjne

Definicja 17.1 Przez *sygnaturę* rozumiemy pewien (zwykle skończony) zbiór *symboli relacyjnych i funkcyjnych*, każdy z ustaloną liczbą argumentów. Sygnaturę Σ przedstawić można jako sumę:

$$\Sigma = \bigcup_{n\in\mathbb{N}} \Sigma_n \ \cup \bigcup_{n\in\mathbb{N}} \Sigma_n^R,$$

gdzie Σ_n jest zbiorem n-argumentowych *symboli funkcyjnych*, a Σ_n^R jest zbiorem n-argumentowych *symboli relacyjnych*. O symbolu $f \in \Sigma_n$ powiemy, że jego *arność* to liczba n i napiszemy $ar(f) = n$. Podobnie dla $r \in \Sigma_n^R$.

Uwaga: Jeżeli sygnatura jest skończona, to prawie wszystkie ze zbiorów Σ_n i Σ_n^R są puste.

Oczywiście n-argumentowe symbole funkcyjne posłużą nam jako nazwy n-argumentowych funkcji, a n-argumentowe symbole będą nazwami n-argumentowych relacji. Przypomnijmy, że n-argumentowa relacja w zbiorze A to dowolny podzbiór iloczynu kartezjańskiego A^n. Inaczej mówiąc, jest to pewien zbiór krotek postaci $\langle a_1, \ldots, a_n\rangle$, gdzie $a_1, \ldots, a_n \in A$. Krotkę $\langle a\rangle$ utożsamiamy z elementem a, tj. uważamy, że A^1 to to samo co A. Natomiast zbiór A^0 ma tylko jeden element, mianowicie pustą krotkę $\langle\,\rangle$. Są więc tylko dwie relacje zero-argumentowe: pusta i pełna.

Skoro zbiór A^0 ma tylko jeden element, to funkcja zeroargumentowa $f : A^0 \to A$ przyjmuje tylko jedną wartość. Będziemy więc każdą taką funkcję nazywać *stałą* i utożsamiać z odpowiednim elementem zbioru A.

Definicja 17.2 *Struktura relacyjna* albo *model* sygnatury Σ to niepusty zbiór (zwany *dziedziną* lub *nośnikiem* struktury) wraz z interpretacją symboli sygnaturowych jako funkcji i relacji o odpowiedniej liczbie argumentów. Dokładniej, jeśli $\Sigma = \{f_1, \ldots, f_n, r_1, \ldots, r_m\}$, to *strukturą relacyjną* (*modelem*) nazywamy krotkę postaci:

$$\mathcal{A} = \langle A, f_1^{\mathcal{A}}, \ldots, f_n^{\mathcal{A}}, r_1^{\mathcal{A}}, \ldots, r_m^{\mathcal{A}}\rangle,$$

gdzie dla dowolnego i:

- $f_i^{\mathcal{A}} : A^k \to A$, jeśli $ar(f_i) = k$ (w szczególności $f_i^{\mathcal{A}} \in A$, gdy $k = 0$);
- $r_i^{\mathcal{A}} \subseteq A^k$, gdy $ar(r_i) = k$.

Definicja 17.3 Jeśli sygnatura Σ nie zawiera żadnych symboli relacyjnych, to modele tej sygnatury nazywamy *algebrami*.

Konwencje notacyjne: Nośnik struktury $\mathcal{A}$ oznaczamy przez $|\mathcal{A}|$. Często przyjmujemy domyślnie, że $|\mathcal{A}| = A$, $|\mathcal{B}| = B$ itd., lub po prostu strukturę i jej nośnik oznaczamy tym samym symbolem. Często też tak samo oznacza się symbol relacyjny (funkcyjny) i odpowiadającą mu relację (funkcję).

Przykład 17.4 Niech $\Sigma = \{+, \bullet, 0, 1, \leq\}$, gdzie $\Sigma_2 = \{+, \bullet\}$, $\Sigma_0 = \{0, 1\}$, oraz $\Sigma_2^R = \{\leq\}$. Modelem dla sygnatury Σ jest oczywiście zbiór liczb rzeczywistych ze zwykłymi działaniami i porządkiem:

$$\mathcal{R} = \langle \mathbb{R}, +^{\mathcal{R}}, \bullet^{\mathcal{R}}, 0^{\mathcal{R}}, 1^{\mathcal{R}}, \leq^{\mathcal{R}} \rangle.$$

Ten model zwykle zapiszemy po prostu tak:

$$\mathcal{R} = \langle \mathbb{R}, +, \cdot, 0, 1, \leq \rangle.$$

Inne modele dla tej sygnatury to np. zbiór liczb naturalnych ze zwykłymi działaniami i porządkiem oraz zbiór wszystkich podzbiorów $\mathbb{R}$ z działaniami mnogościowymi i inkluzją:

$$\mathcal{N} = \langle \mathbb{N}, +, \cdot, 0, 1, \leq \rangle;$$
$$\mathcal{P} = \langle \mathsf{P}(\mathbb{R}), \cup, \cap, \varnothing, \mathbb{R}, \subseteq \rangle.$$

Ale modelem jest też taka struktura:

$$\mathcal{A} = \langle \mathbb{R}, \cdot, f, \pi, 0, \varnothing \rangle,$$

gdzie $f(a, b) = 3$ dla dowolnych liczb a, b (symbol "$+$" jest interpretowany jako mnożenie!).

Przykład 17.5 Modelami dla sygnatury $\Sigma = \{\bullet, 1\}$, gdzie $\Sigma_2 = \{\bullet\}$ i $\Sigma_0 = \{1\}$, są na przykład struktury $\langle \mathbb{N}, +, 0 \rangle$ i $\langle \mathbb{N}, \cdot, 1 \rangle$, oraz algebra słów z konkatenacją i słowem pustym: $\langle \{a, b\}^*, \cdot, \varepsilon \rangle$.

Przykład 17.6 Graf zorientowany $G = \langle V, E \rangle$, gdzie V jest zbiorem wierzchołków, oraz $E \subseteq V \times V$ jest zbiorem krawędzi, jest modelem jednoelementowej sygnatury $\Sigma = \Sigma_2^R = \{r\}$.

Termy

Jak powiedzieliśmy na początku, symbole należące do sygnatury mają nam służyć jako *nazwy* pewnych funkcji i relacji. *Znaczenie* tych nazw zależy oczywiście od wybranego modelu. W szczególności symbole stałych są nazwami ustalonych elementów modelu. Inne elementy modelu też mogą być nazwane, na przykład jeśli $f \in \Sigma_2$ i $c \in \Sigma_0$ to *napis*:

$$\text{„}f(f(c, c), f(c, f(c, c)))\text{”},$$

będzie w modelu $\mathcal{A} = \langle A, \mathrm{f}, \mathrm{c} \rangle$ nazwą elementu $\mathrm{f}(\mathrm{f}(\mathrm{c}, \mathrm{c}), \mathrm{f}(\mathrm{c}, \mathrm{f}(\mathrm{c}, \mathrm{c})))$.

W ten sposób można jednak nazywać tylko elementy generowane przez stałe. Aby nazywać dowolne elementy modelu potrzebujemy *zmiennych*. Ustalmy więc pewien zbiór symboli V, rozłączny z sygnaturą, którego elementy (oznaczane $x, y, \ldots$) będziemy nazywali *zmiennymi indywiduowymi*, lub po prostu *zmiennymi*. Zwykle przyjmuje się, że V jest nieskończonym zbiorem przeliczalnym.

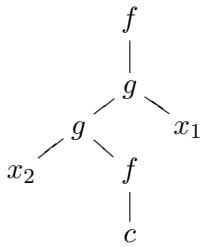
Definicja 17.7 Przez *termy* sygnatury Σ rozumiemy elementy najmniejszego zbioru napisów T_Σ (ozn. też po prostu przez T) spełniającego warunki:

- $V \subseteq T_\Sigma$;
- jeśli $f \in \Sigma_n$ oraz $t_1, \ldots, t_n \in T_\Sigma$ to „$f(t_1, \ldots, t_n)$” $\in T_\Sigma$.

Konwencje notacyjne: Niektóre dwuargumentowe symbole funkcyjne, jak np. „$+$”, „$\cup$” są tradycyjnie pisane w notacji infiksowej, czyli pomiędzy argumentami. Dlatego i my zamiast formalnie poprawnego „$+(2, 2)$” zwykle napiszemy „$2+2$”.

Przykład 17.8 Wyrażenie „$f(g(g(x_2, f(c)), x_1))$” jest termem sygnatury Σ, gdzie $g \in \Sigma_2$, $f \in \Sigma_1$ oraz $c \in \Sigma_0$. Wyrażenie „$(0 + x_1) \bullet 1$” jest termem sygnatury Σ, w której $+, \bullet \in \Sigma_2$ i $0, 1 \in \Sigma_0$.

Często wygodnie jest reprezentować termy za pomocą drzew skończonych, w których liście są etykietowane zmiennymi i stałymi, a wierzchołki wewnętrzne symbolami funkcyjnymi. Oczywiście stopień wyjściowy wierzchołka (liczba dzieci) musi się zgadzać z liczbą argumentów użytego symbolu funkcyjnego. Na przykład term $f(g(g(x_2, f(c)), x_1))$ przedstawiamy jako:



Definicja 17.9 Dla dowolnego termu t, zbiór *zmiennych wolnych* termu t, oznaczany przez $FV(t)$, jest określony przez indukcję:

- $FV(x) = \{x\}$, gdy x jest zmienną;
- $FV(f(t_1, \ldots, t_n)) = FV(t_1) \cup \cdots \cup FV(t_n)$.

Jeśli $X \subseteq V$, to stosujemy takie oznaczenia:

- $T_\Sigma(X) = \{t \in T_\Sigma \mid FV(t) \subseteq X\}$
- $T_\Sigma(n) = T_\Sigma(\{x_1, \ldots, x_n\})$

Zauważmy, że $FV(c) = \varnothing$, gdy c jest symbolem stałej ($c \in \Sigma_0$).

Oczywiście chcemy używać termów jako nazw obiektów indywiduowych (elementów jakiegoś modelu $\mathcal{A}$). To znaczy, że chcemy każdemu termowi przypisać jego *wartość* w modelu $\mathcal{A}$.

Definicja 17.10 *Wartościowaniem* w strukturze $\mathcal{A}$ nazywamy dowolną funkcję $v : V \to |\mathcal{A}|$. Takie wartościowanie v każdemu termowi t przypisuje jego *wartość* oznaczoną $[\![t]\!]_v$, którą definiujemy przez indukcję:

- $[\![x]\!]_v = v(x)$, gdy x jest zmienną;
- $[\![f(t_1, \ldots, t_n)]\!]_v = f^{\mathcal{A}}([\![t_1]\!]_v, \ldots, [\![t_n]\!]_v)$, gdy $f \in \Sigma_n$.

Jeśli v jest wartościowaniem w strukturze $\mathcal{A}$, oraz $a \in |\mathcal{A}|$, to przez $v[x \mapsto a]$ oznaczamy wartościowanie określone tak:

$$v[x \mapsto a](y) = \begin{cases} a, & \text{gdy } y = x; \\ v(y), & \text{w przeciwnym przypadku.} \end{cases}$$

Formuły pierwszego rzędu

Definicja 17.11 *Formuły atomowe* sygnatury Σ są następujące:

- symbole „$\bot$” i „$\top$”;
- napisy postaci „$r(t_1, \ldots, t_n)$”, gdzie $r \in \Sigma_n^R$ oraz $t_1, \ldots, t_n \in \mathcal{T}_\Sigma$;

Definicja 17.12 *Formuły* sygnatury Σ definiujemy jako elementy najmniejszego zbioru $\mathcal{F}_\Sigma$ (ozn. też $\mathcal{F}$) spełniającego warunki:

- formuły atomowe należą do $\mathcal{F}_\Sigma$;
- jeśli $\varphi, \psi \in \mathcal{F}_\Sigma$ to także $(\varphi \to \psi), (\varphi \vee \psi), (\varphi \wedge \psi), (\varphi \leftrightarrow \psi), \neg\varphi \in \mathcal{F}_\Sigma$;
- jeśli $\varphi \in \mathcal{F}_\Sigma$ i $x \in V$ (x jest zmienną indywiduową) to także $(\forall x\varphi), (\exists x\varphi) \in \mathcal{F}_\Sigma$.

Definicja 17.13 Zbiór *zmiennych wolnych* formuły φ, oznaczany przez $FV(\varphi)$, jest określony przez indukcję:

- $FV(r(t_1, \ldots, t_n)) = FV(t_1) \cup \cdots \cup FV(t_n)$;
- $FV(t_1 = t_2) = FV(t_1) \cup FV(t_2)$;
- $FV(\bot) = \varnothing$;
- $FV(\varphi \to \psi) = FV(\varphi \wedge \psi) = FV(\varphi \vee \psi) = FV(\varphi) \cup FV(\psi) = FV(\varphi) \cup FV(\psi)$.
- $FV(\forall x\varphi) = FV(\exists x\varphi) = FV(\varphi) - \{x\}$.

Znaczeniem formuły jest wartość logiczna „prawda” (1) lub „fałsz” (0). Formułom sygnatury Σ możemy przypisywać znaczenia w dowolnej strukturze tej sygnatury, zależnie od wybranego wartościowania termów.

Definicja 17.14 *Wartość* $[\![\varphi]\!]_v$ formuły φ w strukturze $\mathcal{A}$ przy wartościowaniu v definiujemy przez indukcję (ze względu na budowę formuły):

- $[\![\bot]\!]_v = 0$ oraz $[\![\top]\!]_v = 1$;
- $[\![r(t_1, \ldots, t_n)]\!]_v = 1$, gdy $\langle [\![t_1]\!]_v, \ldots, [\![t_n]\!]_v \rangle \in r^{\mathcal{A}}$;
- $[\![r(t_1, \ldots, t_n)]\!]_v = 0$, w przeciwnym przypadku;
- $[\![\neg\alpha]\!]_v = 1 - [\![\alpha]\!]_v$;
- $[\![\alpha \vee \beta]\!]_v = \max\{[\![\alpha]\!]_v, [\![\beta]\!]_v\}$;
- $[\![\alpha \wedge \beta]\!]_v = \min\{[\![\alpha]\!]_v, [\![\beta]\!]_v\}$;
- $[\![\alpha \to \beta]\!]_v = 0$, gdy $[\![\alpha]\!]_v = 1$ i $[\![\beta]\!]_v = 0$;
- $[\![\alpha \to \beta]\!]_v = 1$, w przeciwnym przypadku;
- $[\![\alpha \leftrightarrow \beta]\!]_v = 1 - |[\![\alpha]\!]_v - [\![\beta]\!]_v|$.
- $[\![\forall x\varphi]\!]_v = \min\{[\![\varphi]\!]_{v[x \mapsto a]} \mid a \in |\mathcal{A}|\}$;
- $[\![\exists x\varphi]\!]_v = \max\{[\![\varphi]\!]_{v[x \mapsto a]} \mid a \in |\mathcal{A}|\}$;

Definicja 17.15 Formuła jest *spełnialna (spełnialna w $\mathcal{A}$)* jeśli jest spełniona w pewnym modelu (w modelu $\mathcal{A}$) przez pewne wartościowanie. Zbiór formuł jest *spełnialny* (w $\mathcal{A}$) jeśli wszystkie formuły z tego zbioru są spełnione przez to samo wartościowanie w pewnym modelu (w modelu $\mathcal{A}$).

Formuła φ jest *prawdziwa w $\mathcal{A}$* (piszemy $\mathcal{A} \models \varphi$), jeżeli jest spełniona w $\mathcal{A}$ przez wszystkie wartościowania. Formuła φ jest *prawdziwa* (jest *tautologią*) jeżeli jest prawdziwa w każdym modelu $\mathcal{A}$. Wtedy piszemy po prostu $\models \varphi$.

Tautologie logiki predykatów

Jeśli w tautologii zdaniowej w miejsce symboli zdaniowych wstawimy dowolne formuły, to pozostanie ona zawsze prawdziwa. W tym sensie tautologie zdaniowe są też tautologiami logiki pierwszego rzędu. Ale tautologiami pierwszego rzędu są też na przykład formuły postaci:[28]

22. $\forall x(A(x) \to B(x)) \to (\forall x A(x) \to \forall x B(x))$;
23. $\forall x(A(x) \to B(x)) \to (\exists x A(x) \to \exists x B(x))$;
24. $\neg\forall x A(x) \leftrightarrow \exists x \neg A(x)$;
25. $\neg\exists x A(x) \leftrightarrow \forall x \neg A(x)$;
26. $\forall x(A(x) \wedge B(x)) \leftrightarrow \forall x A(x) \wedge \forall x B(x)$;
27. $\exists x(A(x) \vee B(x)) \leftrightarrow \exists x A(x) \vee \exists x B(x)$;
28. $\forall x \forall y A(x, y) \leftrightarrow \forall y \forall x A(x, y)$;
29. $\exists x \exists y A(x, y) \leftrightarrow \exists y \exists x A(x, y)$;
30. $\exists x \forall y A(x, y) \to \forall y \exists x A(x, y)$;

[28] Przypomnijmy, że dziedzina przebiegana przez zmienne musi być niepusta.

31. $\forall x\, A(x) \to \exists x\, A(x)$.

Jeśli zmienna x nie jest wolna w A, to tautologiami są też formuły:

32. $A \leftrightarrow \exists x A$;
33. $A \leftrightarrow \forall x A$;
34. $\forall x(A \vee B(x)) \leftrightarrow A \vee \forall x B(x)$;
35. $\exists x(A \wedge B(x)) \leftrightarrow A \wedge \exists x B(x)$.

Schematy (24) i (25) nazywamy prawami De Morgana. Tautologie (26) i (27) wskazują na bliski związek kwantyfikatora ogólnego z koniunkcją i kwantyfikatora szczegółowego z alternatywą. Analogiczna rozdzielność kwantyfikatora ogólnego względem alternatywy (34) i kwantyfikatora szczegółowego względem koniunkcji (35) zachodzi pod warunkiem, że zmienna wiązana kwantyfikatorem nie występuje w jednym z członów formuły. Schematy (26–27) i (34–35) możemy nazywać *prawami dystrybutywności* przez analogię do schematu (11). Prawo (34) nazywane też bywa *prawem Grzegorczyka*.

Prawa (28)–(30) charakteryzują możliwości permutowania kwantyfikatorów. Implikacja odwrotna do (30) nie jest tautologią, jako przykład weźmy prawdziwe zdanie $\forall x{:}\mathbb{N}\exists y{:}\mathbb{N}.\, x < y$ (dla każdej liczby naturalnej istnieje liczba od niej większa). Po przestawieniu kwantyfikatorów otrzymamy fałszywe stwierdzenie o istnieniu największej liczby naturalnej.

Dziedzina, którą przebiegają wartości zmiennych jest zawsze niepusta; to założenie powoduje, że formułę (31) trzeba też uznać za tautologię.

Uwaga: Często zakłada się, że znak $=$ zawsze oznacza równość. Przy tej konwencji np. formuła $\forall x \forall y(x = y \to y = x)$ jest uważana za tautologię. Jeśli $=$ może oznaczać jakąkolwiek relację, to oczywiście nie można tak powiedzieć.

Postać normalna formuły: Czasem chcemy by wszystkie kwantyfikatory znajdowały się na początku formuły. Mówimy, że formuła φ jest w *preneksowej postaci normalnej*, gdy

$$\varphi = Q_1 y_1 Q_2 y_2 \ldots Q_n y_n\, \psi,$$

gdzie każde z Q_i to $\forall$ lub $\exists$, a ψ nie zawiera kwantyfikatorów. (Oczywiście n może być zerem.) Stosując równoważności (24–27) i (34–35) możemy każdą formułę przekształcić w równoważną formułę w preneksowej postaci normalnej. Na przykład formuła $\exists y p(y) \to \forall z q(z)$ jest równoważna każdej z następujących formuł:

$$\neg\exists y\, p(y) \vee \forall z\, q(z);$$
$$\forall y\, \neg p(y) \vee \forall z\, q(z);$$
$$\forall y(\neg p(y) \vee \forall z\, q(z));$$
$$\forall y \forall z(\neg p(y) \vee q(z));$$
$$\forall y \forall z(p(y) \to q(z)).$$

Z pomocą praw De Morgana i praw dystrybutywności możemy się czasem łatwo przekonać o tym, że dana formuła jest tautologią. Na przykład z powyższego przykładu wynika od razu, że formuła $(\exists y p(y) \to \forall z q(z)) \to \forall y(p(y) \to q(y))$ jest równoważna tautologii

$$\forall y \forall z(p(y) \to q(z)) \to \forall y(p(y) \to q(y)).$$

18 Dowodzenie twierdzeń

Sprawdzenie, czy dana formuła rachunku zdań jest tautologią, wymaga obliczenia jej wartości przy 2^n różnych interpretacjach, gdzie n jest liczbą zmiennych zdaniowych tej formuły. Dla logiki predykatów nie istnieje w ogóle żaden algorytm sprawdzania czy dana formuła jest tautologią. W obu przypadkach są jednak metody *dowodzenia* pozwalające na uzasadnienie prawdziwości formuły z pomocą pewnego ustalonego systemu reguł wnioskowania. Opiszemy tu nieformalnie jeden z takich systemów, zwany *naturalną dedukcją* i pochodzący od Gerharda Gentzena i Stanisława Jaśkowskiego. Reguły naturalnej dedukcji w dużym stopniu przypominają rzeczywiste sposoby wnioskowania stosowane w matematyce.

Dowód w systemie naturalnej dedukcji polega na wyprowadzaniu kolejnych wniosków z przyjętych założeń. Stosuje się w tym celu *reguły wnioskowania*. Części dowodu, w których wprowadzane są pomocnicze założenia i nazwy umieszczane są w pudełkach Jaśkowskiego. A zatem *dowód* to ciąg formuł i pudełek.

Najprostsza reguła bywa nazywana *aksjomatem naturalnej dedukcji* i stanowi, że każde z założeń może być przywołane w dowolnym miejscu:

Z założenia mamy A.

Powyżej, A może być albo założeniem wolnym (globalnym), przyjętym przed rozpoczęciem dowodu, albo założeniem lokalnym, które obowiązuje w aktualnym pudełku.

Dla każdego operatora logicznego (spójnika lub kwantyfikatora) w systemie naturalnej dedukcji mamy metodę (regułę) *wprowadzania* tego operatora i metodę jego *eliminacji*. Pierwsza pozwala na udowodnienie zdania, w którym dany operator występuje jako główny, druga pokazuje jak można użyć takiego zdania do dowodzenia innych.

Koniunkcja: Koniunkcję można wywnioskować z obu jej składowych. Oto schemat wprowadzania koniunkcji (jeśli tezy A i B już uznaliśmy za prawdziwe, to uznajemy $A \wedge B$):

$$\begin{array}{c} A \\ \vdots \\ B \\ \vdots \\ \text{Ponieważ } A \text{ oraz } B\text{, więc } A \wedge B. \end{array}$$

Schemat eliminacji koniunkcji mówi, że każda składowa jest konsekwencją koniunkcji. Jeśli wcześniej wyprowadziliśmy $A \wedge B$, to możemy wywnioskować A bądź B.

$$\begin{array}{c} A \wedge B \\ \vdots \\ \text{Ponieważ } A \wedge B\text{, więc } A. \end{array} \qquad \begin{array}{c} A \wedge B \\ \vdots \\ \text{Ponieważ } A \wedge B\text{, więc } B. \end{array}$$

Implikacja: Użycie implikacji w dowodzie jest możliwe wtedy, gdy potrafimy też wyprowadzić jej przesłankę. Ten sposób wnioskowania nazywa się *odrywaniem* (*modus ponens*).

$$\begin{array}{c} A \\ \vdots \\ A \to B \\ \vdots \\ \text{Ponieważ } A \text{ oraz } A \to B\text{, więc } B. \end{array}$$

Dowód implikacji $A \to B$ polega na uzasadnieniu B przy założeniu A. Zapiszemy to tak:

Załóżmy A.	*(Cel: B)*
$\vdots$	
Zatem B.	*(Cel osiągnięty)*

Zatem $A \to B$.

Wyprowadzenie B z założenia A zamknęliśmy w pudełku Jaśkowskiego, aby je odseparować od reszty wnioskowania. Założenie A jest bowiem wprowadzone „lokalnie", a w „globalnym" dowodzie (czyli na zewnątrz pudełka) możemy korzystać tylko z konkluzji $A \to B$. Powołanie się na A poza pudełkiem byłoby błędem i jest niedozwolone.

Jako pierwszy przykład udowodnimy tautologię (3).

Załóżmy p	*(Cel 1: $q \to p$)*
Załóżmy q	*(Cel 2: p)*
Z założenia mamy p	*(Cel 2 osiągnięty)*
Zatem $q \to p$	*(Cel 1 osiągnięty)*

Zatem $p \to (q \to p)$

W następnym przykładzie udowodnimy tautologię zdaniową (4). Zrobimy tu pewne uproszczenie: zamiast trzech zagnieżdżonych pudełek narysujemy tylko jedno, wypisując wszystkie potrzebne założenia od razu.

Załóżmy $p \to (q \to r)$, $p \to q$, p.	*(Cel: r)*
Ponieważ p oraz $p \to (q \to r)$, więc $q \to r$.	
Ponieważ p oraz $p \to q$, więc q.	
Ponieważ q oraz $q \to r$, więc r.	

Zatem $(p \to (q \to r)) \to ((p \to q) \to (p \to r))$

Negacja: Kto pamięta, że negacja to właściwie szczególny rodzaj implikacji, ten nie będzie zaskoczony takim schematem wprowadzania negacji:

Załóżmy A	*(Cel: $\bot$)*
$\vdots$	
Zatem $\bot$ (*sprzeczność*).	

Zatem $\neg A$.

Oczywiście eliminacja negacji wygląda tak:

$$\begin{array}{c} A \\ \vdots \\ \neg A \\ \vdots \\ \text{Ponieważ } A \text{ oraz } \neg A\text{, więc } \bot \text{ (sprzeczność).} \end{array}$$

Jako przykład udowodnimy formułę $p \to \neg\neg p$:

Załóżmy p $(Cel:\ \neg\neg p)$

> Załóżmy $\neg p$. $(Cel:\ \bot)$
> Ponieważ p oraz $\neg p$, więc sprzeczność.

Zatem $\neg\neg p$

Zatem $p \to \neg\neg p$

Reguły wprowadzania i eliminacji negacji nie zawsze jednak są wystarczające. Potrzebny jest nam jeszcze schemat wnioskowania przez zaprzeczenie („nie wprost"), np. taki:

> Załóżmy $\neg A$. $(Cel:\ \bot)$
> $\vdots$
> Zatem $\bot$.

Zatem A.

Wnioskowanie przez zaprzeczenie jest niezbędne na przykład w dowodzie formuły $\neg\neg p \to p$:

Załóżmy $\neg\neg p$ $(Cel:\ p)$

> Załóżmy $\neg p$. $(Cel:\ \bot)$
> Ponieważ $\neg p$ oraz $\neg\neg p$, więc sprzeczność.

Zatem p

Zatem $\neg\neg p \to p$

Alternatywa: Alternatywa wynika z każdego ze swoich składników:

$$\begin{array}{c} A \\ \vdots \\ \text{Ponieważ } A\text{, więc } A \vee B. \end{array} \qquad \begin{array}{c} B \\ \vdots \\ \text{Ponieważ } B\text{, więc } A \vee B. \end{array}$$

Sposób użycia alternatywy w dowodzie, to wnioskowanie w rozbiciu na przypadki. Mamy tu dwa „lokalne" założenia obowiązujące tylko wewnątrz swoich pudełek. Oczywiście wcześniej należy udowodnić samą alternatywę.

$$A \vee B$$
$$\vdots$$

Załóżmy A.	*(Cel 1: C)*
$\vdots$	
Zatem C.	

Załóżmy B.	*(Cel 2: C)*
$\vdots$	
Zatem C.	

Ponieważ $A \vee B$, więc C.

Teraz nietrywialne zadanie: jak udowodnić $p \vee \neg p$ (zasadę *tertium non datur*)? Nie wprost.

Załóżmy $\neg(p \vee \neg p)$	*(Cel 1: $\bot$)*
	(Cel 2: $p \vee \neg p$)
	(Cel 3: $\neg p$)
Załóżmy p.	*(Cel 4: $\bot$)*
Ponieważ p, więc $p \vee \neg p$.	
Ponieważ $p \vee \neg p$ oraz $\neg(p \vee \neg p)$, więc sprzeczność.	*(Cel 4 osiągnięty)*
Zatem $\neg p$.	*(Cel 3 osiągnięty)*
Ponieważ $\neg p$, więc $p \vee \neg p$.	*(Cel 2 osiągnięty)*
Ponieważ $p \vee \neg p$ oraz $\neg(p \vee \neg p)$, więc sprzeczność.	*(Cel 1 osiągnięty)*

Zatem $p \vee \neg p$.

Adnotacje o kolejnych celach pomocniczych pokazują taktykę użytą w naszym dowodzie. Aby udowodnić absurdalność założenia $\neg(p \vee \neg p)$, wnioskujemy z niego $p \vee \neg p$, a w tym celu postanowiliśmy wyprowadzić $\neg p$.

Równoważność: Równoważność to dwie implikacje. Wnioskowanie z użyciem równoważności (eliminacja $\leftrightarrow$) jest więc zgodne z schematami:

$$\begin{array}{c} A \\ \vdots \\ A \leftrightarrow B \\ \vdots \end{array}$$

Ponieważ A oraz $A \leftrightarrow B$, więc B.

$$\begin{array}{c} B \\ \vdots \\ A \leftrightarrow B \\ \vdots \end{array}$$

Ponieważ B oraz $A \leftrightarrow B$, więc A.

Natomiast wprowadzenie równoważności polega na dowodzie „w obie strony”:

Załóżmy A.	(*Cel:* B)
⋮	
Zatem B.	

Załóżmy B.	(*Cel:* A)
⋮	
Zatem A.	

Zatem $A \leftrightarrow B$.

Fałsz: Dla stałej $\bot$ nie ma reguły wprowadzania (ostatecznie nikt nie chce udowodnić fałszu).[29] Mamy jednak regułę eliminacji. Pozwala ona na wyprowadzenie z $\bot$ dowolnego wniosku (*ex falso quodlibet*).

$$\begin{array}{c} \bot \\ \vdots \\ \text{Ponieważ } \bot\text{, więc } A. \end{array}$$

Zastosujemy tę regułę w dowodzie formuły $p \to (\neg p \to q)$:

Załóżmy p oraz $\neg p$.	(*Cel 1:* q)
	(*Cel 2:* $\bot$)
Ponieważ p oraz $\neg p$, więc $\bot$.	
Ponieważ $\bot$, więc q.	

Zatem $p \to (\neg p \to q)$

Prawda: Prawda jest mniej ciekawa niż fałsz. Oto schemat wprowadzania prawdy:

Wiadomo, że $\top$. *(I co z tego?)*

Następny przykład to nasze ulubione prawo Peirce'a. Zastosujemy tu zasadę *ex falso quodlibet* i znowu użyjemy metody dowodu przez zaprzeczenie. Uwaga: nie należy jej nadużywać, gdy możliwy jest dowód wprost.

[29] Fałsz (sprzeczność) można udowodnić przez eliminację negacji. Nie jest to jednak wprowadzenie stałej $\bot$, a raczej „odsłonięcie” $\bot$ ukrytego w negacji.

Załóżmy $(p \to q) \to p$ *(Cel 1: p)*

Załóżmy $\neg p$. *(Cel 2: $\bot$)*
(Cel 3: p)
(Cel 4: $p \to q$)

Załóżmy p. *(Cel 5: q)*
(Cel 6: $\bot$)

Ponieważ p oraz $\neg p$, więc $\bot$.
Ponieważ $\bot$, więc q.

Zatem $p \to q$
Ponieważ $p \to q$ oraz $(p \to q) \to p$, więc p.
Ponieważ p i $\neg p$, więc sprzeczność.

Zatem p.

Zatem $((p \to q) \to p) \to p$.

Oczywiście uzasadnianie każdego twierdzenia od samego początku byłoby nieracjonalnie pracochłonne, dlatego w dowodach często powołujemy się na fakty już wcześniej udowodnione. Na przykład w poniższym dowodzie formuły $(p \to q) \to ((\neg p \to q) \to q)$ powołamy się na zasadę wyłączonego środka. Dzięki temu skorzystamy z reguły eliminacji alternatywy (czyli rozważymy dwa przypadki).

Załóżmy $p \to q$ oraz $\neg p \to q$ *(Cel: q)*
Wiadomo, że $p \vee \neg p$

Załóżmy p. *(Przypadek 1)*
Ponieważ p oraz $p \to q$, więc q.

Załóżmy $\neg p$. *(Przypadek 2)*
Ponieważ $\neg p$ oraz $\neg p \to q$, więc q.

Ponieważ $p \vee \neg p$, więc q

Zatem $(p \to q) \to ((\neg p \to q) \to q)$

Kwantyfikator ogólny: Aby udowodnić tezę uniwersalną $\forall x{:}\mathcal{D}.\, A(x)$ należy wywnioskować własność $A(x)$, nie zakładając na temat x niczego poza tym, że należy do $\mathcal{D}$. Dlatego schemat wprowadzania $\forall$ jest taki, jak niżej. (Nazwa x jest „lokalna” i obowiązuje wewnątrz pudełka, w którym została zadeklarowana. Nie może być tam użyta w innym znaczeniu.)

Weźmy dowolne $x : \mathcal{D}$. *(Cel: $A(x)$)*
$\vdots$
Zatem $A(x)$.

Zatem $\forall x{:}\mathcal{D}.\, A(x)$.

Eliminacja $\forall$ polega na zastosowaniu własności ogólnej w szczególnym przypadku:

$$\begin{array}{c} d : \mathcal{D} \\ \vdots \\ \forall x{:}\mathcal{D}.\, A(x) \\ \vdots \\ \text{Ponieważ } \forall x{:}\mathcal{D}.\, A(x) \text{ oraz } d : \mathcal{D}\text{, więc } A(d). \end{array}$$

Jako przykład[30] rozpatrzmy formułę $\forall x{:}\mathcal{D}\, P(x) \wedge \forall x{:}\mathcal{D}(P(x) \to Q) \to Q$, zakładając, że $d \in \mathcal{D}$, i że zmienna x nie występuje w Q.

Załóżmy $\forall x{:}\mathcal{D}.P(x) \wedge \forall x{:}\mathcal{D}(P(x) \to Q)$	(*Cel:* Q)
Ponieważ $\forall x{:}\mathcal{D}.P(x)$, oraz $d : \mathcal{D}$, więc $P(d)$	
Ponieważ $\forall x{:}\mathcal{D}(P(x) \to Q)$, oraz $d : \mathcal{D}$, więc $P(d) \to Q$	
Skoro $P(d) \to Q$ oraz $P(d)$, to Q.	

Zatem $\forall x{:}\mathcal{D}\, P(x) \wedge \forall x{:}\mathcal{D}(P(x) \to Q) \to Q$.

Jeśli dziedzina $\mathcal{D}$ jest domyślna, to zwykle zakładamy, że jest niepusta. Wtedy d nie musi być nazwą konkretnego przedmiotu, ale może też być zmienną. W ten sposób nasza reguła odzwierciedla założenie o niepustości dziedziny.

$$\begin{array}{c} \vdots \\ \forall x.\, A(x) \\ \vdots \\ \text{Ponieważ } \forall x.\, A(x)\text{, więc } A(x). \end{array}$$

Oto przykład użycia tej wersji reguły wprowadzania $\forall$:

Załóżmy $\forall x\, P(x) \wedge \forall x(P(x) \to Q)$	(*Cel:* Q)
Ponieważ $\forall x\, P(x)$, więc $P(x)$	
Ponieważ $\forall x(P(x) \to Q)$, więc $P(x) \to Q$	
Skoro $P(x) \to Q$ oraz $P(x)$, to Q.	

Zatem $\forall x\, P(x) \wedge \forall x(P(x) \to Q) \to Q$.

Kwantyfikator szczegółowy: Wprowadzenie $\exists$ jest możliwe, o ile mamy „świadka”:

$$\begin{array}{c} d : \mathcal{D} \\ \vdots \\ A(d) \\ \vdots \\ \text{Ponieważ } A(d) \text{ oraz } d : \mathcal{D}\text{, więc } \exists x{:}\mathcal{D}.\, A(x) \end{array}$$

„Świadek” d może być dowolnym wyrażeniem, choćby zmienną. Zobaczmy jak można udowodnić zdanie $Q \to \exists x\, (P(x) \to Q)$:

[30] W tym przykładzie pominięto „oczywistą” operację eliminacji koniunkcji.

Załóżmy Q
> Załóżmy $P(x)$.
> Z założenia mamy Q.

Zatem $P(x) \to Q$.
Ponieważ $P(x) \to Q$, więc $\exists x\,(P(x) \to Q)$.
Zatem $Q \to \exists x\,(P(x) \to Q)$.

A oto zasada eliminacji dla kwantyfikatora szczegółowego, określająca jak fakt egzystencjalny może byc wykorzystany w dowodzie. Aby wywnioskować tezę B z założenia $\exists x{:}\mathcal{D}.\, A(x)$ należy udowodnić B zakładając, że x jest typu $\mathcal{D}$ i że zachodzi $A(x)$ ale nie zakładając nic więcej na temat x. Inaczej mówiąc, nieważne jakie jest x, ważne, że istnieje. Oczywiście formuła B nie może zależeć od x (zawierać wolnych wystąpień zmiennej x).

Niech $x : \mathcal{D}$ będzie takie, że $A(x)$ (*Cel:* B)
$\vdots$
Zatem B.

Ponieważ $\exists x{:}\mathcal{D}.\, A(x)$, więc B.

Jako następny przykład udowodnijmy zdanie $\exists x{:}\mathcal{D}\forall y{:}\mathcal{E}\; P(x,y) \to \forall y{:}\mathcal{E}\exists x{:}\mathcal{D}\; P(x,y)$

Załóżmy $\exists x{:}\mathcal{D}\forall y{:}\mathcal{E}\; P(x,y)$ (*Cel:* $\forall y{:}\mathcal{E}\exists x{:}\mathcal{D}\; P(x,y)$)
> Weźmy dowolne $\overline{y} : \mathcal{E}$. (*Cel:* $\exists x{:}\mathcal{D}\; P(x,\overline{y})$)
> (*Użyjemy reguły eliminacji dla* $\exists$)
>> Niech $\overline{x}{:}\mathcal{D}$ będzie takie, że $\forall y{:}\mathcal{E}\; P(\overline{x},y)$. (*Cel:* $\exists x{:}\mathcal{D}\; P(x,\overline{y})$)
>> Ponieważ $\forall y{:}\mathcal{E}\; P(\overline{x},y)$, oraz $\overline{y} : \mathcal{E}$, więc $P(\overline{x},\overline{y})$.
>> Ponieważ $P(\overline{x},\overline{y})$, więc $\exists x{:}\mathcal{D}\; P(x,\overline{y})$.
>
> Ponieważ $\exists x{:}\mathcal{D}\forall y{:}\mathcal{E}\; P(x,y)$, więc $\exists x{:}\mathcal{D}\; P(x,\overline{y})$.

Zatem $\forall y{:}\mathcal{E}\exists x{:}\mathcal{D}\; P(x,y)$.

Zatem $\exists x{:}\mathcal{D}\forall y{:}\mathcal{E}\; P(x,y) \to \forall y{:}\mathcal{E}\exists x{:}\mathcal{D}\; P(x,y)$.

W najmniejszym pudełku mamy dowód tezy $\exists x{:}\mathcal{D}\; P(x,\overline{y})$, otrzymany przy dodatkowym założeniu $\forall y{:}\mathcal{E}\; P(\overline{x},y)$. Nowo wprowadzona zmienna $\overline{x}$, odgrywa tu rolę „świadka" dla tezy egzystencjalnej $\exists x{:}\mathcal{D}\forall y{:}\mathcal{E}\; P(x,y)$. W praktyce matematycznej mamy często do czynienia z takim zabiegiem. Na przykład:

Wiadomo, że istnieją liczby przestępne różne od e i od π. Niech więc ζ będzie taką liczbą . . .

Ostatni przykład to dowód formuły C korzystający z „wolnych" (inaczej „globalnych") założeń:

$$\forall x(P(x) \to C),\;\; B \to \exists y\, P(y),\;\; A \to B,\quad A.$$

Aby wyprowadzić C musimy najpierw udowodnić $\exists y\, P(y)$ i do tej formuły zastosować eliminację $\exists$. Tym razem dowód nie zaczyna się od lokalnych założeń ani deklaracji. Dlatego nie ma zewnętrznego pudełka.

(*Cel 1: C*)
(*Cel 2:* $\exists y\, P(y)$)

Ponieważ A i $A \to B$, więc B.
Ponieważ B i $B \to \exists y\, P(y)$, więc $\exists y\, P(y)$.

(*Użyjemy reguły eliminacji dla* $\exists$)

Niech y będzie takie, że $P(y)$.	(*Cel 2: C*)
Ponieważ $\forall x(P(x) \to C)$, więc $P(y) \to C$.	
Ponieważ $P(y)$ oraz $P(y) \to C$, więc C.	

Ponieważ $\exists y\, P(y)$, więc C

Zasady budowy dowodu

Dowód w naturalnej dedukcji polega na wyprowadzaniu kolejnych wniosków z przyjętych wcześniej założeń. Każdy dowód jest ciągiem skończonym, w którym występować mogą:

- pojedyncze stwierdzenia, tj. założenia i wnioski;
- pudełka zawierające mniejsze dowody.

Pudełka mogą być dowolnie zagnieżdżone. Dowód uważamy za poprawny, jeśli stosuje się do następujących zasad:

- Każdy wniosek jest otrzymany poprzez zastosowanie którejś z reguł wnioskowania. Mogą być przy tym wykorzystane tylko stwierdzenia i nazwy *widoczne* z danego miejsca. W szczególności, zwrotu „*Ponieważ A...*" lub „*Z założenia mamy A...*" wolno użyć tylko wtedy, gdy A jest widoczne.
- W każdym kroku widoczne są tylko stwierdzenia wcześniejsze (już uzasadnione), w tym globalne założenia.
- Zawartość pudełka jest niewidoczna na zewnątrz pudełka.
- Stwierdzenia poprzedzające jakieś pudełko są widoczne wewnątrz tego pudełka. To samo dotyczy wprowadzonych wcześniej nazw zmiennych.
- Nowa zmienna wprowadzona w wewnętrznym pudełku „zasłania" wcześniejsze deklaracje zmiennych o tej samej nazwie i wszystkie stwierdzenia odwołujące się do tej nazwy.

Ćwiczenie 18.1 Metodą naturalnej dedukcji udowodnić tautologie (1–21).

Formalizacja w stylu Gentzena

Pojęcie dowodu można ściśle sformalizować, (mówimy wtedy o *dowodach formalnych*). W logice rozważa się różne systemy dowodzenia, w tym różne warianty naturalnej dedukcji. Nasze dowody z pudełkami Jaśkowskiego są zaledwie pół-formalne, można ten system uściślić wprowadzając bardziej rygorystyczną składnię. Z drugiej strony, zasady naturalnej dedukcji można stosować jako wskazówki dotyczące poprawnej budowy zwykłych dowodów matematycznych pisanych po polsku, czy angielsku. Widzieliśmy już, że logiczny szkielet poprawnego dowodu często w znacznym stopniu przypomina konstrukcje rozważane powyżej. Taki dowód łatwiej

zrozumieć, jeśli przedstawimy sobie jego konstrukcję z pomocą pudełek Jaśkowskiego, zwłaszcza jeśli istotną rolę odgrywają w nim lokalne założenia i zmienne (np. fakt 6.2 i twierdzenie 12.14).

Najczęściej spotykanym ujęciem naturalnej dedukcji jest formalizacja w stylu Gentzena.[31] Rozważamy tu *osądy* postaci $\Gamma \vdash \varphi$, gdzie Γ jest zbiorem formuł (założeń), a φ to formuła (teza). Zamiast „chować" do pudełek lokalne założenia, aktualizujemy na bieżąco dostępny zbiór założeń. Notacja Gentzena jest mniej efektowna od pudełek Jaśkowskiego, ale wygodniejsza wtedy, gdy chcemy analizować dowody w sposób bardziej ścisły. Ponieważ nasz wykład logiki predykatów był i tak wysoce nieformalny, ograniczymy się teraz do rachunku zdań.

Reguły wnioskowania czytamy tak: nad kreską mamy osądy-przesłanki, a pod kreską konkluzję, którą wolno z nich wyprowadzić. Aksjomat (Ax) naturalnej dedukcji nie ma przesłanek, a więc osąd postaci $\Gamma \cup \{\varphi\} \vdash \varphi$ jest zawsze wyprowadzalny. *Dowód formalny* w naturalnej dedukcji, to drzewo etykietowane osądami w ten sposób, że etykieta każdego wierzchołka jest konkluzją pewnej reguły, której przesłanki są etykietami bezpośrednich następników (dzieci) tego wierzchołka.[32] A zatem w liściach dowodu znajdują się trywialne osądy postaci $\Gamma \cup \{\varphi\} \vdash \varphi$. Osąd w korzeniu, to ostateczna konkluzja dowodu. Mamy następujące reguły wnioskowania (litera W oznacza wprowadzanie, a litera E oznacza eliminację):

$$\frac{}{\Gamma \cup \{\varphi\} \vdash \varphi}\,(\mathrm{Ax})$$

$$\frac{\Gamma \vdash A \qquad \Gamma \vdash B}{\Gamma \vdash A \wedge B}\,(\mathrm{W}\wedge) \qquad \frac{\Gamma \vdash A \wedge B}{\Gamma \vdash A}\,(\mathrm{E}\wedge) \qquad \frac{\Gamma \vdash A \wedge B}{\Gamma \vdash B}\,(\mathrm{E}\wedge)$$

$$\frac{\Gamma \vdash A \qquad \Gamma \vdash A \to B}{\Gamma \vdash B}\,(\mathrm{E}\to) \qquad \frac{\Gamma \cup \{A\} \vdash B}{\Gamma \vdash A \to B}\,(\mathrm{W}\to)$$

$$\frac{\Gamma \vdash A}{\Gamma \vdash A \vee B}\,(\mathrm{W}\vee) \qquad \frac{\Gamma \vdash B}{\Gamma \vdash A \vee B}\,(\mathrm{W}\vee)$$

$$\frac{\Gamma \vdash A \vee B \qquad \Gamma \cup \{A\} \vdash C \qquad \Gamma \cup \{B\} \vdash C}{\Gamma \vdash C}\,(\mathrm{E}\vee)$$

$$\frac{\Gamma \cup \{A\} \vdash \bot}{\Gamma \vdash \neg A}\,(\mathrm{W}\neg) \qquad \frac{\Gamma \vdash \neg A \qquad \Gamma \vdash A}{\Gamma \vdash \bot}\,(\mathrm{E}\neg)$$

$$\frac{\Gamma \vdash \bot}{\Gamma \vdash A}\,(\mathrm{E}\bot) \qquad \frac{\Gamma \cup \{\neg A\} \vdash \bot}{\Gamma \vdash A}\,(\mathrm{E}\neg\neg)$$

[31] Nie mylić z tzw. rachunkiem sekwentów Gentzena.

[32] Tym razem, wyjatkowo, należy sobie wyobrażać, że takie drzewa mają korzeń na dole, a liście na górze.

Przykład: Oto dowód osądu $\neg\beta \vdash \neg(\beta \wedge \gamma)$:

$$\dfrac{\dfrac{\neg\beta,\ \beta\wedge\gamma\ \vdash\ \neg\beta \qquad \dfrac{\neg\beta,\ \beta\wedge\gamma\ \vdash\ \beta\wedge\gamma}{\neg\beta,\ \beta\wedge\gamma\ \vdash\ \beta}\,(\mathrm{E}\wedge)}{\neg\beta,\beta\wedge\gamma\ \vdash\ \bot}\,(\mathrm{E}\neg)}{\neg\beta\ \vdash\ \neg(\beta\wedge\gamma)}\,(\mathrm{W}\neg)$$

Notacja i terminologia: Zamiast $\Gamma \cup \{\varphi\}$ zwykle piszemy po prostu Γ, φ. Zamiast $\Gamma \vdash \alpha$, gdzie $\Gamma = \{\beta_1, ..., \beta_n\}$, można napisać $\beta_1, ..., \beta_n \vdash \alpha$. Zamiast $\varnothing \vdash \alpha$ piszemy $\vdash \alpha$.

Przyjmujemy też konfundującą konwencję używania napisu $\Gamma \vdash \alpha$ w dwóch różnych znaczeniach: jako osądu i jako metajęzykowego stwierdzenia „osąd $\Gamma \vdash \alpha$ ma dowód". Użyjemy tej konwencji już w następnej definicji:

Definicja 18.2

- Jeśli $\vdash \alpha$, to formułę α nazywamy *twierdzeniem naturalnej dedukcji*.
- Jeśli $\Gamma \vdash \bot$, to mówimy, że zbiór Γ jest *sprzeczny*.

Poprawność i pełność

Oczywiście to co udowodnimy, powinno być prawdą. Mówimy, że system wnioskowania jest *poprawny* (lub *adekwatny*) jeśli każde jego twierdzenie jest tautologią logiczną.

Naturalna dedukcja jest poprawnym systemem dowodzenia, co intuicyjnie uzasadniamy tym, że reguły wnioskowania pozwalają z prawdziwych stwierdzeń wyprowadzić tylko prawdziwe wnioski. Inaczej: jeśli wszystkie założenia użyte w jakimś dowodzie są spełnione, to także wszystkie wyprowadzone w nim wnioski są spełnione.

Spróbujmy tę obserwację nieco uściślić.

Twierdzenie 18.3 (o poprawności) *System naturalnej dedukcji dla rachunku zdań jest poprawny: dla dowolnych* Γ *i* α*, jeśli* $\Gamma \vdash \alpha$ *, to* $\Gamma \models \alpha$.

Dowód: Dowód przeprowadzimy przez indukcję ze względu na rozmiar... dowodu[33] osądu $\Gamma \vdash \alpha$, tj. liczbę wierzchołków drzewa.

Niech więc $\Gamma \vdash \alpha$ i niech ϱ będzie taką interpretacją, że $[\![\gamma]\!]_\varrho = 1$ dla wszystkich $\gamma \in \Gamma$. Mamy wykazać, że $[\![\alpha]\!]_\varrho = 1$.

Jeśli dowód $\Gamma \vdash \alpha$ polega po prostu na przywołaniu aksjomatu (Ax), to sprawa jest oczywista bo wtedy $\alpha \in \Gamma$. Jako pierwszy przykład rozpatrzmy reguły związane z implikacją.

[33]Paradoksalność tego zdania jest pozorna i wynika z dwojakiego znaczenia słowa „dowód". Raz chodzi tu o dowód rozumiany jako *uzasadnienie* pewnej własności w języku polskim. A drugi raz – o sformalizowany dowód w systemie naturalnej dedukcji, inaczej o *wyprowadzenie* osądu $\Gamma \vdash \alpha$.

Przypuśćmy, że $\Gamma \vdash \alpha$ wywnioskowano przez eliminację implikacji (E$\rightarrow$). To znaczy, że etykiety dwóch poprzedników korzenia mają postać $\Gamma \vdash \beta$ i $\Gamma \vdash \beta \rightarrow \alpha$. Poddrzewa wyrastające z tych dwóch poprzedników stanowią dowody odpowiednich osądów, ale mają mniej wierzchołków. Z założenia indukcyjnego wynika więc, że $\Gamma \models \beta$ oraz $\Gamma \models \beta \rightarrow \alpha$. Stąd $[\![\beta]\!]_\varrho = [\![\beta \rightarrow \alpha]\!]_\varrho = 1$, a to oznacza, że także $[\![\alpha]\!]_\varrho = 1$.

Rozpatrzmy teraz dowód α kończący się zastosowaniem reguły wprowadzania implikacji. W tym przypadku $\alpha = \alpha_1 \rightarrow \alpha_2$, a korzeń dowodu ma jeden poprzednik o etykiecie $\Gamma, \alpha_1 \vdash \alpha_2$. Mamy więc dowód osądu $\Gamma, \alpha_1 \vdash \alpha_2$ o mniejszej liczbie wierzchołków i możemy użyć założenia indukcyjnego. Otrzymujemy więc $\Gamma \cup \{\alpha_1\} \models \alpha_2$. Jeśli teraz $[\![\alpha_1]\!]_\varrho = 1$ to także $[\![\alpha_2]\!]_\varrho = 1$ skąd $[\![\alpha]\!]_\varrho = 1$. A jeśli $[\![\alpha_1]\!]_\varrho = 0$, to także $[\![\alpha]\!]_\varrho = 1$.

W podobny sposób postępujemy w pozostałych przypadkach (gdy dowód $\Gamma \vdash \alpha$ kończy się jakąś inną regułą). Na przykład, jeśli był to dowód przez zaprzeczenie, to istnieje krótszy dowód dla $\Gamma, \neg\alpha \vdash \bot$, do którego stosujemy założenie indukcyjne. Szczegóły pozostawiamy czytelnikowi. □

Ciekawszą (ale trudniejszą) własnością systemu wnioskowania jest pełność. System jest *pełny*, gdy każda tautologia jest jego twierdzeniem.[34] Pełność naturalnej dedukcji dla rachunku zdań uzasadnimy nieformalnie poniżej. W przypadku logiki predykatów problem pełności jest bardziej złożony i wymaga przede wszystkim uściślenia o jaką dokładnie logikę predykatów chodzi. Dla logiki pierwszego rzędu, gdzie kwantyfikatory wiążą tylko zmienne *indywiduowe*, twierdzenie o pełności zachodzi, chociaż jego dowód wymaga innych metod niż użyte poniżej. Jeśli kwantyfikatory mogą wiązać zmienne oznaczające zbiory, funkcje lub relacje, to mamy do czynienia z logiką *wyższego rzędu*. Dla klasycznej logiki wyższego rzędu naturalna dedukcja nie jest pełnym systemem wnioskowania. Co gorsza, żaden skończony system reguł nie jest w tym przypadku pełny.

Pełność rachunku zdań.

Zaczniemy od kilku prostych obserwacji. Ich dowody pozostawiamy jako ćwiczenia.

Lemat 18.4 *Dla dowolnych formuł β i γ zachodzi:*

1. $\beta \vdash \beta$;
2. $\vdash \neg\bot$;
3. $\beta \vdash \neg\neg\beta$;
4. $\neg\beta, \neg\gamma \vdash \neg(\beta \vee \gamma)$;
5. $\beta, \gamma \vdash \beta \wedge \gamma$;
6. $\neg\beta \vdash \neg(\beta \wedge \gamma)$ *oraz* $\neg\gamma \vdash \neg(\beta \wedge \gamma)$;
7. $\beta, \neg\gamma \vdash \neg(\beta \rightarrow \gamma)$;
8. $\neg\beta \vdash \beta \rightarrow \gamma$ *oraz* $\gamma \vdash \beta \rightarrow \gamma$.

Lemat 18.5 *Naturalna dedukcja ma takie własności:*

[34] Zwykle o pełności mówi się tylko w odniesieniu do systemów poprawnych.

1. *Jeśli* $\beta_1, ..., \beta_n \vdash \alpha$, *to także* $\beta_1, ..., \beta_n, \gamma \vdash \alpha$ *(monotoniczność).*

2. *Jeśli* $\beta_1, ..., \beta_n \vdash \alpha$ *oraz* $\beta_1, ..., \beta_n, \alpha \vdash \gamma$ *to* $\beta_1, ..., \beta_n \vdash \gamma$ *(reguła cięcia).*

Lemat 18.6 *Jeśli* $\beta_1, \ldots, \beta_n, \gamma \vdash \alpha$ *oraz* $\beta_1, \ldots, \beta_n, \neg\gamma \vdash \alpha$ *to* $\beta_1, \ldots, \beta_n \vdash \alpha$.

Dla dowolnej interpretacji zdaniowej ϱ i dowolnej formuły α przyjmijmy, że

$$\alpha^\varrho = \begin{cases} \alpha, & \text{jeśli } [\![\alpha]\!]_\varrho = 1; \\ \neg\alpha, & \text{w przeciwnym przypadku.} \end{cases}$$

Na przykład jeśli $\varrho(p) = 0$ i $\varrho(q) = 1$, to $q^\varrho = q$ oraz $((p \to q) \to p)^\varrho = \neg((p \to q) \to p)$.

Lemat 18.7 (Kalmár) *Niech* $p_1, \ldots, p_n$ *będą wszystkimi symbolami zdaniowymi występującymi w formule* α. *Wówczas* $p_1^\varrho, \ldots, p_n^\varrho \vdash \alpha^\varrho$, *dla dowolnej interpretacji* ϱ.

Dowód: Dowód przebiega przez indukcję ze względu na długość formuły α. Jeśli α jest zmienną lub stałą logiczną to teza jest oczywista (patrz lematy 18.4(1,2) i 18.5).

Przypuśćmy, że $\alpha = \beta \to \gamma$. Jeśli $[\![\alpha]\!]_\varrho = 0$ to $[\![\beta]\!]_\varrho = 1$ i $[\![\gamma]\!]_\varrho = 0$. Formuły β i γ są krótsze od α, mamy więc $p_1^\varrho, \ldots, p_n^\varrho \vdash \beta$ oraz $p_1^\varrho, \ldots, p_n^\varrho \vdash \neg\gamma$ z założenia indukcyjnego. Jeśli zaś można udowodnić β i $\neg\gamma$ to można też udowodnić $\neg(\beta \to \gamma)$ (lemat 18.4(7)). Niech więc $[\![\alpha]\!]_\varrho = 1$. Są tu możliwe dwa przypadki: albo $[\![\beta]\!]_\varrho = 0$ albo $[\![\gamma]\!]_\varrho = 1$. W pierwszym przypadku z założenia indukcyjnego wiemy, że $p_1^\varrho, \ldots, p_n^\varrho \vdash \neg\beta$, a w drugim, że $p_1^\varrho, \ldots, p_n^\varrho \vdash \gamma$. W obu przypadkach potrafimy udowodnić $\beta \to \gamma$ (lemat 18.4(8)).

Załóżmy teraz, że $\alpha = \beta \vee \gamma$ i niech $[\![\alpha]\!]_\varrho = 1$. Wtedy jedna z formuł β, γ ma dowód przy założeniach $p_1^\varrho, \ldots, p_n^\varrho$ i stosując zasadę wprowadzania alternatywy od razu dostajemy $p_1^\varrho, \ldots, p_n^\varrho \vdash \beta \vee \gamma$. Jeśli natomiast $[\![\alpha]\!]_\varrho = 0$ to z założenia indukcyjnego wynika, że mamy dowody formuł $\neg\gamma$ i $\neg\beta$. Teraz łatwo jest udowodnić negację alternatywy (lemat 18.4(4)).

Następny przypadek, to $\alpha = \beta \wedge \gamma$. Tu rozumowanie jest podobne i korzysta z lematu 18.4(5,6). Pozostaje jeszcze przypadek negacji, $\alpha = \neg\beta$. Jeśli teraz $[\![\alpha]\!]_\varrho = 1$ to teza wynika natychmiast z założenia indukcyjnego o β, w przeciwnym razie należy z β wyprowadzić podwójną negację $\neg\neg\beta$ (lemat 18.4(3)). □

Twierdzenie 18.8 (o pełności rachunku zdań) *Każda tautologia zdaniowa ma dowód w systemie naturalnej dedukcji.*

Dowód: Niech α będzie tautologią zdaniową. Wtedy $\alpha^\varrho = \alpha$ dla dowolnej interpretacji ϱ. Niech $p_1, \ldots, p_n$ będą wszystkimi symbolami zdaniowymi występującymi w formule α. Udowodnimy, że dla dowolnego $m \leq n$ i dowolnego ϱ zachodzi $p_1^\varrho, \ldots, p_m^\varrho \vdash \alpha$. Przyjmując $m = 0$ otrzymamy $\vdash \alpha$.

Dowód przebiega przez indukcję ze względu na $n - m$. Dla $m = n$ teza wynika z lematu Kalmára. Przypuśćmy więc, że $m < n$. Z założenia indukcyjnego mamy $p_1^\varrho, \ldots, p_m^\varrho, p_{m+1} \vdash \alpha$ oraz $p_1^\varrho, \ldots, p_m^\varrho, \neg p_{m+1} \vdash \alpha$. Ponieważ jednak $p_{m+1} \vee \neg p_{m+1}$ jest twierdzeniem, więc nietrudno jest pokazać, że $p_1^\varrho, \ldots, p_m^\varrho \vdash \alpha$ (lemat 18.6). □

Podziękowania

Za liczne uwagi, które pomogły usunąć z tych notatek rozmaite błędy, dziękuję Pani Karolinie Sołtys oraz Panom: Krzysztofowi Adamkowi, Jarosławowi Apelskiemu, Łukaszowi Bieniaszowi-Krzywiec, Jarosławowi Błasiokowi, Dominikowi Borowcowi, Radosławowi Burnemu, Jackowi Chrząszczowi, Łukaszowi Czajce, Maciejowi Czerwińskiemu, Bartoszowi Dąbrowskiemu, Norbertowi Dojerowi, WoJciechowi Dudkowi, Mateuszowi Dzwonkowi, Oktawianowi Freusowi, Bartłomiejowi Gajewskiemu, Krzysztofowi Gerasowi, Maćkowi Fijałkowskiemu, Andrzejowi Findeisenowi, Dariuszowi Grali, Mateuszowi Greszcie, Danielowi Hansowi, Szczepanowi Hummelowi, Joachimowi Jelisiejewowi, Łukaszowi Kalbarczykowi, Szymonowi Kamińskiemu, Marcinowi Kościelnickiemu, Piotrowi Książkowi, Bolesławowi Kulbabińskiemu, Sławomirowi Lasocie, Grzegorzowi Leszczyńskiemu, Aleksandrowi Lewandowskiemu, Kamilowi Majdanikowi, Szymonowi Matejczykowi, Łukaszowi Marecikowi, Adamowi Michalikowi, Dominikowi Miedzińskiemu, Linhowi Anh Nguyenowi, Michałowi Ogińskiemu, Michałowi Oniszczukowi, Krzysztofowi Opolskiemu, Adamowi Panasiukowi, Adrianowi Panasiukowi, Miłoszowi Piechockiemu, Karolowi Piotrowskiemu, Wojciechowi Przybyszewskiemu, Damianowi Rodziewiczowi, Markowi Rojowi, Maciejowi Różańskiemu, Krzysztofowi Sachanowiczowi, Sławomirowi Sadziakowi, Pawłowi Selwetowi, Michałowi Skrzypczakowi, Adamowi Slaskiemu, Marcinowi Sulikowskiemu, Janowi Szejko, Michałowi Świtakowskiemu, Marcinowi Tatjewskiemu, Szymonowi Toruńczykowi, Bartłomiejowi Wiśniewskiemu, Wojciechowi Wiśniewskiemu, Michałowi Woźniakowi i Maciejowi Zdanowiczowi.