

1 Szablony - definicja, zastosowania

Szablony są jedną z głównych cech języka *C++*. Dzięki nim możemy dostarczać generyczne typy i funkcje, bez kosztów czasu wykonania. Skupiają się na pisaniu kodu w sposób niezależny od konkretnego typu, dzięki czemu wspierają programowanie generyczne. *C++* to bogaty język wspierający polimorficzne zachowania zarówno w czasie wykonania jak i kompilacji. W tym pierwszym używa hierarchii klas i wywołań funkcji wirtualnych by wspierać praktyki zorientowane obiektowo, gdzie wywoływana funkcja zależy od typu obiektu docelowego podczas czasu wykonania. Natomiast w czasie kompilacji szablony wspierają programowanie generyczne, gdzie wywoływana funkcja zależy od statycznego typu czasu kompilacji argumentów szablonu.

Grają kluczową rolę w projektowaniu obecnych, znanych i popularnych bibliotek i systemów. Stanowią podstawę technik programowania w różnych dziedzinach, począwszy od konwencjonalnego programowania ogólnego przeznaczenia do oprogramowywania wbudowanych systemów bezpieczeństwa.

Szablon to coś w rodzaju przepisu, z którego translator *C++* generuje deklaracje.

```
template<typename T>
T kwadrat (T x) {
    return x * x;
}
```

Kod ten deklaruje rodzinę funkcji indeksowanych po parametrze typu. Można odnieść się do konkretnego członka tej rodziny przez zastosowanie konstrukcji `kwadrat<int>`. Mówimy wtedy, że żądana jest specjalizacja szablonu dla funkcji `kwadrat` z listą argumentów szablonu `<int>`. Proces tworzenia specjalizacji nosi nazwę inicjalizacji szablonu, potocznie zwany inicjalizacją. Kompilator *C++* stworzy stosowny odpowiednik definicji funkcji:

```
int kwadrat(int x) {
    return x * x;
}
```

Argument typu `int` jest podstawiony za parametr typu `T`. Kod wynikowy jest sprawdzany pod względem typu, by zapewnić brak błędów wynikających z podmiany. Inicjalizacja szablonu jest wykonywana tylko raz dla danej specyfikacji nawet jeśli program zawiera jej wielokrotne żądania.

W przeciwieństwie do języków takich jak *Ada* czy *System F*, lista argumentów szablonu może być pominięta z żądania inicjalizacji szablonu funkcji. Zazwyczaj, wartości parametrów szablonu są dedukowane.

```
double d = kwadrat(2.0);
```

Argument typu jest dedukowany na `double`. Warto zauważyć, że odmiennie niż w językach takich jak *Haskell* czy *System F*, parametry szablonu w *C++* nie są ograniczone względem typów.

Szablonów używa się do zmniejszania kar abstrakcji i zjawiska *code bloat* w systemach wbudowanych w stopniu, który jest niepraktyczny w standardowych systemach obiektowych. Robi się to z dwóch powodów:

- Po pierwsze, inicjalizacja szablonu łączy informacje zarówno z definicji, jak i z kontekstu użycia. To oznacza, że pełna informacja zarówno z definicji jak i z wywołanych kontekstów (włączając w to informacje o typach) jest udostępniana generatorowi kodu. Dzisiejsze generatory kodu dobrze sobie radzą z używaniem tych informacji w celu zminimalizowania czasu wykonania i przestrzeni kodu. Różni się to od zwykłego przypadku w języku obiektowym, gdzie wywołujący i wywoływany są kompletnie oddzieleni przez interfejs, który zakłada pośrednie wywołania funkcji.
- Po drugie, szablon w *C++* jest zazwyczaj domyślnie tworzony tylko jeśli jest używany w sposób niezbędny dla semantyki programu, automatycznie minimalizując miejsce w pamięci, które wykorzystuje aplikacja. W przeciwieństwie do języka *Ada* czy *System F*, gdzie programista musi wyraźnie zarządzać inicjalizacjami.

1.1 Parametryzacja szablonów

Parametry szablonu są specyfikowane na dwa sposoby:

1. *parametry szablonu* – wyraźnie wspomniane jako parametry w deklaracji szablonu
2. *nazwy zależne* - wywnioskowane z użycia parametrów w definicji szablonu

W *C++* nazwa nie może być użyta bez wcześniejszej deklaracji. To wymaga od użytkownika ostrożnego traktowania definicji szablonów. Np. w definicji funkcji `kwadrat` nie ma widocznej deklaracji symbolu `*`. Jednak, podczas inicjalizacji szablonu `kwadrat<int>` kompilator może sprowadzić symbol `*` do (wbudowanego) operatora mnożenia dla wartości `int`. Dla wywołania `kwadrat(zespolona(2.0))`, operator `*` zostałby rozwiązany do (zdefiniowanego przez użytkownika) operatora mnożenia dla wartości `zespolona`. Symbol `*` jest więc *nazwą zależną* w definicji funkcji `kwadrat`. Oznacza to, że jest to ukryty parametr definicji szablonu. Możemy uczynić z operacji mnożenia formalny parametr:

```
template<typename Multiply , typename T>
T square(T x) {
```

```
    return Multiply() (x,x);  
}
```

Pod-wyrażenie `Multiply()` tworzy obiekt funkcji, który wprowadza operacje mnożenia wartości typu `T`. Pojęcie *nazw zależnych* pomaga utrzymać liczbę jawnych argumentów.

1.2 Inicjalizacje i sprawdzanie

Minimalne przetwarzanie semantyczne odbywa się, gdy po raz pierwszy widzi definicję szablonu lub jego użycie. Pełne przetwarzanie semantyczne jest przesuwane na czas inicjalizacji (tuż przed czasem linkowania), na podstawie każdej instancji. Oznacza to, że założenia dotyczące argumentów szablonu nie są sprawdzane przed czasem inicjalizacji. Np.

```
string x = "testowy tekst";  
kwadrat(x);
```

Bezsensowne użycie zmiennej `string` jako argumentu funkcji `kwadrat` nie jest wyłapane w momencie użycia. Dopiero w czasie inicjalizacji kompilator odkryje, że nie ma odpowiedniej deklaracji dla operatora `*`. To ogromny praktyczny błąd, bo inicjalizacja może być przeprowadzona przez kod napisany przez użytkownika, który nie napisał definicji funkcji `kwadrat` ani definicji `string`. Programista, który nie znał definicji funkcji `kwadrat` ani `string` miałby ogromne trudności w zrozumieniu komunikatów błędów związanych z ich interakcją (np. "illegal operand for *").

Istnienie symbolu operatora `*` nie jest wystarczające by zapewnić pomyślną kompilację funkcji `kwadrat`. Musi istnieć operator `*`, który przyjmuje argumenty odpowiednich typów i ten operator `*` musi być bezkonkurencyjnym dopasowaniem według zasad przeciążania C++. Dodatkowo funkcja `kwadrat` przyjmuje argumenty przez wartość i zwraca swój wynik przez wartość. Z tego wynika, że musi być możliwe skopiowanie obiektów dedukowanego typu. Potrzebny jest rygorystyczny framework do opisywania wymagań definicji szablonów na ich argumentach.

Doświadczenie podpowiada nam, że pomyślna kompilacja i linkowanie może nie gwarantować końca problemów. Udana budowa pokazuje tylko, że inicjalizacje szablonów były poprawne pod względem typów, dostając argumenty które przekazaliśmy. Co z typami argumentów szablonu i wartościami, z którymi nie próbowaliśmy użyć naszych szablonów? Definicja szablonu może zawierać przypuszczenia na temat argumentów, które przekazaliśmy ale nie zadziała dla innych, prawdopodobnie rozsądnych argumentów. Uproszczona wersja klasycznego przykładu:

```
template<typename FwdIter>  
bool czyJestPalindromem(FwdIter first, FwdIter last){
```

```

    if (last <= first) return true;
    if (*first != *last) return false;
    return czyJestPalindromem(++first, --last);
}

```

Testujemy czy sekwencja wyznaczona przez parę iteratorów do jego pierwszego i ostatniego elementu, jest palindromem. Przyjmuje się, że te iteratory są z kategorii *forward iterator*. To znaczy, że powinny wspierać co najmniej operacje takie jak: `*`, `!=` i `++`. Definicja funkcji `czyJestPalindromem` bada czy elementy sekwencji zbiegają z początku i końca do środka. Możemy przetestować tę funkcję używając `vector`, tablicę w stylu `C` i `string`. W każdym przypadku nasz szablon funkcji zainicjalizuje się i wykona się poprawnie. Niestety, umieszczenie tej funkcji w bibliotece byłoby dużym błędem. Nie wszystkie sekwencje wspierają `--` i `<=`. Np. listy pojedyncze nie wspierają. Eksperci używają wyszukanych, regularnych technik by uniknąć takich problemów. Jednakże, fundamentalny problem jest taki, że definicja szablonu nie jest (według siebie) dobrą specyfikacją jego wymagań na jego parametry.

1.3 Wydajność

Szablony grają kluczową rolę w programowaniu w `C++` dla wydajnych aplikacji. Ta wydajność ma trzy źródła:

- eliminacja wywołań funkcji na korzyść *inliningu*
- łączenie informacji z różnych kontekstów w celu lepszej optymalizacji
- unikanie generowania kodu dla niewykorzystanych funkcji

Pierwszy punkt nie odnosi się tylko do szablonów ale ogólnie do cech funkcji *inline* w `C++`. Jakkolwiek, *inlining* jest istotny dla drobno-granularnej parametryzacji, którą powszechnie stosuje się w bibliotece *STL* i innych bibliotekach bazujących na generycznych technikach programowania. Wydajność ta przekłada się zarówno na czas wykonania jak i pamięć. Szablony mogą równocześnie zmniejszyć obie wydajności. Zmniejszenie rozmiaru kodu jest szczególnie ważne, ponieważ w przypadku nowoczesnych procesorów zmniejszenie rozmiaru kodu pociąga za sobą zmniejszenie ruchu w pamięci i poprawienie wydajności pamięci podręcznej.

```

template<typename FwdIter, typename T>
T suma(FwdIter first, FwdIter last, T init){
    for(FwdIter cur = first, cur != last, T init)
        init = init + *cur;
    return init;
}

```

Funkcja `suma` zwraca sumę elementów jej sekwencji wejściowej używając trzeciego argumentu ("akumulatora") jako wartości początkowej

```
vector<zespolona<double>> v;  
zespolona<double> z = 0;  
z = suma(v.begin(), v.end(), z);
```

By wykonać swoją pracę, `suma` użyje operatorów dodawania i przypisania na elementach typu `zespolona<double>` i dereferencji iteratorów `vector<zespolona<double>>`. Dodanie wartości typu `zespolona<double>` pociąga za sobą dodanie wartości typu `double`. By zrobić to wydajnie wszystkie te operacje muszą być *inline*. Zarówno `vector` jak i `zespolona` są typami zdefiniowanymi przez użytkownika. Oznacza to, że typy te jak i ich operacje są zdefiniowane gdzie indziej w kodzie źródłowym *C++*. Obecne kompilatory *C++* radzą sobie z tym przykładem, dzięki czemu jedyne wygenerowane wywołanie to wywołanie funkcji `suma`. Dostęp do pól zmiennej `vector` staje się prostą operacją maszyny ładującej, dodawanie wartości typu `zespolona` staje się dwiema instrukcjami maszyny dodającej dwa elementy zmiennoprzecinkowe. Aby to osiągnąć, kompilator potrzebuje dostępu do pełnej definicji `vector` i `zespolona`. Jednak wynik jest ogromną poprawą (prawdopodobnie optymalną) w stosunku do naiwnego podejścia generowania wywołania funkcji dla każdego użycia operacji na parametrze szablonu. Oczywiście instrukcja dodawania wykonuje się znacznie szybciej niż wywołanie funkcji zawierającej dodawanie. Poza tym, nie ma żadnego wstępu wywołania funkcji, przekazywanie argumentów itd., więc kod wynikowy jest również wiele mniejszy. Dalsze zmniejszanie rozmiaru generowanego kodu uzyskuje się nie wysyłając kodu niewykorzystywanych funkcji. Klasa szablonu `vector` ma wiele funkcji, które nie są wykorzystywane w tym przykładzie. Podobnie szablon klasy `zespolona` ma wiele funkcji i funkcji nieskładowych (nienależących do funkcji klasy). Standard *C++* gwarantuje, że nie jest emitowany żaden kod dla tych niewykorzystanych funkcji.

Aby kontrastować, rozważ bardziej konwencjonalny przypadek, w którym argumenty są dostępne za pośrednictwem interfejsów zdefiniowanych jako wywołania funkcji pośrednich. Każda operacja staje się wtedy wywołaniem funkcji w pliku wykonywalnym generowanym dla kodu użytkownika, takiego jak `suma`. Co więcej, byłoby wyraźnie nietypowe unikać odkładania kodu nieużywanych (wirtualnych) funkcji składowych. Jest to poza zdolnością obecnych kompilatorów *C++* i prawdopodobnie pozostanie takie dla głównych programów *C++*, gdzie oddzielna kompilacja i łączenie dynamiczne jest normą. Ten problem nie jest wyjątkowy dla *C++*. Opiera się on na podstawowej trudności w ocenieniu, która część kodu źródłowego jest używana, a która nie, gdy jakkolwiek forma wysyłki czasu wykonania ma miejsce. Szablony nie cierpią na ten problem bo ich specjalizacje są rozwiązywane w czasie kompilacji.

Przykład funkcji `suma` nie jest idealny do zilustrowania subtelności generowania kodu obiektu z kodu źródłowego znalezionej w różnych częściach programu. Nie polega na niejawnym konwersjach lub nietypowych przeciążaniach. Jednak, rozważ wariant gdzie wartości `int` są sumowane w obiekcie `zespolona<double>`:

```
vector<int> v;  
zespolona<double> s = 0;  
s = suma(v.begin(), v.end(), s);
```

Tu dodawanie jest wykonane przez konwertowanie wartości `int` do wartości `double` i potem dodawanie tego do akumulatora `s`, używając operatora `+` typu `zespolona<double>` i `double`. To podstawowe dodawanie zmiennoprzecinkowe. Kwestia jest taka, że operator `+` w funkcji `suma` zależy od dwóch parametrów szablonu i leży to w kwestii kompilatora by wybrać bardziej odpowiedni operator `+` bazując na informacji o tych dwóch argumentach. Byłoby możliwe utrzymanie lepszego rozdzielania między różnymi kontekstami przez zawsze przekształcanie typu elementu w typ akumulatora. W takim przypadku spowodowałoby to powstanie dodatkowego `zespolona<double>` dla każdego elementu i dodania dwóch wartości typu `zespolona`. Rozmiar kodu i czas wykonywania byłyby większe niż dwukrotnie.

Nie spodziewalibyśmy się zobaczyć tego ostatniego przykładu bezpośrednio w kodzie źródłowym. Gdybyśmy go zobaczyli, uznalibyśmy, że jest on źle napisany. Jednakże, równoważny kod jest powszechny w wyniku zagnieżdżonych abstrakcji. Jest to szczególnie ważne by generować dobry kod w takich przypadkach ponieważ nie robienie tego byłoby zniechęcające dla abstrakcji.

Warto zauważyć, że te optymalizacje są wspólnym miejscem. Duże ilości prawdziwego oprogramowania zależą od nich. W konsekwencji udoskonalone sprawdzanie typu, co zostało obiecanie przy użyciu konceptów, nie może kosztować tych optymalizacji.