

JiPP zadanie 2 - „Interpreter”

W ramach zadania 2 należy zaimplementować interpreter języka programowania.

Za zadanie można uzyskać maksymalnie 30 punktów. Na ocenę ma wpływ zarówno zakres projektu („wielkość” i „trudność” języka wybranego do implementacji) jak i jakość rozwiązania.

Harmonogram

Zadanie podzielone jest na trzy fazy (trzecia faza jest opcjonalna).

1. Deklaracja języka do implementacji – do czwartku 6 kwietnia 2023.

Należy oddać (przez Moodle'a) dokument zawierający opis języka wybranego do implementacji. Format pliku: PDF, płaski tekst lub ZIP zawierający pliki wspomnianego typu. Prosimy o nazwy plików postaci `imie_nazwisko.pdf`. Zawarte mają być:

- gramatyka języka – można ją podać w notacji EBNF (szczególnie zalecane osobom chcącym skorzystać z BNFC) lub w dowolnej rozsądnej postaci „na papierze” (można zacząć od pewnego poziomu abstrakcji – nie definiować literałów, identyfikatorów itp.),
- kilka przykładowych programów ilustrujących planowane konstrukcje składniowe,
- tekstowy opis języka z podkreśleniem nietypowych konstrukcji; uwaga! nie opisujemy rzeczy oczywistych, np. w przypadku zapożyczeń ze znanych języków jak Pascal, C, Haskell wystarczy je wymienić, bez dokładnego opisu
- wypełnioną odpowiednią *tabelkę cech* (dołączone są do treści zadania) oraz łączną liczbę punktów, jaką autor spodziewa się dostać za interpreter, o ile poprawnie zaimplementuje całą podaną funkcjonalność – patrz (dużo) niżej.

Za niewykonanie w terminie tego etapu zadania od oceny końcowej zostanie odjętych 8 punktów, a za wykonanie niepełne (np. brak gramatyki) do 8 punktów.

W odpowiedzi na deklarację języka sprawdzający potwierdzi (lub nie :) maksymalną ocenę, jaką można uzyskać za poprawne zrealizowanie takiego języka. Ostateczny zakres projektu może jednak jeszcze zostać zmieniony w porozumieniu ze sprawdzającym.

2. Pierwsza wersja interpretera – do niedzieli 14 maja.

Działająca implementacja interpretera. Należy oddać plik `imie_nazwisko.zip`, o zawartości opisanej poniżej. Oddanie rozwiązania przez Moodle w terminie jest obowiązkowe. Dodatkowo sprawdzający może poprosić o osobistą prezentację rozwiązania.

3. Ostateczna wersja interpretera – do niedzieli 4 czerwca.

Opcjonalnie, w porozumieniu ze sprawdzającym – dodatkowe funkcjonalności w celu podwyższenia oceny.

Uwaga, uzupełnianie istotnych błędów i braków znalezionych przez sprawdzającego może nie rekompensować w pełni odjętych punktów. Na pewno nie można liczyć na bardzo dobrą ocenę oddając początkowo „wydmuszkę” i uzupełniając rozwiązanie dopiero w drugiej iteracji.

Implementacja i zawartość paczki

Interpreter należy zaimplementować w Haskellu.

Jako rozwiązanie należy oddać plik `imie_nazwisko.zip`, który po rozpakowaniu tworzy katalog `imie_nazwisko`, w którym wywołanie polecenia `make` buduje bez przeszkód (na maszynie `students`, przy użyciu zainstalowanego tam kompilatora Haskell w wersji 8.8.4) działający interpreter. Ma się on uruchamiać poleceniem `./interpreter program`, gdzie `program` oznacza plik z programem do interpretacji. Ponadto, jeśli w języku nie ma obsługi standardowego wejścia, interpreter wywołany bez parametru może wczytywać program ze standardowego wejścia. Wyniki mają wypisywać się na standardowe wyjście, komunikaty o błędach na standardowe wyjście błędów. Domyślnie interpreter nie powinien wypisywać żadnych dodatkowych komunikatów

diagnostycznych.

Rozwiązanie ma również zawierać plik **README** z ogólnym opisem rozwiązania i sposobu uruchomienia (zwłaszcza gdyby był nieoczywisty), zaktualizowaną *tabelkę cech* oraz przykładowe programy:

- w podkatalogu **good** przykłady poprawnych programów ilustrujących **wszystkie** punktowane konstrukcje języka, w miarę możliwości testy powinny nazywać się nn-nazwa, lub nn-mm-nazwa, gdzie nn i mm to numery konstrukcji z jednej z list wymagań dla standardowych projektów poniżej, np. `06-13-przekazywanie-tablic.prg`
- w podkatalogu **bad** przykłady niepoprawnych programów, aby zilustrować działanie interpretera w sytuacjach wyjątkowych, na przykład (ale w zależności od języka listę należy wydłużyć):
 - błędy składniowe (nie za dużo),
 - nieznany identyfikator, nieznana funkcja itp.,
 - zła liczba argumentów,
 - błędy typów i inne błędy wykrywane statycznie (jeśli rozwiązanie wspiera statyczną kontrolę typów),
 - błędy czasu wykonania (dzielenie przez zero, odwołanie do indeksu spoza tablicy itp.).

Język

Nie ma wstępnych ograniczeń na język - może być imperatywny, funkcyjny, obiektowy, logiczny albo mieszany. Szczególnie cenimy nieszablonowe pomysły i dla takich „ciekawych” języków sprawdzający mogą podejmować indywidualne decyzje odnośnie oceny nawet jeśli język nie spełnia podanych poniżej wymagań dla projektów standardowych. Nie będziemy natomiast wysoko cenić języków, których oryginalność ogranicza się do wymyślnej składni.

Dla ułatwienia studentom decyzji, a sprawdzającym ujednolicenia ocen, przedstawiamy poniżej dwie linie standardowych projektów i wymagania na poszczególne oceny (przy założeniu poprawnej i porządnie napisanej implementacji). Do treści zadania dołączone są *tabelki cech*: jedną z nich należy wypełnić na etapie projektu oraz zaktualizować na etapie realizacji języka, wstawiając znak **+** przy planowanych / zrealizowanych funkcjonalnościach.

Oczywiście w celu uzyskania pełnej liczby punktów na danym poziomie należy oprócz danych cech języka również zaimplementować naturalnie wynikające kombinacje tych cech, np. implementując przekazywanie przez zmienną oraz tablice, należy również zaimplementować przekazywanie przez zmienną tablic itp.

Docelowa gramatyka języka powinna być w miarę możliwości wolna od konfliktów. Jeśli niektórych konfliktów nie da się (w miarę łatwo) uniknąć, należy je udokumentować - podać przykłady wyrażen prezentujących konflikt oraz skutki rozwiązania konfliktu przyjętego przez użyty generator parserów. Nieuudokumentowane konflikty w gramatyce mogą spowodować obniżenie oceny.

Powinna być możliwość umieszczania w programach komentarzy.

Ze względu na łatwość obsługi błędów, polecamy BNFC z opcją `--functor` jako generator parsera. Program ten jest dostępny pod adresem <http://bnfc.digitalgrammars.com/> a także na maszynie students w katalogu `/home/students/inf/PUBLIC/MRJP/bin`. Polecana opcja `--functor` powoduje, że w drzewie składni abstrakcyjnej umieszczane są pozycje odpowiednich fragmentów pliku wejściowego. To znacznie ułatwia wypisywanie użytecznych komunikatów o błędach, bez czego interpreter nie może uzyskać wysokiej oceny.

Osobom, które planują napisać interpreter języka statycznie typowanego radzimy pisać type-checker i interpreter równolegle lub type-checker najpierw. Pisanie type-checkera na końcu jest dość niewdzięcznym zadaniem, ponieważ okazuje się wtedy, że większość błędów, które zgłasza type-checker i tak jest już zgłaszana przez interpreter (np. że próbuję podzielić liczbę przez napis).

Język imperatywny

Język imperatywny, najlepiej w składni opartej o C lub Javę. W przypadku braku własnych oryginalnych pomysłów można wzorować się na języku Latte

<https://www.mimuw.edu.pl/~ben/Zajecia/Mrj2022/Latte/>

Na 15 punktów

1. Co najmniej trzy typy wartości: `int`, `bool` i `string`
(to znaczy `if 2+2 then _` parsuje się, ale wyrażenie ma niepoprawny typ).
2. Literały, arytmetyka, porównania.
3. Zmienne, operacja przypisania
4. Jawne wypisywanie wartości na wyjście (instrukcja lub wbudowana procedura `print`).
5. `while`, `if` (z `else` i bez, może być też składnia `if _ elif _ else _ endif`).
6. Funkcje lub procedury (bez zagnieżdżania), rekurencja.
Jedna rzecz z poniższej listy lub coś o porównywalnej trudności:
7. co najmniej dwa sposoby przekazywania parametrów (przez zmienną / przez wartość / in/out),
8. zmienne „read-only” i użycie ich np. w implementacji pętli `for` w stylu Pascala
(`for i = pocz to kon` - wewnątrz pętli nie można zmienić wartości zmiennej sterującej, wartość `kon` liczona tylko raz - przed wejściem do pętli)

Na 20 punktów

J.w., a ponadto:

9. Przesłanie identyfikatorów ze statycznym ich wiązaniem (zmienne lokalne i globalne lub zagnieżdżone procedury/funkcje).
10. Obsługa błędów wykonania, np. dzielenie przez zero (może być elegancki komunikat i zatrzymanie interpretera).
11. Funkcje przyjmujące i zwracające wartość dowolnych obsługiwanych typów (tzn. nie tylko procedury; za to mogą być tylko funkcje – jak w języku C).

Do 30 punktów wg cennika...

12. Statyczne typowanie (tj. zawsze terminująca faza kontroli typów przed rozpoczęciem wykonania programu) – 4pkt,
13. Dowolnie zagnieżdżone definicje funkcji / procedur z zachowaniem poprawności statycznego wiązania identyfikatorów (jak w Pascalu) – 2 pkt,
14. Rekordy albo tablice indeksowane `int` albo coś à la listy – 1pkt,
ALBO tablice wielowymiarowe przekazywane i przypisywane "przez wskaźnik" (jak w Javie), a nie "przez kopię" – 2pkt
15. Dowolnie zagnieżdżone krotki z przypisaniem jak w Pythonie (składnia wedle uznania) – 2pkt,
16. Operacje przerywające pętlę `while` - `break` i `continue` – 1pkt,
17. Funkcje jako parametry,
zwracanie funkcji w wyniku, domknięcia à la JavaScript.
funkcje anonimowe – 4pkt,
18. Procedury generujące i składnia do ich używania (np. jak w Pythonie - instrukcja `yield` oraz `next`, a także `for x in generator(...)`) – 3pkt.

Język funkcyjny

Język funkcyjny, najlepiej w składni opartej o składnię SML/Camla lub Haskellu (lub Lisp/Scheme dla prostszych).

Na 20 punktów

1. Co najmniej dwa typy wartości w wyrażeniach: `int` i `bool`
(to znaczy `if 2+2 then _` parsuje się, ale wyrażenie ma niepoprawny typ).
2. Arytmetyka, porównania.
3. Wyrażenie warunkowe `if`.
4. Funkcje wieloargumentowe, rekurencja.
5. Funkcje anonimowe, częściowa aplikacja, funkcje wyższego rzędu, domknięcia.
6. Obsługa błędów wykonania, np. dzielenie przez zero (może być elegancki komunikat i zatrzymanie interpretera).
7. Statyczne wiązanie identyfikatorów przy dowolnym poziomie zagnieżdżenia definicji.

Listy z:

8. pattern matchingiem `[] | x:xs` (składnia może być inna niż w Haskellu),
9. **lub** zestawem wbudowanych operacji: `empty`, `head`, `tail`,
10. mile widziany lukier syntaktyczny do wpisywania stałych list, np.: `[1, 2, 3]`.

Na 25 punktów

J.w., a ponadto:

11. Listy dowolnego typu, także listy zagnieżdżone i listy funkcji
12. **Lub** ogólne rekurencyjne typy algebraiczne (jak `data` w Haskellu) z pattern matchingiem.
Mogą być monomorficzne a pattern matching jednopoziomowy.
13. Statyczne typowanie (tj. zawsze terminująca faza kontroli typów przed rozpoczęciem wykonania programu). Na tym poziomie można wymagać jawnego podawania typów.

Na 30 punktów

J.w., a ponadto:

14. Ogólne **polimorficzne i rekurencyjne** typy algebraiczne.
Mile widziane wykonane w samym języku definicje typów `List` (składnia może być inna niż w Haskellu, lukier syntaktyczny nie wymagany), `Maybe` i `Either` oraz ich zastosowania w przykładowych programach.
15. Dowolne zagnieżdżenie wzorców w pattern matchingu.

Bonus do 6 punktów

16. Typy polimorficzne w stylu ML (jak Caml lub Haskell bez klas) z algorytmem rekonstrukcji typów. Nie jest konieczna składnia do deklarowania (skrótów) typów – 4 pkt
17. Sprawdzanie kompletności pattern matchingu i ostrzeżenie przy próbie zdefiniowania funkcji częściowej (wymaga pattern matchingu o dowolnym zagnieżdżeniu) – 2pkt
Za interpreter implementujący ten algorytm można dostać więcej niż 30 punktów.

Uwaga! Otrzymanie punktów w II terminie za rekonstrukcję typów jest możliwe tylko w przypadku gdy interpreter w I terminie zawierał sprawdzanie typów (statyczne typowanie) i został oceniony na (prawie) 30 punktów (liczenie na „wyciągnięcie” kiepskiego interpretera za pomocą algorytmu rekonstrukcji typów w przeszłości zwykle kończyło się rozczarowaniem).

W razie poważnych kłopotów...

Za realizację części w/w funkcjonalności w wersji z tylko jednym typem **Int** (instrukcje warunkowe z porównaniami lub interpretacją 0/1 jak w C), ocena zostanie obniżona o 5 punktów.

Zapożyczenia

Projekt zaliczeniowy ma być pisany samodzielnie. Wszelkie przejawy niesamodzielności będą karane. W szczególności nie wolno oglądać kodu innych studentów, pokazywać, ani w jakikolwiek sposób udostępniać swojego kodu.

Dopuszczalne są jawne (z podaniem źródła i poszanowaniem praw autorskich) zapożyczenia elementów rozwiązania nie negujące własnego wkładu pracy i intelektu w całokształt projektu, np.:

- wykorzystanie ogólnie dostępnej gramatyki języka,
- oparcie się w swoim rozwiązaniu o dostępny (i wskazany w rozwiązaniu) opis pewnej techniki lub algorytmu (np. realizacja przesłaniania identyfikatorów, algorytm rekonstrukcji typów); kod ma być w tym przypadku napisany samodzielnie.

W rozwiązaniu można do woli korzystać z **własnych** projektów realizowanych przy innej okazji, np. na potrzeby innych przedmiotów czy w poprzedniej edycji JiPP. Należy jednak pamiętać, że oczekiwany jest interpreter (a nie np. kompilator) i że projekt będzie oceniany od nowa.