

MASARYKOVA UNIVERZITA
FAKULTA INFORMATIKY



Studie nástrojů pro trasování a testování programů v Javě

BAKALÁRSKA PRÁCA

Matej Majdiš

Brno, 2015

Prehlásenie

Prehlasujem, že táto bakalárska práca je mojím pôvodným autorským dielom, ktoré som vypracoval samostatne. Všetky zdroje, pramene a literatúru, ktoré som pri vypracovaní používal alebo z nich čerpal, v práci riadne citujem s uvedením úplného odkazu na príslušný zdroj.

Vedúci práce: RNDr. Adam Rambousek

Zhrnutie

TODO...

Klíčové slová

TODO...

Pod'akovanie

TODO...

Obsah

1	Úvod	2
1.1	Cieľ práce	3
1.2	Členenie práce	3
2	Bajtkód	4
2.1	Štruktúra class súboru	4
2.1.1	Dátové typy	6
2.1.2	Premenné tried a inšancií	7
2.1.3	Metódy	8
2.1.4	Atribúty	9
2.2	Inštrukčná sada bajtkódu	9
	Literatúra	10
A	Tabuľky	11

Kapitola 1

Úvod

Java je dnes jedným z najpoužívanějších programovacích jazykov. Od syntakticky podobných programovacích jazykov ako napríklad C++, alebo C# sa líši prekladom zdrojových tried do medzikódu často označovaného ako bajtkód (*bytecode*, *p-code*, *portable code*).

Preklad a spustenie programu napísaných v programovacom jazyku Java prebieha v nasledujúcich fázach:

1. Preklad do medzikódu: Java compiler ¹ preloží zdrojový kód do bajtkódu. V praxi to znamená, že každej triede, alebo rozhraniu je priradený súbor *class*, ktorý obsahuje inštrukcie popisujúce fungovanie danej triedy.
2. Načítanie a Interpretácia: Virtuálny stroj Javy (ďalej len JVM ²) načíta inštrukcie *class* súboru potrebnej triedy, ktoré ďalej spracúva jedným z nasledujúcich spôsobov:
 - JIT prekladač (*Just In Time compiler*): Štandardne je z bajtkódu najskôr vygenerovaný strojový (*machine code*) konkrétného zariadenia, ktorý je následne interpretovaný priamo vykonávaný procesorom.
 - Java interpreter: Ďalším spôsobom spracovania bajtkódu je využitie Java interpretu, ktorý bajtkódkód spracováva a sám interpretuje.

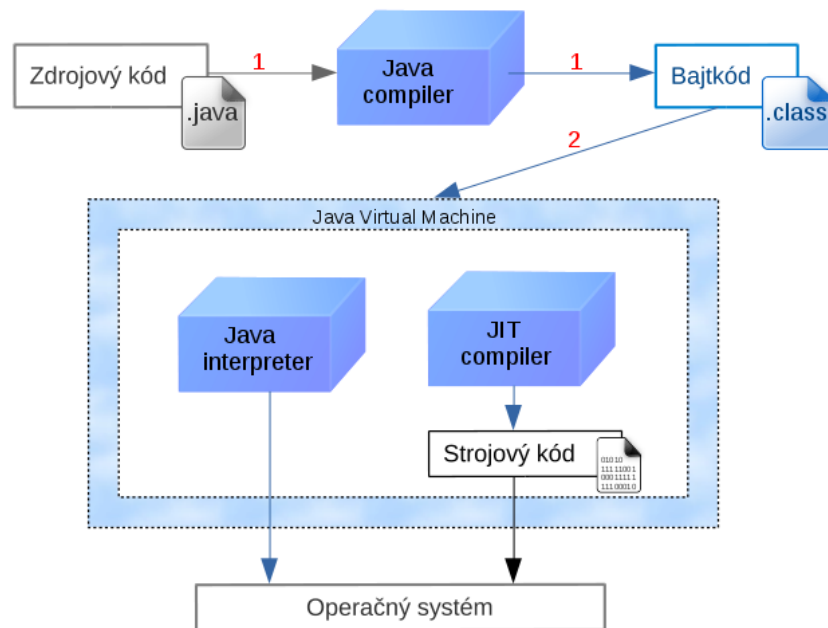
Výhodou prekladu do bajtkódu je jeho a prenositeľnosť. Samotný bajtkód je platformovo nezávislý. Program teda nieje nutné prispôbovať jednotlivým operačným systémom, ktoré sa líšia len v implementácii JVM.

Class súbory obsahujúce bajtkód je možné za behu programu modifikovať. Jednotlivé triedy a rozhrania aplikácie uložené v týchto súboroch

1. Najčastejšie využívaným *javac*om je *javac*, ktorý je súčasťou JDK.

2. Java Virtual Machine, špecifikácia je dostupná na <http://docs.oracle.com/javase/specs/jvms/se7/html>

podľa potreby načítava JVM. Vkladanie nových metód a tried na úrovni bajtkódu, pred načítaním *class* súboru do JVM sa nazýva injekcia bajtkódu (*bytecode injection*, ďalej len BI). Pridávanie novej funkcionality pomocou BI bez nutnosti zastavenia behu programu je často využívané pri testovaní a trasovaní (*tracing*) programov.



Obr. 1.1: Grafické znázornenie prekladu a spustenia programu, zdroj: vlastné spracovanie

1.1 Cieľ práce

TODO...

1.2 Členenie práce

TODO...

Kapitola 2

Bajtkód

Po preklade zdrojových kódov prekladačom *javac* je každej triede, prípadne rozhraniu programu priradený jeden *class* súbor popisujúci jej funkcionality.

Pri načítavaní *class* súbru JVM dostane takzvaný prúd inštrukcií bajtkódu (*bytecode stream*) pre každú metódu triedy. V prípade volania konkrétnej metódy za behu programu sú inštrukcie danej metódy vykonávané. Každá z inštrukcií bajtkódu je reprezentovaná číselnou hodnotou nazývanou *opcode*. Zároveň má každá inštrukcia aj textovú podobu (*mnemonic*), ktorá je jej menom. V *class* súbooroch sú inštrukcie uložené v numerickej podobe.

Táto kapitola popisuje formát *class* súboru a následne stručne charakterizuje inštrukčnú sadu bajtkódu.¹

2.1 Štruktúra *class* súboru

Class súbor pozostáva z jednej *ClassFile* štruktúry. *ClassFile* štruktúra jednoznačne identifikuje konkrétnu triedu, prípadne rozhranie, definuje jej premenné a metódy.

Nasledujúci popis definuje sadu datových typov. Typy *u1*, *u2*, a *u4* reprezentujú neznamienkové jedno, dvoj, alebo štvorbajtové číslo. *ClassFile* je zobrazená ako pseudoštruktúra v notácii jazyka C. Obsah štruktúry je popísaný ako po sebe nasledujúce položky.

Formát *ClassFile* štruktúry

```
ClassFile {  
    u4 magic;  
    u2 minor_version;  
    u2 major_version;  
    u2 constant_pool_count;  
    cp_info constant_pool[constant_pool_count-1];  
}
```

1. Nasledujúci text vychádza zo 4. až 6. kapitoly špecifikácie JVM [LYBB13].

```

u2 access_flags;
u2 this_class;
u2 super_class;
u2 interfaces_count;
u2 interfaces[interfaces_count];
u2 fields_count;
field_info fields[fields_count];
u2 methods_count;
method_info methods[methods_count];
u2 attributes_count;
attribute_info attributes[attributes_count];
}

```

Konštanta *magic* identifikuje formát súboru *class*, jej hodnota je 0xCA-FEBABE.

Položky *minor_version* a *major_version* určujú verziu *class* súboru. Napríklad *minor_version* s hodnotou *m* a *major_version* s hodnotou *M* indikujú verziu s hodnotou *M.m*.

Hodnota položky *constant_pool_count* je rovná počtu záznamov v *constant_pool[]* plus jeden.

Úložisko záznamov *constant_pool[]* (v podobe poľa štruktúr) zahŕňa rôzne konštanty: mená tried a rozhraní, mená premenných a iné. Každý záznam *constant_pool[]* sa skladá zo značky (*tag*) a indexu (*name index*). Značka určuje typ záznamu. Tabuľka značiek je uvedená v prílohe A.1. Pomocou unikátneho indexu, je možné odkazovať sa na záznamy v ďalších častiach bajtkódu. Existuje niekoľko typov štruktúr² reprezentujúcich rôzne druhy záznamov. Napríklad štruktúra *CONSTANT_String_info* reprezentuje objekty typu *String* zatiaľ čo štruktúry *CONSTANT_Methodref_info* a *CONSTANT_InterfaceMethodref_info* reprezentujú metódy danej triedy, alebo rozhrania.

Hodnota *access_flags* popisuje oprávnenia prístupu k informáciám a vlastnosti tejto triedy, respektíve rozhrania pomocou indikátorov. Napríklad nastavenie indikátora *ACC_INTERFACE* znamená, že *class* súbor popisuje rozhranie. Tabuľka indikátorov je uvedená v prílohe A.2.

Položka *this_class* obsahuje index *constant_pool[]* odkazujúci na štruktúru typu *CONSTANT_Class_info*³. Reprezentuje triedu, respektíve rozhranie, definované týmto *class* súborom.

Hodnotou *super_class* je taktiež index *constant_pool[]* odkazujúci na štruktúru typu *CONSTANT_Class_info*. Reprezentuje priamu nadtriedu triedy

2. Všetky štruktúry *constant_pool[]* sú popísané v špecifikácii JVM [LYBB13].

3. *CONSTANT_Class_info* je štruktúra *constant_pool*, ktorá reprezentuje triedu, alebo rozhranie.

definovanej týmto *class* súborom. V prípade, že tento *class* súbor popisuje rozhranie, index odkazuje na triedu *Object*. Trieda *Object* má ako jediná hodnotu *super_class* nulovú.

Počet rozhraní, ktoré trieda implementuje vyjadruje položka *interface_count*, v prípade rozhrania je táto položka rovná počtu priamych nadrozhraní.

Pole *interfaces[]* obsahuje indexy *constant_pool[]* odkazujúce na štruktúru typu *CONSTANT_Class_info*. Zahŕňa indexy všetkých rozhraní, ktoré sú implementované triedou, prípadne priamymi nadrozhraniami *class* súboru.

Položka *fields_count* je rovná počtu premenných triedy a premenných inštancií (*fields*) *class* súboru.

Štruktúry typu *field_info* sú združené v poli *fields[]*. Toto pole zahŕňa každú premennú danej triedy, respektíve rozhrania. Nezahŕňa zdedené atribúty. Podrobne sa štruktúrou *field_info* sa zaoberá kapitola 2.1.2.

Hodnota položky *methods_count* vyjadruje počet štruktúr *method_info* v poli *methods[]*.

Položka *methods[]* je pole štruktúr typu *method_info*. Každá štruktúra *method_info* popisuje metódu tejto triedy, respektíve rozhrania. Zahŕňa konštruktory, metódy triedy a metód inštancií. Neobsahuje však žiadne zdedené metódy. Štruktúru *method_info* popisuje kapitola 2.1.3.

Hodnota *attributes_count* je rovná počtu atribútov poľa *attributes[]* *class* súboru.

Pole *attributes[]* obsahuje štruktúry typu *attribute_info*. Atribútmi štruktúry *ClassFile* sú napríklad: *SourceFile*, *Deprecated*, *InnerClasses* a iné. Atribút *SourceFile* slúži na reprezentáciu mena *class* súboru. Pole *attributes[]* *class* súboru môže obsahovať maximálne jeden takýto atribút. Atribút *Deprecated* môže byť použitý v prípade, že bola daná trieda nahradená (*deprecated*). Pri volaní takejto triedy môže prekladač upozorniť užívateľa, že sa odkazuje na nahradenú triedu ⁴. Vo všeobecnosti sa štruktúre *attribute_info* sa venuje kapitola 2.1.4.

2.1.1 Dátové typy

Dátové typy sú v *class* súboroch reprezentované vo formáte reťazcov s kódovaním *UTF-8*. Delíme ich na:

- dátové typy premenných
 - primitívne dátové typy
 - referenčné dátové typy

4. Rovnakým spôsobom je možné atribút *Deprecated* aplikovať aj na premenné a metódy.

- polia
- dátové typy metód

Primitívnym dátovým typom (*byte*, *integer*, ...) je priradený popis v podobe znaku (*B*, *I*, ...). Napríklad premenná typu *int* je reprezentovaná znakom: *I*.

Referenčné dátové typy reprezentuje popis v tvare: *L<classname>;*, kde *classname* je meno triedy, alebo rozhrania daného referenčného dátového typu. Premenná typu *Object* je interpretovaná ako *java/lang/Object;*.

Identifikačný reťazec jednorozmerného poľa typu *T* sa značí [*T*, pričom počet znakov *I* je rovný dimenzii poľa. Napríklad premenná typu: *double d[][][]* generuje reťazec: *[[[D*.

Reťazec dátového typu metódy sa skladá z reťazcov pre dátový typ parametrov, ohraničených v zátvorkách (*P**) a reťazca pre dátový typ návratovej hodnoty *R*. Tvar reťazca dátového typu metódy je potom (*P**)*R*. V prípade návratovej hodnoty *null* je reťazcom návratovej hodnoty znak *V*. Napríklad metódu *boolean long pow(int n, int k)* reprezentuje reťazec: *(II)I*, v prípade metódy *Object method(byte b)* by šlo o reťazec: *(B)Ljava/lang/Object;*. Komplexný prehľad reprezentácie dátových typov je uvedený v prílohe A.3.

2.1.2 Premenné tried a inštancií

Premenné tried inštancií (*fields*) *class* súboru sú v poli *fields[]* reprezentované pomocou štruktúry *field_info*. Formát štruktúry *field_info* je nasledovný:

Štruktúra *field_info*

```
field_info {
    u2 access_flags;
    u2 name_index;
    u2 descriptor_index;
    u2 attributes_count;
    attribute_info attributes[attributes_count];
}
```

Položka *access_flags* je indikátorom oprávnenia prístupu k danej premennej. Mená indikátorov spolu s ich interpretáciou a hodnotou sú uvedené v prílohe A.4.

Dvoj bajtová hodnota *name_index* je index *constant_pool[]* reprezentujúci meno premennej

Podobne ako *name_index* aj *descriptor_index* je dvojbajtová položka odkazujúca sa na štruktúru v *constant_pool*. Na rozdiel od mena premennej však popisuje datový typ premennej. Reprezentáciou datových typov sa zaoberá kapitola 2.1.1.

Položka *attributes_count* vyjadruje počet atribútov v poli *attributes[]*.

Pole *attributes[]* môže obsahovať ľubovoľné množstvo atribútov popisujúcich premennú. Štruktúra reprezentujúca atribút je daná všeobecným predpisom *attributeq_info*. Atribúty premenných musia byť reprezentované jednou zo štruktúr *ConstantValue*, *Synthetic*, *Signature*, *Deprecated*, *RuntimeVisibleAnnotations*, alebo *RuntimeInvisibleAnnotations*. Atribút *ConstantValue* popisuje konštantné statické premenné, *Synthetic* je používaný u položiek, ktoré sa nevyskytujú v zdrojovom kóde. Štruktúrou *attribute_info* sa zaoberá kapitola 2.1.4.

2.1.3 Metódy

Každá metóda triedy, prípadne rozhrania je v poli *methods[]* uložená pomocou štruktúry *method_info*. Štruktúra *method_info* má nasledujúci formát:

Štruktúra *method_info*

```
method_info {
    u2 access_flags;
    u2 name_index;
    u2 descriptor_index;
    u2 attributes_count;
    attribute_info attributes[attributes_count];
}
```

Indikátor *access_flags* zahŕňa nastavenia prístupových práv a vlastností metódy. Tabuľka indikátorov *access_flags* štruktúry *method_info* sa nachádza v prílohe A.5.

Položky *name_index* a *descriptor_index* sú podobne ako u štruktúry *field_info* indexmi do *constant_pool*. Tieto indexy v *constant_pool* odkazujú na štruktúry popisujúce meno a datový typ metódy. Reprezentácia dátových typov je popísaná v kapitole 2.1.1.

Hodnotou položky *attributes_count* je počet atribútov poľa *attributes[]*.

Pole *attributes[]* zahŕňa dodatočné atribúty (položky) danej metódy. Každá položka poľa je reprezentovaná všeobecným predpisom *attributes_info*. Počet štruktúr v poli nieje obmedzený, každá položka však musí byť jednou zo štruktúr: *Code*, *Exceptions*, *Synthetic*, *Signature*, *Deprecated*, *RuntimeVisibleAnnotations*, *RuntimeInvisibleAnnotations*, *RuntimeVisibleParameterAnno-*

tations, *RuntimeInvisibleParameterAnnotations*, alebo *AnnotationDefault*. Atribút *Code* je jedným z najdôležitejších. Obsahuje inštrukcie bajtkódu popisujúce fungovanie metódy. Okrem metód deklarovaných ako abstraktná, alebo natívna musí každá metóda obsahovať práve jeden atribút *Code*. Atribút *Exceptions* zahŕňa indexy výnimiek, ktoré metóda vyhadzuje. Popisom formátu štruktúry *attributes_info* sa zaoberá kapitola 2.1.4.

2.1.4 Atribúty

Pojem atribút v tomto texte vyjadruje atribúty používané v poli *attributes[]* štruktúr *field_info*, *method_info* a *Code_attributes*. Všeobecný predpis všetkých atribútov je vyjadrený štruktúrou *attribute_info*. Existuje niekoľko základných preddefinovaných atribútov: *SourceFile*, *ConstantValue*, *Code*, *Exceptions*, *InnerClasses*, *Synthetic*, *LineNumberTable*, *LocalVariableTable*, *Deprecated* a iné. Líšia sa funkcionalitou a využitím jednotlivými časťami *class* súboru. Všetky atribúty vychádzajú z už spomínaného všeobecného predpisu *attribute_info*, ktorý má nasledujúci formát:

Štruktúra *attribute_info*

```
attribute_info {
    u2 attribute_name_index;
    u4 attribute_length;
    u1 info[attribute_length];
}
```

Položka *attributes_name_index* je indexom do *constant_pool* odkazujúcim na meno atribútu.

Štvorbajtová položka *attribute_length* je rovná hodnote vyjadrujúcej dĺžku následných informácií uložených v *info[attribute_length]*. Informácie sa líšia na základe odlišnej funkcionality a využitia jednotlivých atribútov.

2.2 Inštrukčná sada bajtkódu

TODO...

Literatúra

- [LYBB13] Tim Lindholm, Frank Yellin, Gilad Bracha, and Alex Buckley. *The Java Virtual Machine Specification, Java SE 7 Edition*. Addison-Wesley Professional, 1st edition, 2013. <http://docs.oracle.com/javase/specs/jvms/se7/html/>.

Dodatok A

Tabuľky

Zdrojom nasledujúcich tabuliek je špecifikácia JVM JVM [LYBB13].

Constant Type	Value
CONSTANT_Class	7
CONSTANT_Fieldref	9
CONSTANT_Methodref	10
CONSTANT_InterfaceMethodref	11
CONSTANT_String	8
CONSTANT_Integer	3
CONSTANT_Float	4
CONSTANT_Long	5
CONSTANT_Double	6
CONSTANT_NameAndType	12
CONSTANT_Utf8	1
CONSTANT_MethodHandle	15
CONSTANT_MethodType	16
CONSTANT_InvokeDynamic	18

Tabuľka A.1: Tabuľka značiek určujúcich typ záznamu v *constant_pool*. Stĺpec *Constant Type* označuje názov typu, stĺpec *value* priradí uje každému typu číselnú hodnotu.

Meno Indikátora	Hodnota	Interpretácia
ACC_PUBLIC	0x0001	Deklarovaná ako verejná; prístupná aj mimo balíka.
ACC_FINAL	0x0010	Deklarovaná ako final; žiadne podtriedy po inicializácii.
ACC_SUPER	0x0020	Volá metódu nadtriedy, hlavne inštrukcia invokespecial.
ACC_INTERFACE	0x0200	Je rozhranie, nie trieda.
ACC_ABSTRACT	0x0400	Deklarovaná ako abstraktná, nemôže byť inštanciovaná.
ACC_SYNTHETIC	0x1000	Deklarovaná ako synthetic, nieje prítomná v zdrojovom kóde.
ACC_ANNOTATION	0x2000	Deklarovaná ako typ annotation.
ACC_ENUM	0x4000	Deklarovaná ako typ enum.

Tabuľka A.2: Tabuľka indikátorov prístupových práv *ClassFile* štruktúry.

Reprezentácia pomocou reťazca	Typ	Interpretácia
B	byte	znamienkové celé číslo veľkosti jedného bajtu
C	char	Znak s kódovaním UTF-16
D	double	číselná hodnota s dvojitou presnosťou a plávajúcou desatinnou čiarkou
F	float	číselná hodnota s plávajúcou desatinnou čiarkou
I	int	celé číslo
J	long	celé číslo väčšieho rozsahu
L ClassName ;	referencia	inštancia triedy ClassName
S	short	znamienkové celé číslo krátkeho rozsahu
Z	boolean	pravda alebo nepravda
[reference	jednorozmerné pole

Tabuľka A.3: Tabuľka reprezentácie datových typov pre premenné.

Meno Indikátora	Hodnota	Interpretácia
ACC_PUBLIC	0x0001	Deklarovaná ako verejná; prístupná aj mimo balíka.
ACC_PRIVATE	0x0002	Deklarovaná ako privátna; použiteľná len v rámci triedy, v ktorej bola definovaná.
ACC_PROTECTED	0x0004	Deklarovaná ako protected; prístupná aj podtriedam.
ACC_STATIC	0x0008	Deklarovaná ako statická.
ACC_FINAL	0x0010	Deklarovaná ako final; žiadne ďalšie priradenia po inicializácii.
ACC_VOLATILE	0x0040	Deklarovaná ako volatile; nemôže byť uložená do medzipamäte.
ACC_TRANSIENT	0x0080	Deklarovaná ako transient; nieje čítaná ani modifikovaná objektovým manažérom.
ACC_SYNTHETIC	0x1000	Deklarovaná ako synthetic, nieje prítomná v zdrojovom kóde.
ACC_ENUM	0x4000	Deklarovaná ako prvok objektu enum

Tabuľka A.4: Tabuľka indikátorov prístupových práv a vlastností štruktúry *field_info*.

Meno Indikátora	Hodnota	Interpretácia
ACC_PUBLIC	0x0001	Deklarovaná ako verejná; prístupná aj mimo balíka.
ACC_PRIVATE	0x0002	Deklarovaná ako privátna; použiteľná len vrámci triedy, v ktorej bola definovaná.
ACC_PROTECTED	0x0004	Deklarovaná ako protected; prístupná aj podtriedam.
ACC_STATIC	0x0008	Deklarovaná ako statická.
ACC_FINAL	0x0010	Deklarovaná ako final; nemôže byť prepísaná.
ACC_SYNCHRONIZED	0x0020	Deklarovaná ako synchronized; pri volaní je zabalená za použitia monitora.
ACC_BRIDGE	0x0040	Bridge metóda; je generovaná prekladačom.
ACC_VARARGS	0x0080	Deklarovaná s dynamickým počtom argumentov.
ACC_NATIVE	0x0100	Deklarovaná ako natívna; implementovaná v inom jazyku ako Java.
ACC_ABSTRACT	0x0400	Deklarovaná ako abstraktná, nieje implementovaná.
ACC_STRICT	0x0800	Deklarovaná ako strictfp, výpočty s plávajúcou čiarkou sú FP - strict.
ACC_SYNTHETIC	0x1000	Deklarovaná ako synthetic, nieje prítomná v zdrojovom kóde.

Tabuľka A.5: Tabuľka indikátorov prístupových práv a vlastností štruktúry *method_info*.