|  |  |
| --- | --- |
| etf_logo_2010_cir | **УНИВЕРЗИТЕТ У БЕОГРАДУ**  **ЕЛЕКТРОТЕХНИЧКИ ФАКУЛТЕТ**  Програмски преводиоци 1 – 2019/2020.  БулеваркраљаАлександра 73, ПФ 35-54, 11120 Београд, Србија  телефон: 011/3218-321, [dekanat@etf.bg.ac.rs](mailto:dekanat@etf.bg.ac.rs) |

ПРОЈЕКТНИ ЗАДАТАК

***КОМПАЈЛЕР ЗА МИКРОЈАВУ***

Програмски преводиоци 1

Студент:

Богдан Мићовић, 2016/0106

Београд, август 2020.

Садржај

[О пројектном задатку 3](#_Toc29682049)

[Опис команди за генерисање јава кода алатима, превођење кода компајлером и покретање 3](#_Toc29682050)

[Опис новоуведених класа 4](#_Toc29682051)

[Опис тест примера 4](#_Toc29682052)

# О пројектном задатку

Израда решења пројектног задатка урађена је под оперативним системом Windows у програмском језику Java. Циљ пројектног задатка је рализација компајлера за програмски језик Микројаву. Компајлер омогућава превођење синтаксно и семантички исправних Микројава програма у Микројава бајткод који се извршава на виртуелној машини за Микројаву. Да ли је код синтаксно и семантички исправан одређује се на основу спецификације која је дефинисана поставком пројектног задатка.

Програмски преводилац има четири основне функционалности: лексичку анализу, синтаксну анализу, семантичку анализу и генерисање кода.

Лексички анализатор треба да препознаје језичке лексеме и да врати скуп токена изднојених из изворног кода, који се даље разматрају у оквиру синтаксне анализе.

Синтаксни анализатор има задатак да утврди да ли издојени токени могу да формирају граматички исправне сентенце. Као резултат добија се апстаракно синтаксно стабло, на основу граматике која се имплементира у овој фази.

Семантички анализатор се формира на основу добијеног апстарактног синтаксног стабла. Спроводи се имплементацијом метода за посећивање чворова у оквиру који се испитује одговарајућа семантичка правила.

Генератор кода преводи синтаксно и семантички исправне програме у извршни облик за одабрано извршно окружење Микројава ВМ.

# Опис команди за генерисање јава кода алатима, превођење кода компајлером и покретање

За покретање кода користи се класа Run из библиотеке mj-runtime-1.1.jar и као аргумент јој се прослеђује путања до одговарајућег .obj фајла који се тестира. Да би се генерисао потребан .obj фајл користи се додатно имплементирана класа Compiler. Пре покретање наведене класе неоходно је покренути build.xml који ће уз помоћ алата (JFlex, AST-CUP)изгенерисати неопходне јава класе, које уз помоћ осталих, додатно имплементираних класа покрећу цео код.

build.xml се састоји од циљева који међусобно могу али не морају бити зависни, у оквиру којих се позивају претходно наведини алата и дефинише место на којем треба да се смести резултат њиховог извршавања.

# Опис новоуведених класа

Новоуведене класе Compiler, MJParserTest, MJParserTestSintaksa и МЈTest слиже за тестирање генерисања кода, семантичке анализе, синтаксне анализе и лексичке анализе респективно.

Поред њих као нове класе јављају се CodeGeneratorи SemanticAnalyzerу оквиру којих се преклапањем visit метода одговарајућих чворава врши генерисање кода и семантичка анализа.

# Опис тест примера

* test301.mj

Користи се при провери функционалности А фазе

* test302.mj

Користи се за проверу функционалности Б фазе

Vezano za kod:

Designator nigde ne postavlja svoju vrednost tj. adresu ako je to niz vec to rade za njega roditeljske smene gde ima dodela i Loperation smena u koliko je element niza i to rade samo Assign smene ali za pocetnu vrednost dodele kod Designator Assignop Expr pa za taj levi Designator

Za ostale Designatore koji su u izrazima oni pripadaju FactorDesignator smeni i on ih ucitava

Imaju 2 vrste Designatora - one u Expru i one van njega

oni u njemu imaju Kind = Elem a ovi van Kind = Var Type = array

kod a = new int[5] pa ovo a ili a = 3 + 2 \*5 tj. svuda ged je a sa leve strane

Smene MyAssignop(=) i MyAssignopRight(\*=,+=,...) imaju zadatak i da ucitaju na stek vrednost levog operanda npr a = 5; ili a+=5 ili a[5] +=3;

MjParser:

Ako ima smena koja nema | moze da ima isto ime u zagradama sa desne strane ::= kao i sa leve strane, sta vise i treba da ima

Ne zaboraviti da ne smes da imas 2 iste smene

Ne zaboraviti da se doda nova smena kao neterminal u .cup fajlu

Prevodjenje:

Kada se doda novi simbol u lexeru, onda mora i u parseru da se doda taj terminal i prevodi se prvo repackage, pa lexer gen pa compile u build fajlu

ascii code 0 **NULL** (Null character)

ascii code 1 **SOH** (Start of Header)

ascii code 2 **STX** (Start of Text)

ascii code 3 **ETX** (End of Text)

ascii code 4 **EOT** (End of Transmission)

ascii code 5 **ENQ** (Enquiry)

ascii code 6 **ACK** (Acknowledgement)

ascii code 7 **BEL** (Bell)

ascii code 8 **BS** (Backspace)

ascii code 9 **HT** (Horizontal Tab)

ascii code 10 **LF** (Line feed)

ascii code 11 **VT** (Vertical Tab)

ascii code 12 **FF** (Form feed)

ascii code 13 **CR** (Carriage return)

ascii code 14 **SO** (Shift Out)

ascii code 15 **SI** (Shift In)

ascii code 16 **DLE** (Data link escape)

ascii code 17 **DC1** (Device control 1)

ascii code 18 **DC2** (Device control 2)

ascii code 19 **DC3** (Device control 3)

ascii code 20 **DC4** (Device control 4)

ascii code 21 **NAK** (Negative acknowledgement)

ascii code 22 **SYN** (Synchronous idle)

ascii code 23 **ETB** (End of transmission block)

ascii code 24 **CAN** (Cancel)

ascii code 25 **EM** (End of medium)

ascii code 26 **SUB** (Substitute)

ascii code 27 **ESC** (Escape)

ascii code 28 **FS** (File separator)

ascii code 29 **GS** (Group separator)

ascii code 30 **RS** (Record separator)

ascii code 31 **US** (Unit separator)

ascii code 32 (space)

ascii code 33 **!** (exclamation mark)

ascii code 34 **"** (Quotation mark)

ascii code 35 **#** (Number sign)

ascii code 36 $ (Dollar sign)

ascii code 37 % (Percent sign)

ascii code 38 & (Ampersand)

ascii code 39 ' (Apostrophe)

ascii code 40 ( (round brackets or parentheses)

ascii code 41 ) (round brackets or parentheses)

ascii code 42 \* (Asterisk)

ascii code 43 + (Plus sign)

ascii code 44 , (Comma)

ascii code 45 - (Hyphen)

ascii code 46 . (Full stop , dot)

ascii code 47 / (Slash)

ascii code 48 0 (number zero)

ascii code 49 1 (number one)

ascii code 50 2 (number two)

ascii code 51 3 (number three)

ascii code 52 4 (number four)

ascii code 53 5 (number five)

ascii code 54 6 (number six)

ascii code 55 7 (number seven)

ascii code 56 8 (number eight)

ascii code 57 9 (number nine)

ascii code 58 : (Colon)

ascii code 59 ; (Semicolon)

ascii code 60 < (Less-than sign )

ascii code 61 = (Equals sign)

ascii code 62 > (Greater-than sign ; Inequality)

ascii code 63 ? (Question mark)

ascii code 64 @ (At sign)

ascii code 65 A (Capital A )

ascii code 66 B (Capital B )

ascii code 67 C (Capital C )

ascii code 68 D (Capital D )

ascii code 69 E (Capital E )

ascii code 70 F (Capital F )

ascii code 71 G (Capital G )

ascii code 72 H (Capital H )

ascii code 73 I (Capital I )

ascii code 74 J (Capital J )

ascii code 75 K (Capital K )

ascii code 76 L (Capital L )

ascii code 77 M (Capital M )

ascii code 78 N (Capital N )

ascii code 79 O (Capital O )

ascii code 80 P (Capital P )

ascii code 81 Q (Capital Q )

ascii code 82 R (Capital R )

ascii code 83 S (Capital S )

ascii code 84 T (Capital T )

ascii code 85 U (Capital U )

ascii code 86 V (Capital V )

ascii code 87 W (Capital W )

ascii code 88 X (Capital X )

ascii code 89 Y (Capital Y )

ascii code 90 Z (Capital Z )

ascii code 91 [ (square brackets or box brackets)

ascii code 92 \ (Backslash)

ascii code 93 ] (square brackets or box brackets)

ascii code 94 ^ (Caret or circumflex accent)

ascii code 95 \_ (underscore, under strike, underbar or low line)

ascii code 96 ` (Grave accent)

ascii code 97 a (Lowercase a )

ascii code 98 b (Lowercase b )

ascii code 99 c (Lowercase c )

ascii code 100 d (Lowercase d )

ascii code 101 e (Lowercase e )

ascii code 102 f (Lowercase f )

ascii code 103 g (Lowercase g )

ascii code 104 h (Lowercase h )

ascii code 105 **i** (Lowercase i )

ascii code 106 j (Lowercase j )

ascii code 107 k (Lowercase k )

ascii code 108 l (Lowercase l )

ascii code 109 m (Lowercase m )

ascii code 110 n (Lowercase n )

ascii code 111 o (Lowercase o )

ascii code 112 p (Lowercase p )

ascii code 113 q (Lowercase q )

ascii code 114 r (Lowercase r )

ascii code 115 s (Lowercase s )

ascii code 116 t (Lowercase t )

ascii code 117 u (Lowercase u )

ascii code 118 v (Lowercase v )

ascii code 119 w (Lowercase w )

ascii code 120 x (Lowercase x )

ascii code 121 y (Lowercase y )

ascii code 122 z (Lowercase z )

ascii code 123 { (curly brackets or braces)

ascii code 124 | (vertical-bar, vbar, vertical line or vertical slash)

ascii code 125 } (curly brackets or braces)

ascii code 126 ~ (Tilde ; swung dash)

ascii code 127 DEL (Delete)

ascii code 128 Ç (Majuscule C-cedilla )

ascii code 129 ü (letter "u" with umlaut or diaeresis ; "u-umlaut")

ascii code 130 é (letter "e" with acute accent or "e-acute")

ascii code 131 â (letter "a" with circumflex accent or "a-circumflex")

ascii code 132 ä (letter "a" with umlaut or diaeresis ; "a-umlaut")

ascii code 133 à (letter "a" with grave accent)

ascii code 134 å (letter "a" with a ring)

ascii code 135 ç (Minuscule c-cedilla)

ascii code 136 ê (letter "e" with circumflex accent or "e-circumflex")

ascii code 137 ë (letter "e" with umlaut or diaeresis ; "e-umlaut")

ascii code 138 è (letter "e" with grave accent)

ascii code 139 ï (letter "i" with umlaut or diaeresis ; "i-umlaut")

ascii code 140 î (letter "i" with circumflex accent or "i-circumflex")

ascii code 141 ì (letter "i" with grave accent)

ascii code 142 Ä (letter "A" with umlaut or diaeresis ; "A-umlaut")

ascii code 143 Å (letter "A" with a ring)

ascii code 144 É (Capital letter "E" with acute accent or "E-acute")

ascii code 145 æ (Latin diphthong "ae")

ascii code 146 Æ (Latin diphthong "AE")

ascii code 147 ô (letter "o" with circumflex accent or "o-circumflex")

ascii code 148 ö (letter "o" with umlaut or diaeresis ; "o-umlaut")

ascii code 149 ò (letter "o" with grave accent)

ascii code 150 û (letter "u" with circumflex accent or "u-circumflex")

ascii code 151 ù (letter "u" with grave accent)

ascii code 152 ÿ (letter "y" with diaeresis)

ascii code 153 Ö (letter "O" with umlaut or diaeresis ; "O-umlaut")

ascii code 154 Ü (letter "U" with umlaut or diaeresis ; "U-umlaut")

ascii code 155 ø (slashed zero or empty set)

ascii code 156 £ (Pound sign ; symbol for the pound sterling)

ascii code 157 Ø (slashed zero or empty set)

ascii code 158 × (multiplication sign)

ascii code 159 ƒ (function sign ; f with hook sign ; florin sign )

ascii code 160 á (letter "a" with acute accent or "a-acute")

ascii code 161 í (letter "i" with acute accent or "i-acute")

ascii code 162 ó (letter "o" with acute accent or "o-acute")

ascii code 163 ú (letter "u" with acute accent or "u-acute")

ascii code 164 ñ (letter "n" with tilde ; enye)

ascii code 165 Ñ (letter "N" with tilde ; enye)

ascii code 166 ª (feminine ordinal indicator )

ascii code 167 º (masculine ordinal indicator)

ascii code 168 ¿ (Inverted question marks)

ascii code 169 ® (Registered trademark symbol)

ascii code 170 ¬ (Logical negation symbol)

ascii code 171 ½ (One half)

ascii code 172 ¼ (Quarter or one fourth)

ascii code 173 ¡ (Inverted exclamation marks)

ascii code 174 « (Guillemets or angle quotes)

ascii code 175 » (Guillemets or angle quotes)

ascii code 176 ░

ascii code 177 ▒

ascii code 178 ▓

ascii code 179 │ (Box drawing character)

ascii code 180 ┤ (Box drawing character)

ascii code 181 Á (Capital letter "A" with acute accent or "A-acute")

ascii code 182 Â (letter "A" with circumflex accent or "A-circumflex")

ascii code 183 À (letter "A" with grave accent)

ascii code 184 © (Copyright symbol)

ascii code 185 ╣ (Box drawing character)

ascii code 186 ║ (Box drawing character)

ascii code 187 ╗ (Box drawing character)

ascii code 188 ╝ (Box drawing character)

ascii code 189 ¢ (Cent symbol)

ascii code 190 ¥ (YEN and YUAN sign)

ascii code 191 ┐ (Box drawing character)

ascii code 192 └ (Box drawing character)

ascii code 193 ┴ (Box drawing character)

ascii code 194 ┬ (Box drawing character)

ascii code 195 ├ (Box drawing character)

ascii code 196 ─ (Box drawing character)

ascii code 197 ┼ (Box drawing character)

ascii code 198 ã (letter "a" with tilde or "a-tilde")

ascii code 199 Ã (letter "A" with tilde or "A-tilde")

ascii code 200 ╚ (Box drawing character)

ascii code 201 ╔ (Box drawing character)

ascii code 202 ╩ (Box drawing character)

ascii code 203 ╦ (Box drawing character)

ascii code 204 ╠ (Box drawing character)

ascii code 205 ═ (Box drawing character)

ascii code 206 ╬ (Box drawing character)

ascii code 207 ¤ (generic currency sign )

ascii code 208 ð (lowercase "eth")

ascii code 209 Ð (Capital letter "Eth")

ascii code 210 Ê (letter "E" with circumflex accent or "E-circumflex")

ascii code 211 Ë (letter "E" with umlaut or diaeresis ; "E-umlaut")

ascii code 212 È (letter "E" with grave accent)

ascii code 213 ı (lowercase dot less i)

ascii code 214 Í (Capital letter "I" with acute accent or "I-acute")

ascii code 215 Î (letter "I" with circumflex accent or "I-circumflex")

ascii code 216 Ï (letter "I" with umlaut or diaeresis ; "I-umlaut")

ascii code 217 ┘ (Box drawing character)

ascii code 218 ┌ (Box drawing character)

ascii code 219 █ (Block)

ascii code 220 ▄

ascii code 221 ¦ (vertical broken bar )

ascii code 222 Ì (letter "I" with grave accent)

ascii code 223 ▀

ascii code 224 Ó (Capital letter "O" with acute accent or "O-acute")

ascii code 225 ß (letter "Eszett" ; "scharfes S" or "sharp S")

ascii code 226 Ô (letter "O" with circumflex accent or "O-circumflex")

ascii code 227 Ò (letter "O" with grave accent)

ascii code 228 õ (letter "o" with tilde or "o-tilde")

ascii code 229 Õ (letter "O" with tilde or "O-tilde")

ascii code 230 µ (Lowercase letter "Mu" ; micro sign or micron)

ascii code 231 þ (capital letter "Thorn")

ascii code 232 Þ (lowercase letter "thorn")

ascii code 233 Ú (Capital letter "U" with acute accent or "U-acute")

ascii code 234 Û (letter "U" with circumflex accent or "U-circumflex")

ascii code 235 Ù (letter "U" with grave accent)

ascii code 236 ý (letter "y" with acute accent)

ascii code 237 Ý (Capital letter "Y" with acute accent)

ascii code 238 ¯ (macron symbol)

ascii code 239 ´ (Acute accent)

ascii code 240 ¬ (Hyphen)

ascii code 241 ± (Plus-minus sign)

ascii code 242 ‗ (underline or underscore)

ascii code 243 ¾ (three quarters)

ascii code 244 ¶ (paragraph sign or pilcrow)

ascii code 245 § (Section sign)

ascii code 246 ÷ (The division sign ; Obelus)

ascii code 247 ¸ (cedilla)

ascii code 248 ° (degree symbol )

ascii code 249 ¨ (Diaeresis)

ascii code 250 • (Interpunct or space dot)

ascii code 251 **¹** (superscript one)

ascii code 252 **³** (cube or superscript three)

ascii code 253 **²** (Square or superscript two)

ascii code 254 **■** (black square)

ascii code 255 **nbsp** (non-breaking space or no-break space)