# תורת הקומפילציה

Flex תרגיל בית 1 – בניית מנתח לקסיקלי באמצעות מתרגל אחראי לתרגיל: אדם בוטח

יש להפנות שאלות על התרגיל במייל ל- <u>botach@campus.technion.ac.il</u> עם "236360" בתחילת כותרת ההודעה.

ההגשה ביחידים או בזוגות.

לתרגיל ייפתח דף FAQ באתר הקורס. כל הבהרה שתופיע בו עד יומיים ממועד ההגשה תהווה הוראה מחייבת.

# הנחיות כלליות

- בתרגיל זה תממשו מנתח לקסיקלי עבור <u>וריאציה מצומצמת</u> של השפה Swift.
   השפה פותחה ע"י חברת Apple ושוחררה לראשונה בשנת 2014 בכדי להציע אלטרנטיבה מודרנית לשפה Objective-C כשפת הפיתוח המרכזית עבור מרבית הפלטפורמות של החברה (מכשירי iOS, iPadOS, MacOS) וכו׳).
- נחוצים MacOS אינם השפה ו/או שימוש במחשב המריץ לבות עם השפה ו/או שימוש במחשב המריץ לצורך פתרון התרגיל.
- התרגיל ייבדק אוטומטית. **הקפידו למלא אחר ההוראות במדויק.** הבדיקה תתבצע על csComp.
  - יש להשתמש ב-flex בלבד (ולא ב-lex).



#### הגדרת מושגים כלליים

- רווח לבן: אחד מבין: רווח (ספייס), טאב, CR (התו LF,(\r והתו n)).
- LF ,(0x09), או רווחים לבנים: טאב (0x09), או רווחים לבנים: טאב (0x09), LF ,(0x09), או רווחים לבנים: טאב (0x09), (0x0E) (0x0D) (0x0D) (0x0D) (0x0D) (0x0D) (0x0D)
  - ניתן לקרוא על תווים ניתנים להדפסה בהרחבה בוויקיפדיה בערך הבא: o <a href="https://en.wikipedia.org/wiki/ASCII#Printable">https://en.wikipedia.org/wiki/ASCII#Printable</a> characters
  - רובף מילוט (escape sequence): לוכסן אחורי (התו \) ואחריו תו או יותר שביחד מפורשים כתו אחר.
    - טאב.  $\t$  ירידת שורה, \n טאב.
    - ניתן לקרוא על רצפי מילוט בהרחבה בוויקיפדיה בערך הבא: https://en.wikipedia.org/wiki/Escape sequences in C
- רצף מילוט של תו ASCII¹ רצף בעל התבנית הבאה: \u{n} כאשר n הינו מספר בייצוג הקסדצימלי (המורכב ASCII מהספרות 0-9 והאותיות a-f גדולות/קטנות) באורך 1 עד 6 תווים. הרצף מהווה מילוט לתו שערך ה-u{II שלו בייצוג הקסדצימלי הינו n. דוגמא: \u{41} הינו רצף מילוט עבור התו 'A'.
   הערה: בהמשך יוסבר אופן הטיפול בערך ascii שאינו חוקי או אינו ניתן להדפסה.

בטבלה בעמוד הבא מפורטים האסימונים (tokens) שאיתם נעבוד בתרגיל וההגדרות שלהם.

שימו לב – אלא אם כן נכתב במפורש אחרת כל הגדרות האסימונים הינן case-sensitive.

ASCII משמשת לייצוג תווי Swift בתרגיל זה אנו נתמוך באופן חלקי בתווי  $\ensuremath{\mbox{\mbox{$\sc u$}}} \ensuremath{\mbox{\mbox{$\sc u$}}} \ensuremath{\mbo$ 

# <u>הגדרת האסימונים</u>

דוגמאות (לקסמות המתאימות לאסימון זה)	ערכים אפשריים	תיאור	שם האסימון
a _compi _Z123ab _cC1aA	מתחיל באות (גדולה או קטנה) <u>או</u> קו תחתון, ואחריה רצף של אותיות (גדולות או קטנות) או ספרות. שימו לב – אם התו הראשון הינו קו תחתון הרצף שמגיע אחריו <u>אינו ריק</u> .	שם של מזהה	ID
0b1010 0b0	מספר המתחיל בתווים 0b ואחריהם רצף <u>לא</u> <u>ריק</u> המורכב מהספרות 0 או 1.	ערך מספרי שלם <u>אי-שלילי</u> בבסיס בינארי (בסיס 2)	BIN_INT
0o700	מספר המתחיל בתווים 00 ואחריהם רצף <u>לא</u> <u>ריק</u> המורכב מהספרות 0-7.	ערך מספרי שלם <u>אי-שלילי</u> בבסיס אוקטלי (בסיס 8)	OCT_INT
123 0 42	מספר המורכב מספרה דצימלית ( <u>אחת או</u> <u>יותר</u> ).	ערך מספרי שלם <u>אי-שלילי</u> בבסיס דצימלי (בסיס 10)	DEC_INT
0x10 0x7FFFFFF 0xFf	מספר המתחיל בתווים 0x ואחריהם רצף לא ריק של האותיות a-f (גדולות או קטנות) והספרות 0-9.	ערך מספרי שלם <u>אי-שלילי</u> בבסיס הקסדצימלי (בסיס 16)	HEX_INT
0.0 .2 8. 42.125 1.e-8 1.E+10	אחד משני הבאים: 1. מספר המורכב מספרה דצימלית אחת או יותר, עם נקודה עשרונית אחת לפניו, אחריו או בתוכו. 2. מספר ממשי כמתואר בסעיף 1 ובתוספת אקספוננט בסופו. האקספוננט מורכב מהאות e (גדולה או קטנה), אחריה סימן + או – ואז מספר שלם בייצוג דצימלי.	מספר ממשי <u>אי-</u> <u>שלילי</u> בבסיס דצימלי	DEC_REAL
0xFp-2 0x6P+3	מספר המתחיל בתווים $0x$ , אחריהם רצף לא ריק המורכב מהאותיות $a$ - $f$ , ואז אקספוננט קטנות) והספרות $0$ - $f$ , ואז אקספוננט המורכב מהאות $p$ (גדולה או קטנה), אחריה $p$ סימן $p$ או - ואז מספר שלם בייצוג $p$ סימן $p$ או - ואז מספר שלם בייצוג $p$ סימן $p$ או $p$ הלקסמה $p$	מספר ממשי <u>אי-</u> <u>שלילי</u> בבסיס הקסדצימלי בייצוג Floating Point	HEX_FP
"this is a \t legal string \n"  "also legal \n\r\\"  "we can \"quote\" inside!"  "\u{54}\u{69}\u{4D}\u{65}"  "\u{57}\u{61}\u{53}\u{74}\u{45}\u{64}"	אוסף תווים ניתנים להדפסה באורך אפס או יותר, בתוך מרכאות כפולות, בשורה אחת.  1. צירוף האותיות n\ יכול להופיע (כשני תווים, לוכסן אחורי ואחריו האות n) במחרוזת, אך ירידת השורה עצמה כתו אחד אינה יכולה להופיע. כנ״ל עבור r\. 2. המחרוזת אינה יכולה להכיל מרכאות כפולות ״ ולוכסן אחורי \ אלא אם הם חלק מרצף מילוט חוקי כמפורט בסעיף הבא.	ליטרל המייצג מחרוזת	STRING

מהתרות תומנת מרצפי מילוט באופן מור למתוב   מרצי את התויח משתחר מדי   מרצי את מדי מו הביא מן מימו לב: כל רצי מילוט שאינו בהשם, בחלק של התרוזת בלי ברשימת הבתחלה ובסוף לא עולה על   מרצי המילוט יוסבר   מרצי מדי מו מרצי מדי מדי מרצי מדי מו מרצי מדי מדי מדי מדי מדי מדי מדי מדי מדי מד	=	=	מציין השמה	ASSIGN
אחר הווים שמותר לכתוב אחר היוס (מאחרי בתורות ונתמוך אחר היוס (מאחרי בתורות ונתמוך ר   י   בששת המקרים הבאים בלבד: אופן הטיפול ברצפי המילוט יוסבר בפשת המקרים הבאים בלבד: ממוס ברצי (מיש של תווים. אופן הטיפול ברצפי המילוט יוסבר בחימה הכיל אינו מהווה קלט חוקי. שמול בל כרצף מילוט שאינו בהתחלה ובסוף לא עולה על אניתן להכיח שהאורך של מחדות בה'י שמול בל כרצי מילוט ובסוף לא עולה על אחרון הבאים (שימו לב שהשמות הינם וחד מן בהצימ (שימו לב שהשמות הינם וחד מן בהצימ (שימו לב שהשמות הינם וחד מן במצימ (שימו לב שהשמות הינם וחד מו במצימ (שימו לב שומו לב שומו לב ומצימ וחדשה	]	]		
אחרי הלוכסן האחורי במחרזות ונתמוך האחרי באוחרי באחרזות ונתמוך האחרי באחרזות ונתמוך האחרי באחרזות ונתמוך   אר האחרי   אר האווי   אר האחרי   אר האווי   אר האוו	[	[	שמאלי	
אחרי הלוכסן האחורי במחרוזת ונתמוך אחרי האחרי במחרוזת ונתמוך האחרי האחרי במחרוזת ונתמוך בששת המקרים הבאים בלבד:    "   "   "   "   "   "   "   "   "	}	}		RBRACE
אחרי הלוכסן האחורי במחרוזת ונתמוך האחרי בשלו האחרי במחרוזת ונתמוך האחרי	{	{		LBRACE
אחרי הלוכסן האחורי במחרוזת ונתמוך אחרי הלוכסן האחורי במחרוזת ונתמוך אחרי הלוכסן האחורי במחרוזת ונתמוך אחרי בששת המקרים הבאים בלבד:  - * * * * * * * * * * * * * * * * * *	)	)	סוגר מעוגל ימני	RPAREN
אופן המקיר את התווים שמותר לכתוב באשר המקרים הבאים בלבד: אחרי הלוכסן האחורי במחרוזת נתמוך   ארי הלוכסן האחורי במחרוזת נתמוך   ארי פישה המקרים הבאים בלבד: אופן הטיפול ברצפי המילוט יוסבר   ברשימה הנ"ל אינו מהווה קלט חומי.   ברשימה הנ"ל אינו מהווים   ברשימה הנ"ל אינו מהווים   ברשימה הנ"ל אינו מהווים   ברשימה הנ"ל אינו מהווים   ברשימה להצהרה   ברשימה להצהרה   ברשימי ליבוא   ברשימה להצהרה   ברשימה להצהרה   ברשימה להצהרה   ברשימה ליבוא   ברשימה להצהרה   ברשימה ליבוא   ברשימה ליבוא אופן ווחסר   ברשימה ליבוא   ברשימה ליבוא אופן ווחסר   ברשימה ליבוא אופן ווחסר   ברשימה ליבוא אופן ווחסר   ברשימי ליבוא   ברשימה ליבוא   ברשימה הוו   ברשימה ליבוא   ברשימה הוו   ברשימה ליבוא   ברשימה ליבוא אוופ   ברשימה ליבוא אוופ   ברשימה ליבוא   ברשימה ליבוא   ברשימה אוופ   ברשימה ליבור   ברשימה ליבוא אוופ   ברשימה ליבור   ברשימה ליבור ליום אומים   ברשימה ליבור ליים אומים   ברשימה ליבור ליים אומים   ברשימה ליבור ליים אומים אומים אומים   ברשימה ליבור ליים אומים אומים   ברשימה ליבור ליים אומים אומי	(	(		LPAREN
אחרי הלוכסן האחורי במחרוזת נתמון אחרי במחרוזת נתמון החלי. נגביל את התווים שמותר לכתוב בששת המקרים הבאים בלבד:  אחרי הלוכסן האחורי במחרוזת נתמון או או או המקרים הבאים בלבד:  מבשת המקרים רצף מילוט של תו "ה"  מבשת המקרים הציים בלבד:  מבישה הני"ן או אופן הטיפול ברצפי המילוט יוסבר בהמשך. בחלק של הדפסת אסימונים. בהמשך. בחלק של הדפסת אסימונים. בהמשך. בחלק של הדפסת אסימונים. ברשימה הני"ל אינו מהווה קלט חוקי.  ברשימה הני"ל אינו מהווה קלט חוקי. ברשימה הני"ל אינו מהווה קלט חוקי. ברשימה הני"ל אינו מהווה קלט חוקי. ברשימה הני"ל אינו מהווה הלינו חוקי. ברשימה הני"ל אינו מהווה הינים וווים. ברשימה לבאבים (שימו לב שהשמות הינים וווים. ברשימה לבאבים (שימו לב שהשמות הינים וווים. במיסי שם של טיפוס (שימו לב שהשמות הינים וווים. במשמש להצהרה על מבוע חדש משמש להצהרה ודש משמש להצהרה ודש משמש להצהרה ודש משמש להצהרה ודש משמש לייבוא וווחסד וווים. במשמש לייבוא שמשל לייבוא וווחסד וווים. במשמש לייבוא שמשל לייבוא שווחסד וווים. במשמש לייבוא שמשל לייבוא שווחסד וווים. במשמש לייבוא הוון הוון הוון הוון הוון הוון הוון הו	,	,		COMMA
אחרי הלוכסן האחורי במחרוזת ונתמוך האחרי במחרוזת ונתמוך האחרי במחרוזת ונתמוך האחרי במחרוזת ונתמוך האחרי הלוכסן האחורי במחרוזת ונתמוך האחרי בשת המקרים הבאים בלבד:  מביש המביע היש היש ליש ליש ליש האחרי של מחרוזת בלי ברצים המילוט שאנו ברמשה ההיל אינו מהווה קלט חוקי. ברמשה המבאורך של מחרוזת בלי המרכאות בהתחלה ובסוף לא עולה על שימו לב שהשמות הינם (case sensitive Int, Ulnt, Double, Float, Bool, String, Character  Var var var שמשש להצהרה על משתבה חדש על קבוע חדש שמש להצהרה (character with a part of the part	;	;	•	SC
אחרי הלפוס האחורי במחרוזת ונתמוך   האחורי במחרוזת ביי   האחורי ביי מילוט של תו ascii   האחופר לעיל.				
אחרי הלוכסן האחורי במחרוזת ונתמוך   אחרי הלוכסן האחורי במחרוזת ונתמוך   בששת המקרים הבאים בלבד:   •     •         •         •				
אחרי הלוכסן האחורי במחרוזת ונתמוך האחרי במחרוזת ונתמוך האחרי במחרוזת ונתמוך בששת המקרים הבאים בלבד:  אחרי הלוכסן האחורי במחרוזת ונתמון אחרי בששת המקרים הבאים בלבד:  אופן הטיפול ברצף מילוט של תו ascii ווויסבר כפי שהוסבר לעיל.  מפי שהוסבר לעיל.  בפי שהוסבר לעיל.  אופן הטיפול ברצף מילוט של תו incar במהשך, בחלק של הדפסת אסימונים. שימו לב: כל רצף מילוט שינו בהתחלה ובסוף לא עולה על ברצי מהרבית בהתחלה ובסוף לא עולה על א ברצי מהרבית בהתחלה ובסוף לא עולה על א עולה על במחרוזת בלי במרכאות בהתחלה ובסוף לא עולה על במחרוזת בלי במיסים במהשות הינם וחדש במיסים במהשות הינם במיסים במהשמות הינם במשם של טיפוס במהשמות הינם במחלה (Case sensitive Int, Ulnt, Double, Float, Bool, String, Character על משתנה חדש על קבוע חדשה במשם להצהרה ודש במשם להצהרה ודשה במשם להצהרה ודשה במשם של לייבוא ווmport import import import import import NIL				
אחרי הלוכסן האחורי במחרוזת ונתמוך אחרי האחרי האחרי בששת המקרים הבאים בלבד:  מצופן הטיפול ברצפי המילוט יוסבר כפי שהוסבר לעיל. אופן הטיפול ברצפי המילוט יוסבר כפי שהוסבר לעיל. במל בצף מילוט שאינו בהמשך, בחלק של הדפסת אסימונים. בשימו לב: כל רצף מילוט שאינו ברשימו לב: כל רצף מילוט שאינו ברשימו הינים ברשימה הנייל אינו מהווה קלט חוקי. ברשימה הנייל אינו מהווה קלט חוקי. אחד מן הבאים (שימו לב שהשמות הינים אחד מן הבאים (שימו לב שהשמות הינים אחד מן הבאים (שימו לב שהשמות הינים ות) אחד מן הבאים (שימו לב שהשמות הינים ות) אחד מן הבאים (שימו לב במפולוני) בסיסי ב				
חלקי. נגביל את התווים שמותר לכתוב אחרי הלוכסן האחורי במחרוזת ונתמוך במחרוזת ונתמוך פששת המקרים הבאים בלבד:  " " " " " " " " " " " " " " " " " " "	•		ספריות	NII
אחרי הלוכסן האחורי במחרוזת ונתמוך אחרי הלוכסן האחורי במחרוזת ונתמוך בששת המקרים הבאים בלבד:  אחרי הלוכסן האחורי במחרוזת ונתמון אר פששת המקרים הבאים בלבד:  מבששת המקרים הבאים בלבד:  מביש הוסבר לעיל.  מביש לו במולה במחלה ובסוף לא עולה על א. ביתן להנית שהאורך של מחרוזת בלי.  מביסי שו מביסי שו מבים (במבי במבים) במבים (בסיסי שו מבים (בסיסי שו מ			משמש לייבוא	
אחרי הלוכסן האחורי במחרוזת ונתמוך בששת המקרים הבאים בלבד: אחרי הלוכסן האחורי במחרוזת ונתמוך בששת המקרים הבאים בלבד: \rangle \	func	func	משמש להגדרת	FUNC
חלקי. נגביל את התווים שמותר לכתוב אחרי הלוכסן האחורי במחרוזת ונתמוך בששת המקרים הבאים בלבד: \n \\ \n \\ \t \\ \t \\ \t \\ \n \\ \t \\ \n	let	let	משמש להצהרה	LET
אחרי הלוכסן האחורי במחרוזת ונתמוך אחרי הלוכסן האחורי במחרוזת ונתמוך בששת המקרים הבאים בלבד:   מ	var	var		VAR
חלקי. נגביל את התווים שמותר לכתוב אחרי הלוכסן האחורי במחרוזת ונתמוך פארוי מקרים הבאים בלבד: ר י מין ר י מיין ר י מין ר י מיין ר י מייין ר י מייין ר י מייין ר י מיייייייייייייייייייייייייייייייייי	Double	תווים. אחד מן הבאים (שימו לב שהשמות הינם case sensitive: Int, UInt, Double, Float, Bool, String,		TYPE
ן לי דומון מור ומחלור בר צב. מטוט באופו		חלקי. נגביל את התווים שמותר לכתוב אחרי הלוכסן האחורי במחרוזת ונתמוך ח   \n \r \r \r \t \t \r \t \r		

== >=	==   !=   <   >   <=   >=		אופרטורים רלטיביים	RELOP
&& 	&&		אופרטורים לוגיים	LOGOP
+ -	+   -   *   /   %		אופרטורים מתמטיים בינאריים	BINOP
true	true		ערך לוגי 1	TRUE
false	false		ערך לוגי 0	FALSE
->	->		מציין ערך החזרה בהגדרת פונקציה	ARROW
:	:		משמש לציון explicit טיפוסים	COLON
/* this is a comment */	אחד מהשניים הבאים: מתחיל ב */ ומסתיים ב /*. בין הפותח	.1		
/* this is also a valid comment. Note that it is spread across multiple lines. */ // this is a single line comment	והסוגר יכול להופיע <u>כל תו</u> שניתן להדפסה (כולל ירידות שורה) פרט לצירוף */.	.2	הערות	COMMENT

# פעולת המנתח והפלט הנדרש

המנתח יתעלם מכל הרווחים הלבנים, חוץ מבתוך מחרוזות.

כאשר המנתח מזהה אסימון, יש לפלוט שורה בפורמט הבא (יש לדאוג לרווח יחיד בין כל רכיב שורה ולירידת שורה ע"ד (n) LF) בלבד לאחר הרכיב האחרון):

line number> <token name> <value>

#### :כאשר

- line number: מספר השורה בה האסימון מסתיים.
- token שם האסימון שזוהה (לפי השמות בחלק "הגדרת אסימונים" למעלה). ●
- value ערך האסימון שזוהה, כלומר הלקסמה, פרט למקרה של <u>מחרוזות, מספרים שלמים והערות</u> כפי שיוסבר בהמשך.

# הדפסת הלקסמה של מחרוזות:

מחרוזות יודפסו <u>ללא</u> המרכאות הכפולות המקיפות אותן. בנוסף נרצה לטפל <u>ברצפי המילוט</u> באופן הבא:

- יוחלפו בסוג המתאים של רווח לבן (טאב, LF ,CR ). (µ,\r,\t
  - \/ יוחלף בלוכסן אחורי יחיד (\).
- <u>רצף מילוט של תו ascii</u>: אם הרצף (המיוצג כאמור ע״י n בתבנית (המוה ייצוג הקסדצימלי <u>חוקי</u> של <u>תו שניתן להדפסה,</u> אז יש להדפיס את התו המתאים במקום רצף המילוט. אחרת, יש להדפיס שגיאה כפי שיפורט בהמשך תחת ׳<u>טיפול בשגיאות</u>׳.

#### דוגמה – המחרוזת הבאה:

 $\u{49}\u{20}\L\u{6F}ve\L\u{00070}i!\nCompi Loves Me\u{21}$ 

תודפס בפורמט הנדרש באופן הבא:

```
1 STRING I "Love" Compi!
Compi Loves Me!
```

1 STRING To Infinity, and Beyond!

# הדפסת הלקסמה של מספרים שלמים:

יש להדפיס את ערכם המספרי של מספרים <u>שלמים</u> (המיוצגים ע״י האסימונים: BIN\_INT, OCT\_INT, DEC\_INT יש להדפיס את ערכם המספרי של מספרים <u>שלמים</u> (המיוצגים ע״י האסימונים: HEX\_INT) <u>כערך דצימלי</u> (יש להדפיס ערכים חיוביים ללא סימן ה-'+').

עבור מספרים ממשיים (המיוצגים ע״י האסימונים DEC REAL, HEX FP) תודפס הלקסמה כפי שהופיעה בקלט.

#### הדפסת הלקסמה של הערות:

במקום תוכן ההערה, יודפס מספר השורות בהערה, כלומר מספר המופעים של ירידות שורה, פלוס 1. זכרו: ירידת שורה עשויה להיות אחד הרצפים הבאים: r\n (\r\n (\r\n (\r\n) CR (\r\n) CR, (\r\n) CRLF), מחשב לירידת שורה אחת.

שימו לב שההוראות האלה מתייחסות להערות <u>משני הסוגים</u> שפורטו לעיל.

#### <u>דוגמה</u>

עבור הקלט:

```
var x : String = "\u{54}o \u{6C}nf\u{69}n\u{69}ty, a\u{6E}d B\u{65}y\u{6F}nd!"
while true {
    print(x)
}
// something important

1 VAR var
1 ID x
1 COLON :
1 TYPE String
1 ASSIGN =
```

2 WHILE while

2 TRUE true

2 LBRACE {

3 ID print

3 LPAREN (

3 ID x

3 RPAREN)

4 RBRACE }

5 COMMENT 1

# הערות נוספות על תווים בקובץ

ניתן להניח כי קבצי הדוגמאות הם קבצי ascii בלבד (כלומר: אינם UTF-8 או UTF-16). בהכינכם קבצי בדיקה, וודאו cascii בי אתם מכוונים את ה-Encoding של הקובץ ל-ASCII או ASCII, או מבצעים save as כי אתם מכוונים את ה-ASCII של הקובץ להלן מפתח של התווים המוזכרים בתרגיל וערכי ה-ASCII שלהם:

(hex) ASCII ערך	סימן	שם	
7B	{	סוגר מסולסל שמאלי	
7D	}	סוגר מסולסל ימני	
3A	:	נקודתיים	
3D	=	שווה	
5C	\	לוכסן אחורי	
3B	•	נקודה פסיק	
2D	-	מינוס / מקף	
2B	+	פלוס	
2C	,	פסיק	
5F		קו תחתון	
2E		נקודה	
22	"	מרכאות כפולות	
0D	CR	Carriage return	
0A	LF	Line feed	
20		רווח	
09		טאב	
3C	<	סוגר משולש שמאלי	
3E	>	סוגר משולש ימני	
2A	*	כוכבית	
2F	/	לוכסן (סלש)	
25	%	אחוז	
26	&	אמפרסנד	
7C		קו ניצב	

קבצי הטסט זמינים בקובץ zip ומומלץ תמיד להוריד ולהעביר אותם כ-zip על מנת למנוע שינוי אוטומטי של ירידות השורה על ידי תכנות להעברת קבצים.

# טיפול בשגיאות

**הערה:** אחרי הדפסת ההודעה המתאימה לשגיאה <u>הראשונה</u> בה נתקלתם, יש <u>לסיים את התכנית</u> (לשם כך היעזרו בפקודה (exit(0).

1. כאשר המנתח נתקל בתו לא חוקי יש להדפיס:

Error <char>\n

:כך שעבור הקלט הבא

@

:הודעת השגיאה תהיה

Error @

(רווח בודד בין Error ל @, וירידת שורה בסוף השורה על ידי LF בלבד.)

2. כאשר שורה מסתיימת באמצע מחרוזת, יש להדפיס:

Error unclosed string\n

3. כאשר הקלט מסתיים באמצע הערה (מהסוג הראשון המופיע בטבלת האסימונים), יש להדפיס:

Error unclosed comment\n

4. כאשר מחרוזת מכילה רצף מילוט שלא מופיע בהגדרת התרגיל, יש להדפיס:

Error undefined escape sequence <sequence>\n

כאשר <sequence> הינו התו הראשון ברצף המופיע לאחר הלוכסן האחורי. לדוגמא:

עבור מחרוזת המכילה את הרצף הלא חוקי: , \q יודעת השגיאה תהיה:

Error undefined escape sequence q

עבור מחרוזת המכילה את הרצף הלא חוקי:  $u\{000\}$ , הודעת השגיאה תהיה:

Error undefined escape sequence u

כאשר בתוך הערה נקרא הרצף \*/, יש להדפיס:

Warning nested comment\n

שימו לב: אין להדפיס את האסימון של COMMENT במקרה זה, אלא יש כאמור לסיים את התכנית מיידית. הערה: מטרת בדיקה זו היא להימנע מטעות של מתכנתים רבים: אין לקנן הערות, ברוב שפות התכנות. הנה דוגמה בעייתית בשפת C:

```
int nest = /* /*/ 0 */* */ 1; // our guess: int nest = 1;
int nest = /* /*/ 0 */* */ 1; // actual result: int nest = 0 * 1;
```

#### הוראות הגשה

עליכם להגיש קובץ zip המכיל קובץ אחד בלבד בשם hw1.lex דרך אתר הקורס.

#### דרישות נוספות והערות

על המנתח להיבנות <u>על השרת csComp</u> בעזרת הפקודות הבאות:

flex hw1.lex gcc -II lex.yy.c

מנתח שלא יבנה בהצלחה בעזרת הפקודות הללו יקבל 0 אוטומטית.

בתרגיל זה (כמו בתרגילים אחרים בקורס) ייבדקו העתקות. אנא כתבו את הקוד שלכם בעצמכם.

שימו לב: מומלץ (מאוד) להיוועץ ב-manual של flex לצורך ביצוע התרגיל. קל יותר לבצע אותו על ידי שימוש ביכולות .debug mode- מתקדמים, ו-regex patterns ,start conditions שלא נלמדו בתרגולים, כגון

**טיפ:** תוכלו להשתמש באתר <u>https://regexr.com</u> שעוזר בהבנה ובבנייה של תבניות regex מורכבות.

#### בדיקת המנתח

t2.in, t2.out ו-t1.in, t1.out באתר הקורס מופיע קובץ zip המכיל קבצי בדיקה לדוגמה

ניתן ואף רצוי לבדוק את עצמכם באופן הבא:

בנו את המנתח על ידי הפקודות לעיל על השרת csComp. העבירו את קובץ ה-zip של הקבצים לדוגמא לשרת ובצעו unzip. נניח שקובץ ההרצה של המנתח הוא a.out, אזי יש להריץ:

./a.out < t1.in >& t1.res diff t1.res t1.out

ולבדוק שמתקבל diff ריק.

<u>שימו לב</u> - במידה והמנתח שלכם לא עובר את כל קבצי הבדיקה שסופקו מראש לא תתאפשר הגשה חוזרת של התרגיל. בכדי למנוע מצבים כאלה - באתר מופיע script לבדיקה עצמית לפני הגשה בשם selfcheck-hw1 בו אתם נדרשים להשתמש בו על מנת לוודא את תקינות ההגשה שלכם. את הסקריפט ניתן להריץ על השרת csComp <u>בלבד</u> באמצעות הפקודה:

./selfcheck-hw1 hw1.zip

Compiler: Error at line 40

Me: "What? How? My code only has 30 lines

Compiler:



בהצלחה!

