

תורת הקומפילציה

תרגיל בית 1 – שימוש ב Flex

מתרגל אחראי לתרגיל: עידן שורץ 236360.cs@gmail.com
הגשה בזוגות.

הנחיות כלליות

- בתרגיל זה נממש מנתח לקסיקלי פשוט לשפת XML.
 - התרגיל יבדק בדיקה אוטומטית. **הקפידו למלא אחר ההוראות במדויק.** הבדיקה תתבצע על שרת תואם t2.
- יש להשתמש ב Flex בלבד (ולא ב Lex)

הגדרת האסימונים

שם האסימון	תיאור	דוגמאות
SS_TAG	התחלה של תג פותח. תג פותח מתחיל בסימן < ובצמוד אליו שם התג. שם התג מתחיל באות ויכול להכיל רצף כלשהו של אותיות וספרות.	<tag <tag2
SE_TAG	סוף של תג פותח, הסימן >	>
ATTR	מאפיין. מתחיל באות ואחריה רצף כלשהו של אותיות וספרות, בסוף הרצף יש להופיע סימן = (אין רווחים מסביב לסימן שווה). לאחר הסימן שווה מופיע בין גרשיים ביטוי שיכול להכיל כל תו מלבד ' \ ' ו ' ' ' (ייתכן גם ריק).	size="3" color="red"
E_TAG	תג סוגר. תג סוגר מתחיל ברצף הסימנים </ ובצמוד אליהם שם התג ובסוף הסימן > ללא רווחים. שם התג מתחיל באות ויכול להכיל רצף כלשהו של אותיות וספרות.	</tag>
SINGLE_TAG	תג יחיד. תג יחיד מתחיל בסימן < ובצמוד אליו שם התג. שם התג מתחיל באות ויכול להכיל רצף כלשהו של אותיות וספרות. בסיום מופיע רווח שלאחריו רצף הסימנים />	<tag />
STRING	מחרוזת. מחרוזת מתחילה באות, ולאחריה רצף כלשהו של אותיות.	string
PUNCTUATION	סימן פיסוק. תו בודד מהבאים: פסיק, רווח, נקודה, או קו מפריד.	, . -
DATE	תאריך. יורכב מרצף של ארבע ספרות המתחיל ב 19 או 20, לאחריו סימן – לאחריו רצף של שתי ספרות מ 01 עד 12,	2000-12-16 1920-01-30

	ולאחריו סימן – לאחר מכן עוד רצף של שתי ספרות מ01 עד 30.	
123 1	מספר שלם. הספרה 0 או מספר שמתחיל בספרה שונה מ 0 ואחריו רצף של אפס או יותר ספרות.	INT
.12 .0 1.2 01.01 1e-5 1e-09 .1e4 120e-3 0.0e2 0.00e2 0.000e2	מספר עשרוני. מספר עשרוני מכיל לפחות את אחד הרכיבים הבאים: 1. שבר – רצף ספרות, נקודה עשרונית, ואחריה רצף של ספרה אחת או יותר. 2. אקספוננט – האות e, אחריה הסימן – או + או אף אחד מהם, ואחר כך רצף של ספרה אחת או יותר. אם רק הרכיב הראשון קיים אז לא חייבת להיות ספרה לפניו. כאשר רק הרכיב השני קיים אז לפניו חייבת להיות ספרה אחת שונה מ 0 לפחות. (כדי להקטין את כמות הייצוגים העודפים לאפס) כאשר שני הרכיבים קיימים, רכיב 1 יבוא לפני רכיב 2. מספר עשרוני יכול להכיל אפסים מובילים או עוקבים (כלומר אחרי הנקודה).	FLOAT
<!-- comment -->	הערה. מתחילה ברצף התווים - - ! < ומסתיימת ברצף הראשון של התווים - - > (שני הרצפים לא מכילים רווחים). כל התווים באמצע הם חלק מההערה (כולל רווחים טאבים וירידות שורה)	COMMENT

פעולת המנתח והפלט הנדרש

כאשר המנתח מזהה אסימון יש לבצע את הפעולות הבאות:

- לפלוט שורה בפורמט הבא: (יש לדאוג לרווח יחיד בין כל רכיב בשורה, ולירידת שורה לאחר הרכיב האחרון)

```
<value> <token> <line_number>
```

כאשר:
 - line_number: מספר השורה שבה זוהה האסימון, כלומר השורה בו הוא מסתיים.
 - token: שם האסימון שזוהה (לפי השמות בחלק "הגדרת אסימונים" למעלה).
 - value: ערך האסימון שזוהה, כלומר הלקסמה.

דוגמה

עבור הקלט:

```
<tag>3</tag>
```

פלט המנתח יהיה:

```
1 SS_TAG <tag
```

```
1 SE_TAG >
1 INT 3
1 E_TAG </tag>
```

טיפול בשגיאות

כאשר המנתח נתקל בתו לא חוקי יש להדפיס את הודעת השגיאה הבאה **ולסיים** את ריצת התוכנית(העזרו בפקודת `(exit(0)`):

```
Error <char>\n
```

למשל עבור הקלט הבא:

```
$
```

הודעת השגיאה תהיה:

```
Error $
```

(רווח בודד בין Error ל \$)

הוראות הגשה

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

דרישות נוספות והערות

על המנתח להבנות על השרת `t2.technion.ac.il` בעזרת הפקודות הבאות:

```
flex hw1.lex
gcc -ll lex.yy.c
```

מנתח שלא יבנה בהצלחה בעזרת הפקודות הללו יקבל **0** אוטומטית.

הערה: שימו לב שבתרגיל זה לא נאכף התנאי ש `SS_TAG` ו `E_TAG` הם אכן זהים. נסו לנמק לעצמכם מדוע אי אפשר לדאוג לכך בשלב זה. נפתור בעיה זה בהמשך הקורס בחלק הסמנטי.

בדיקת המנתח

באתר הקורס קבצי בדיקה לדוגמה. `t1.in`, `t1.out`.

ניתן לבצע את הבדיקה באופן הבא:

נניח שקובץ ההרצה של המנתח הוא `a.out`, אזי יש להריץ:

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

בהצלחה! ☺