

מושגים בשפות תכנות

תרגיל 4

להגשה עד 15/06/2015

בתרגיל זה נבנה Parser ו Interpreter ל λ -calculus בשפת OCaml.

הנחיה חדשה שימו לב:

תרגיל זה ניתן בשתי גרסאות. הגרסה המקורית שפורסמה תלויה בספריית Core שאינה זמינה למערכת הפעלה Windows. גרסה זו נמצאת ב:

http://www.cs.tau.ac.il/~msagiv/courses/pl15/pl15_ex4.zip

והיא הגרסה המומלצת למי שעובד במערכת הפעלה Linux / Mac OS, לאחר התקנת Core. גרסה נוספת, שאינה תלויה בספריית Core, נמצאת ב:

http://www.cs.tau.ac.il/~msagiv/courses/pl15/pl15_ex4_nocore.zip

וניתן להשתמש גם בה לפתרון התרגיל למי שעובד בסביבת Windows.

ההבדלים בין הגרסאות הם קטנים וטכניים:

- הגרסה שלא תלויה ב Core מכילה ב utils.ml מספר פונקציות שימושיות שאינן חלק מהספרייה הסטנדרטית של OCaml.
- בגרסה שמתמשת ב Core המודול עבור קבוצות של מחרוזות בפתרון צריך להיות String.Set, ובגרסה השניה הוא צריך להיות StringSet שמוגדר ב utils.ml (ראו הערה בשאלות 2, 3).
- בגרסה שאינה משתמשת ב Core הטיפוס של פונקציות ספרייה (List.map) מעט שונה - הארגומנטים הם ללא תווית.
- יש הבדל בפקודת הקומפילציה, כפי שמוסבר בקובץ build.txt בכל אחת מהגרסאות.

כל קבוצה רשאית לפתור את התרגיל באמצעות אחת משתי הגרסות. בעת ההגשה עליכם להבהיר בגוף המייל באיזה גרסה השתמשתם (core / nocore) - והקוד ייבדק בהתאם.

הנחיות כלליות:

- בתרגיל נעשה שימוש ב OCaml בגרסת 4.01. כל הפתרונות צריכים לרוץ עם גרסה זו ולהתקמפל ללא שגיאות עם הפקודות שמפורטות ב build.txt. הוראות התקנה מופיעות ב: <https://github.com/realworldocaml/book/wiki/Installation-Instructions>
- בתרגיל זה כל הקוד צריך להיות "טהור", כלומר ללא side-effects, למעט הדפסות פלט.
- סגנון התכנות צריך להיות בהתאם למה שנלמד בשיעורים ובתרגול: שימוש נרחב ב pattern-matching ופונקציות רקורסיביות (במקום ביטויי תנאי ולולאות).
- כל השינויים בקבצים צריכים להיות במקומות המסומנים בהם. אין לשנות בקבצים דבר מלבד במקומות אלה.
- באחריותך לבדוק את הקוד שכתבת על דוגמאות נוספות ולוודא את נכונותו.
- מומלץ לקרוא את התרגיל עד סופו לפני שמתחילים לפתור אותו.

1. בשאלה זו נבנה Parser ל λ -calculus מורחב שכולל גם let expressions. התחביר הקונקרטי (concrete syntax) מוגדר ע"י הדקדוק הבא (אנו משתמשים בסמל \ במקום λ):

$$t ::= id \mid (\lambda id. t) \mid (t1 \ t2) \mid (t) \mid \text{let } id=t1 \text{ in } t2$$

שימו לב שהדקדוק מחייב סוגריים מסביב ל λ -abstractions ו applications. לדוגמה, המחרוזת הבאה היא מילה חוקית בשפה:

```
let tru = (\t. (\f. t)) in
let fls = (\t. (\f. f)) in
let and = (\b. (\c. ((b c) fls))) in
((and tru) fls)
```

הייצוג הפנימי לביטויים בשפה הוא AST, שניתן ע"י התחביר האבסטרקטי (abstract syntax) הבא:

$$\text{term} ::= id \mid \lambda id.\text{term} \mid \text{term1 term2}$$

הקובץ lexer.ml מכיל את ה Lexer המלא עבור שפה זו (אין לשנות קובץ זה), ומגדיר את הטיפוס token. הקובץ parser.ml מגדיר את הטיפוס הבא, שמשמש לייצוג ה AST (אין לשנות טיפוס זה):

```
type term = Variable of string
          | Abstraction of string * term
          | Application of term * term
```

בקובץ parser.ml עליכם לממש את הפונקציות הבאות:

```
parse_term : token list -> term * token list
parse : string -> term
format_term : term -> string
```

- הפונקציה parse_term מקבלת רשימה של tokens ומחזירה term ורשימה של של tokens שנשארו, בדומה לדוגמה שראינו בתרגול עבור ביטויים רגולריים. הפונקציה צריכה לזרוק SyntaxError במידה וה parsing נכשל. הפונקציה מטפלת בביטויים מהצורה let x=t1 in t2 ע"י ייצוגם בתחביר האבסטרקטי כך:

$$(\lambda x. t2) t1$$

(השתכנעו שזהו אכן ייצוג שמשמר את המשמעות של let expressions כפי שאנו מכירים אותם).

- הפונקציה parse מקבלת מחרוזת ומחזירה את ה term שהיא מייצגת, או זורקת SyntaxError במידה והמחרוזת אינה מכילה מילה בשפה (לפי הדקדוק של התחביר הקונקרטי).

- הפונקציה `format_term` מקבלת `term` ומחזירה ייצוג שלו באמצעות מחרוזת (נניח לצורך הדפסה). הייצוג צריך להיות מילה חוקית בשפה - כך שהפעלה של `parse` על התוצאה של `format_term` תחזיר `term` זהה ל `term` המקורי.

בקובץ `reducer.ml` נבנה `interpreter` עבור λ -calculus בשלבים, בשאלות הבאות.

2. הוסיפי לקובץ `reducer.ml` את הפונקציה:

```
fv : term -> String.Set.t
```

שמקבלת `term` ומחזירה את קבוצת המשתנים החופשיים בו. את הקבוצה יש לייצג באמצעות המודול `String.Set`. כזכור, את קבוצת המשתנים החופשיים ניתן להגדיר באופן אינדוקטיבי כך:

$$FV(x) = \{x\}$$

$$FV(\lambda x. t) = FV(t) \setminus \{x\}$$

$$FV(t1 t2) = FV(t1) \cup FV(t2)$$

3. לצורך מימוש `alpha-renaming`, אנו זקוקים לפונקציה שתחזיר שם של משתנה חדש שאינו בקבוצה של משתנים בשימוש. לצורך כך הקובץ `reducer.ml` מכיל את הערך `possible_variables : string list`, שמכיל רשימה של שמות משתנים אפשריים. הוסיפי לקובץ `reducer.ml` את הפונקציה:

```
fresh_var : String.Set.t -> string
```

הפונקציה צריכה לקבל קבוצה של שמות משתנים בשימוש, ולהחזיר שם חדש מתוך הרשימה `possible_variables`. במידה וכל השמות ברשימה בשימוש, הפונקציה צריכה לזרוק `OutOfVariablesError`.

הערה לשאלות 2, 3:

הטיפוסים בגרסת `nocore`, צריכים להיות:

```
fv : term -> StringSet.t
```

```
fresh_var : StringSet.t -> string
```

4. הוסיפי לקובץ reducer.ml את הפונקציה:

```
substitute : string -> term -> term -> term
```

על פונקציה זו לממש החלפה (substitution) כולל alpha-renaming במקרה הצורך. סדר הפרמטרים הוא כזה שהביטוי $t_1 t_2$ "x" substitute יחזיר את:

$$[x \mapsto t_1] t_2$$

כלומר t_2 כאשר כל המופעים של המשתנה x הוחלפו ב t_1 (ולא להיפך!).

הפונקציה צריכה לבצע את ההחלפה בכל מקרה, תוך שהיא מבצעת alpha-renaming במקרה הצורך. את פעולת הפונקציה ניתן להגדיר אינדוקטיבית באופן הבא:

$$[x \mapsto s] x = s$$
$$[x \mapsto s] y = y \quad \text{if } y \neq x$$
$$[x \mapsto s] (t_1 t_2) = ([x \mapsto s] t_1) ([x \mapsto s] t_2)$$
$$[x \mapsto s] (\lambda x. t) = \lambda x. t$$
$$[x \mapsto s] (\lambda y. t) = \lambda y. ([x \mapsto s] t) \quad \text{if } y \neq x \text{ and } y \notin FV(s)$$
$$[x \mapsto s] (\lambda y. t) = \lambda z. ([x \mapsto s] ([y \mapsto z] t)) \quad \text{if } y \neq x \text{ and } y \in FV(s), \\ \text{when } z \notin FV(s) \cup FV(t) \cup \{x\}$$

כאשר המקרה האחרון מבצע alpha-renaming (השתכנעי שזוהי אכן הגדרה נכונה להחלפה).

5. הוסיפי לקובץ reducer.ml את הפונקציה:

```
reduce_strict : term -> term option
```

על פונקציה זו לממש צעד אחד של חישוב (reduction) לפי סמנטיקת call-by-value. הפונקציה מחזירה ערך מטיפוס term option, כיוון שלא על כל term ניתן לבצע reduction. משמעות ערך החזרה היא:

$$(\text{reduce_strict } t) = \text{Some } t' \quad \text{if } t \Rightarrow t' \text{ in call-by-value}$$
$$(\text{reduce_strict } t) = \text{None} \quad \text{if } t \not\Rightarrow, \text{ i.e. } t \text{ is not reducible in call-by-value}$$

הפונקציה צריכה לממש את הכללים שנלמדו בשיעור ובתרגול, כאשר הערכים היחידים הם abstractions.

6. הוסיפי לקובץ reducer.ml את הפונקציה:

`reduce_lazy : term -> term option`

על פונקציה זו לממש צעד אחד של חישוב (reduction) לפי סמנטיקת lazy-evaluation. הפונקציה מחזירה ערך מטיפוס term option, ומשמעות ערך החזרה לפי ההסבר בשאלה 5, הפעם עבור סמנטיקת lazy-evaluation. סמנטיקת lazy-evaluation מוגדרת ע"י הכללים הבאים:

$$\frac{t_1 \rightarrow t'_1}{t_1 t_2 \rightarrow t'_1 t_2} \quad (\text{E-APP1})$$

$$(\lambda x. t_{12}) t_2 \rightarrow [x \mapsto t_2] t_{12} \quad (\text{E-APPABS})$$

7. הוסיפי לקובץ reducer.ml את הפונקציה:

`reduce_normal : term -> term option`

על פונקציה זו לממש צעד אחד של חישוב (reduction) לפי סמנטיקת normal-order. הפונקציה מחזירה ערך מטיפוס term option, ומשמעות ערך החזרה לפי ההסבר בשאלה 5, הפעם עבור סמנטיקת normal-order. סמנטיקת normal-order מוגדרת ע"י הכללים הבאים לפי הסדר, כלומר הכלל העליון ביותר מופעל במקרה בו ניתן להפעיל יותר מכלל אחד (בניסוח אחר, תמיד בוחרים את ה-redex החיצוני ביותר והשמאלי ביותר):

$$(\lambda x. t_{12}) t_2 \rightarrow [x \mapsto t_2] t_{12} \quad (\text{E-APPABS})$$

$$\frac{t_1 \rightarrow t'_1}{\lambda x. t_1 \rightarrow \lambda x. t'_1} \quad (\text{E-ABS})$$

$$\frac{t_1 \rightarrow t'_1}{t_1 t_2 \rightarrow t'_1 t_2} \quad (\text{E-APP1})$$

$$\frac{t_2 \rightarrow t'_2}{t_1 t_2 \rightarrow t_1 t'_2} \quad (\text{E-APP2})$$

8. הקובץ tests.ml מכיל את הפונקציה הבאה:

```
evaluate : verbose:bool -> (term -> term option) -> term -> term
```

פונקציה זו מקבלת את אחת הפונקציות שמימשת בשאלות 5-7, ומקבלת term, ומחשבת אותו איטרטיבית עד לצורה שהיא irreducible תוך שימוש בפונקציה הנתונה. אם הפרמטר verbose הוא true, הפונקציה גם מדפיסה את תהליך החישוב. הקובץ tests.ml גם מכיל שני קלטי בדיקה ראשוניים. הרחיבי את הקובץ כדי שיכלול בדיקות נוספות, והשתמשי בו במהלך הפיתוח של כל השאלות הקודמות כדי לבדוק את המימוש.

9. **בונוס א'**

הוסיפי לקובץ parser.ml את הפונקציות הבאות:

```
parse_term_conv : token list -> term * token list  
parse_conv : string -> term
```

הפונקציות צריכות לתפקד כמו הפונקציות משאלה 1, אבל לתמוך במוסכמות התחביריות (syntax conventions) שהגדרנו בכיתה, כלומר לאפשר להשמיט סוגריים, כך שלדוגמה הביטוי הבא יעבור parsing בהצלחה:

```
\x. \y. \z. x y z          (* parses like: (\x (\y. (\z. ((x y) z)))) *)
```

10. **בונוס ב'**

הוסיפי לקובץ parser.ml את הפונקציה:

```
format_term_conv : term -> string
```

שתפעל בדומה ל format_term משאלה 1, אבל לא תדפיס סוגריים מיותרים, כשהיא לוקחת בחשבון את המוסכמות התחביריות שהגדרנו בכיתה.

בהצלחה!