Python Slicer (Projector)

Oded Elbaz, Tomer Greenwald

Program Analysis 2016

Lecturer: Prof. Mooly Sagiv

תוכן

[מבוא 2](#_Toc441481796)

[איך להריץ? 3](#_Toc441481797)

# מבוא

בחרנו לממש Python Slicer עפ"י הצגת הפרויקט עבור Java בשיעור 12 ואישורך במייל לבצע את הפרויקט על קוד פיית'ון. קראנו לתוכנית Projector כיוון שהיא מקרינה את קוד הקלט על פני משתנה שבחר המשתמש.

הProjector מבצע אנליזה על הקוד ויוצר שני גרפים בעלי רשימת קודקודים משותפת.

כל קודקוד ברשימה מתאים לשורת קוד בתוכנית המקורית. ישנם שני סוגי קודקודים:

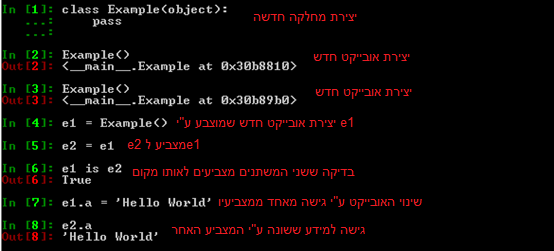
* Statement Node: מתאים לשורת השמה או ביטוי.
* Control Node: מתאים לשורת קוד שמשפיעה על הflow בתוכנית. לדוגמא if, else, while.

באשר לקשתות, סט אחד מתאר את הControl Flow בתוכנית, והשני מתאר את התלות הלוגית (Dependency) בין שתי שורות קוד.

הכוח הגדול של האנליזה טמון בעובדה שהיא מזהה שינויים באובייקט עליו אנחנו מצביעים, גם כאשר השינויים בוצעו ממצביע אחר.

תזכורת על שפת פיית'ון:

כל קריאה לConstructor יוצרת אובייקט חדש. שני משתנים יכולים להצביע אל אותו אובייקט. שינוי באובייקט משפיע על האובייקט ולא על המצביע, לכן נוכל לראות את השינוי ע"י גישה מכל אחד מהמשתנים שמצביעים אליו:

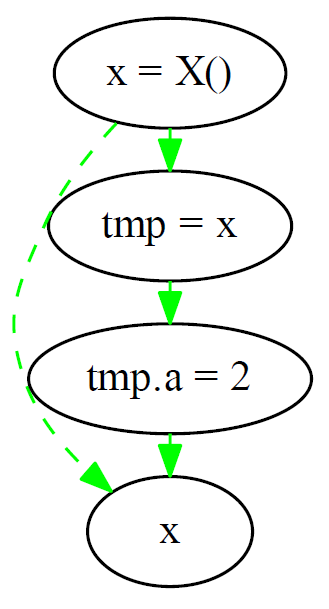


האנליזה יודעת לזהות ששני משתנים מצביעים לאותו אובייקט, וכן שורות קוד שגורמות לשינוי באובייקט זה. אנחנו עושים זאת באמצעות Points to Analysis כפי שיוסבר בהמשך. דוגמא (דוגמאות מורכבות יותר יינתנו בהמשך):



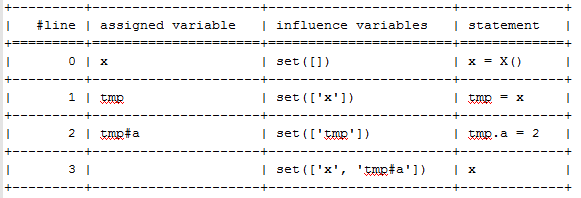
בדוגמא הזאת אנחנו מייצרים אובייקט חדש מסוג X שמצביעים אליו המשתנים x ו-tmp. אנחנו משנים את האובייקט דרך tmp, ורוצים לוודא שהמשתנה x מודע לשינויים הללו, כך שכאשר נקרין את הקוד על פני x, גם השורה tmp.a=2 תהיה בפלט (וכן השורה שמקשרת את tmp לאובייקט שעליו מצביע x).

זוהי תוצאת האנליזה:



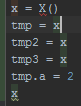
אנחנו רואים ש-x בשורה האחרונה מושפע על ידי ההשמה הראשונה למשתנה וע"י ההשמה tmp.a. tmp.a מושפע מההשמה למשתנה tmp.

"מאחורי הקלעים" האנליזה מייצרת את הטבלה הבאה:



ניתן לראות שהאנליזה יודעת שx בשורה 3 מושפע מx ו-tmp.a.

למרות שהשורה tmp=x משפיעה על x בשורה 3, tmp לא מופיע ברשימת המשתנים המשפיעים. בנינו את האנליזה כך כיוון שאנחנו לא רוצים להוסיף את כל המשתנים שמצביעים אל האובייקט, אלא רק אלו שמשפיעים על התוצאה. השורה tmp=x תתווסף לפלט מכיוון שtmp.a=2 תלוי בtmp. כך לדוגמא, אם נסתכל על הקוד הבא:



ההשמות של האוביקט אל tmp2 ו-tmp3 לא ישפיעו על x.

הנחת העבודה היא שהקוד בצורת SSA, כך שההשמה לכל משתנה היא יחידה. לכן, אם, לדוגמא, tmp.a נמצא ברשימת המשתנים המשפיעים, עלינו לחפש היכן בוצעה ההשמה היחידה לtmp.a.

כפי שניתן לראות כבר מהדוגמא הפשוטה הנ"ל, הקוד יודע לטפל בגישה לAttributes של משתנה. נשים לב שזוהי הדרך היחידה לשנות אובייקט. במילים אחרות, שינוי אובייקט הוא שינוי הAttributes שלו.

לאחר שבנינו את הגרפים, הקרנת הקוד היא כבר עניין טכני – אנחנו הולכים מהסוף להתחלה עד אשר אנחנו מזהים את הפעם הראשונה שבה המשתנה המוקרן מושפע. משם אנחנו מטיילים אחורה על פני קשתות הDependency. כל פעם שאנחנו מגיעים לקודקוד חדש, אנחנו בודקים האם יש Control Flow בין שני הקודקודים. במידה ויש, אנחנו מוסיפים את הקודקוד החדש לקוד המוקרן וכן את כל הControl Nodes בדרך.

לאחר שהאנליזה הסתיימה אנחנו יוצרים שלושה סוגי פלטים:

1. רשימה מלאה של תוצאת האנליזה (analysis\_result.txt)
   1. טבלה המתארת איזה משתנים משפיעים ומושפעים עבור כל שורת קוד.
   2. רשימת Control Edges
   3. רשימת Dependency Edges
2. המחשה גרפית של הגרפים שהאנליזה יצרה:
   1. Out.gv\_control.pdf: גרף הControl Flow בתוכנית
   2. Out.gv\_dep.pdf: גרף הDependency בתוכנית
   3. Out.gv\_all.pdf: איחוד של שני הגרפים הקודמים
3. הקוד המוקרן על פני אחד המשתנים (projected\_code.py)

# איך להריץ?

התוכנית נבדקה ועובדת על פני פלטפורמת Windows, ופיית'ון 2.7.

החבילות הבאות צריכות להיות מותקנות על מנת שיהיה אפשר להריץ את התוכנית:

* Astor – חבילה שבאמצעותה אנחנו הופכים אובייקטי AST לשורת קוד:

<https://pypi.python.org/pypi/astor/0.5>

* Graphviz – חבילה שבאמצעותה אנחנו מדפיסים בצורה גרפית את הגרפים שאנחנו יוצרים:

<https://pypi.python.org/pypi/graphviz/0.4.8>

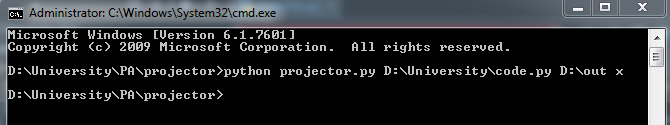
אם ברצונך להריץ את הטסטים, עליך להתקין pytest:

<https://pypi.python.org/pypi/pytest/2.8.5>

כאשר כל התלויות מוכנות, ניתן להריץ את האנליזה בצורה הבאה:

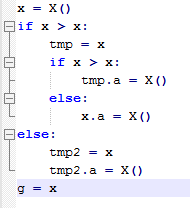
python projector.py <original\_code> <output\_dir> <projected variable>

דוגמאת הרצה:



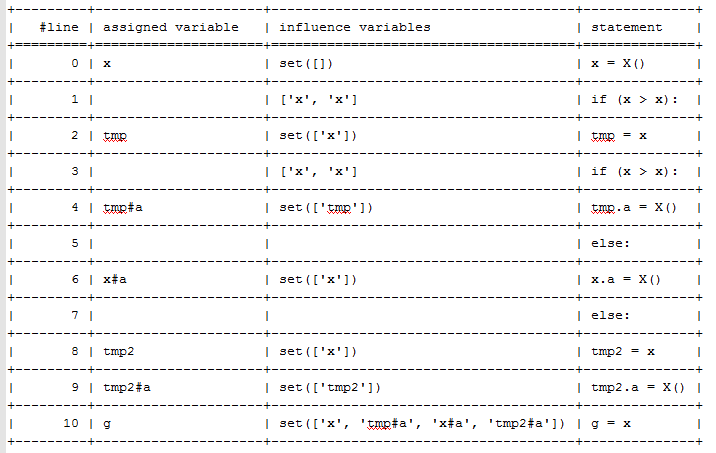
# כיצד לקרוא את הפלט

על מנת להדגים את קריאת הפלט נריץ את האנליזה על התוכנית הבאה:



ההטלה שבחרתי להדגים היא על פני משתנה g.

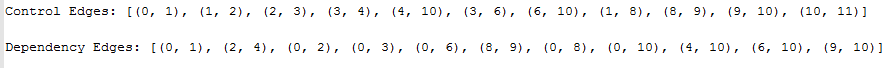
ראשית נתרכז בטבלה שבפלט analysis\_result.txt:



ניתן לראות שישנה שורה עבור כל שורת קוד.

כל שורה מציגה איזה משתנה מושפע כתוצאה מהפעלת השורה (assigned variable), ואיזה משתנים משפיעים על הרצת השורה (influence variable). כך לדוגמא אם נסתכל על שורה 2, המשתנה המושפע הוא tmp כי זוהי שורת השמה למשתנה זה, והמשתנה המשפיע עליו הוא x. עבור שורה 1, אף משתנה לא מושפע מהרצת שורה זאת (זוהי שורת if), אך המשתנים המשפיע על שורה זאת הוא x.

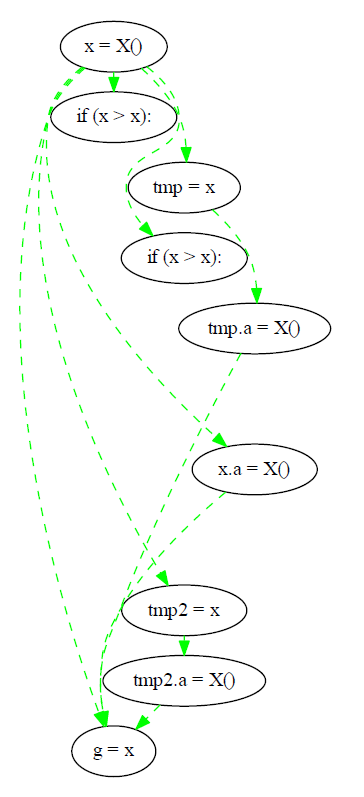
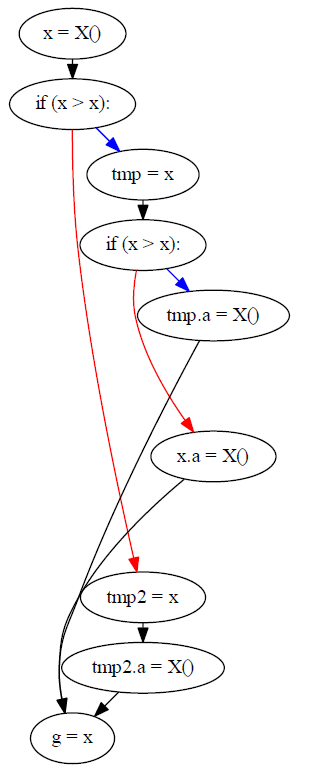
כעת נתרכז ברשימות הקשתות המופיעות באותו קובץ:



Control Edges מתארת את הControl Flow של הקוד. אם הקוד טורי אזי יש קשת בין שורת קוד לשורה הבאה אחריה, לדוגמא 0 ל-1. אם יש קפיצות (לדוגמא בif או while), אזי יש קשת עבור כל קפיצה מתאימה. לדוגמא 1 ל-2 ו1 ל-8.

Dependency Edges היא רשימה המחזיקה את התלויות הלוגיות. לדוגמא זיהינו בשורה 10 את התלות בx, tmp.a, x.a, tmp2.a, ולכן יש קשת אל שורה 10 מההשמות אל משתנים אלו (נזכיר שאנחנו מניחים שהקוד בצורת SSA).

קבצי הpdf שנוצרו בתיקיית הפלט מציגים את שני הגרפים בצורה גרפית:



ולבסוף ניתן לראות את הקוד המוקרן בקובץ projected\_code.py.

הערה:

כפי שניתן לראות, הטבלה שבקובץ analysis\_result הינה ליבת האנליזה, כל שאר הפעולות הן טכניות מעליה. לכן מעתה נציג במסמך רק את הטבלה הזאת.

כמו כן, הדוגמאות שנציג יהיו קצרות מאוד, כי אנחנו רוצים להתרכז במהות ולא לבלבל עם מספר רב של שורות קוד. ניתן להרחיב את הדוגמאות והקונספטים ליצירת קוד מורכב יותר.

# הנחות עבודה

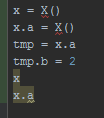
* קוד הקלט בצורת SSA.
* כל משתנה מחזיק אובייקט או Integer (ניתן להרחיב כך שהמשתנה יחזיק גם טיפוסים נוספים, אך זהו עניין טכני).
* האנליזה היא intraprocedural.
* הקוד אינו מכיל הגדרות מכל סוג שהוא, ההנחה היא שהכל מוגדר כפי שצריך (לדוגמא אם קוראים לConstructor של המחלקה X אזי המחלקה מוגדרת כראוי).
* האנליזה מטפלת בהשמה של ערך או פעולה בינארית (ניתן להרחיב, זהו הליך טכני).
* הקוד יודע לטפל במבנה של If-else, while וקינון ביניהם.
* אובייקטים
  + האובייקטים הם מהצורה:



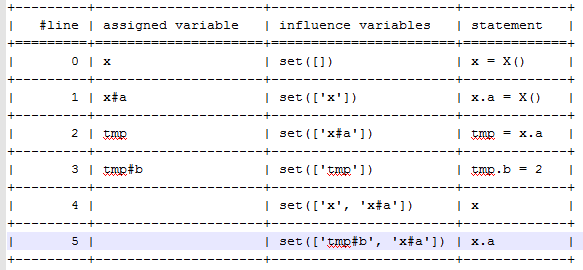
כלומר בברירת מחדל ללא Attributes, ניתן להוסיף אותם דינמית.

* + לא ניתן לגשת לAttribute של Attribute, לצורך כך יש להשתמש במשתנה עזר.
  + הפרויקט עובד על Shallow Pointers, לכן לדוגמא x.a.a משפיע על x.a ולא על x (בצורה דומה x.a משפיע על x).
  + הקריאה לConstructor חוקית.
  + דוגמא לסעיף אובייקטים:

נניח נתון הקוד הבא:



אזי האנליזה תייצר את הטבלה הבאה עבורו:



* הקוד לא יכול להיגמר במשתנה אותו רוצים לבדוק

הוספת המשתנה

# Points to analysis

# אלגוריתם האנליזה

# דוגמאות