Goals

The goals of my project “Syntatic Rule Useage in Python” is to recreate a study done by Dong Qiu, Bixin Lia, Earl T.Barr and ZhendongSu titled “Understanding the syntactic rule usage in Java”. My study will try to answer questions posed in Dong Qui et als work such as How are syntactic rules used in practice? How are syntactic rules used in practice over time? How strongly do rule usage in practice depend on context?

I hope to parse enough python files to generate a sample that accuratly mirrors real world code. From this sample I intend to count the various Syntax rule useages through out all the projects and ind the individual projects. To date I have 49 python applications to parse yeilding 29002 individual python files. These applications hail from the Qualitas corpus. However i would like to add a bout 50 more python applicions to the parse list. I would like these new 50 applications to have more of a home grown feel and not a comersial one. I plan on trawling git hub to collect these python application.

Background

i hadnty an overly extensive bckground in Python development besides a course on CodeAcademy. Although i was (and sill am) a relitive novice in python, i was excided to undertake a project based in Python. This way i get to grips with an extremely popular programming language whilst also gaining an insight to how other developers utalise the varying syntax in python

Progress

to achieve my goals i have to complete a numbe of steps

* Aquire files to parse

this is in my eyes half completed

* write python code that can trawl through a directory amd its subsiquent sub-directories and pick out the desired python files

This is complete, i have working code capeable of this

* Be able to parse python files and get at their Abstract Syntax Tees

can do

* be able to walk through the asts and pick out and identify the nodes

can do this however ill have to type up comparrison statements for each type of node eg. isinstance(node, ast.IfExp): identifies an If statement

Problems

So far I havnt encountered any major problems. One problem I did encounter is that if the Python file does not contain valid python syntax but is named with the suffix .py the entire process of trawling through the files and parseing them will be halted. Whilst look ing through the python documentation for pythons ast module i noticed some new classes (such as MatMult; the @ operator for matrix multiplication ) were introduced in python 3.5. When checking the trees for this class an error was thrown, AttributeError: 'module' object has no attribute 'MatMult'.

From what i can deduccefrom this error my IDE is running an earlier version of python(2.7).

This leads me to wonder what if i preformed the analysis utalising the newer python would results vary and if yes by how much and what would this indecate

Planned compleetion steps

To complete the progeat i will need to complie a results list from all the parsed python files then use a tool such as R to annalyse the data and construct graphs to back up arguements and answer questions posed by the study. I aslo Need to gain an understanding of the nodes so i can analyse the data professionally.