Practical 8. Working With Information

Xin Chen

Semester 2, 2018/19

8.1 Objectives

- Learn to program with the help of online references

8.2 Introduction

In this practical, you will write python codes for one problem. You may need to look for online references to complete this assignment, which is also a good practice on how to ask right questions. You are expected to use version control systems (e.g. git) to store your codes. After this practical, the tutorial session may give you an opportunity to share your codes to your peers and receive critical reviews for improvement of your codes. You may continue to make them better.

In class, your instructors are there to help. Please don't be shy about asking them. Once you have completed the assignment, you are free to leave.

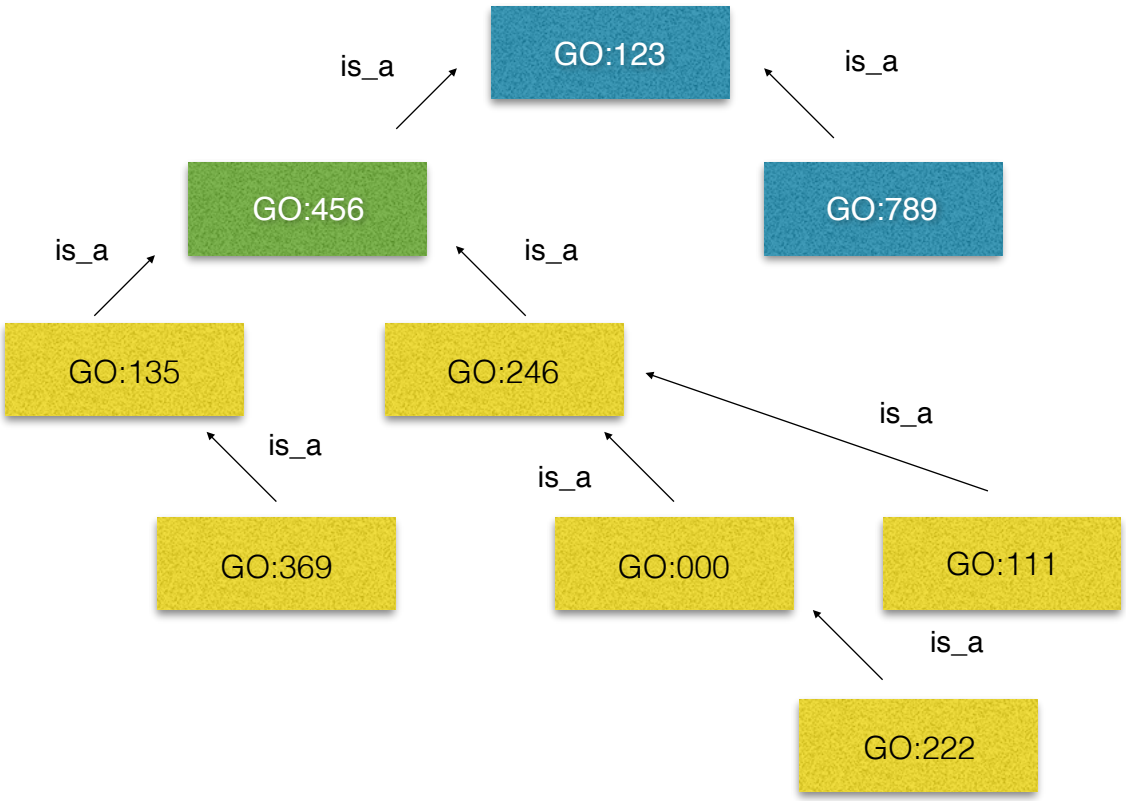
8.3 Problems to solve

Find GO terms!

Problem: You have interest in biological process related with '**autophagosome**'. Try to find genes that related with 'zinc'. In the xml document, if text in `<defstr>` contains the word 'autophagosome', we suppose this gene is related with 'autophagosome'. And you also have to find the number of **childNodes** of the target gene

(`<is_a>`tag means subclass, e.g. condensed chromosome 'is-a' chromosome)

Suppose GO:456 is related with 'autophagosome', the number of child nodes is 6 (`count until you reach the bottom`). (Figure 1)



(Figure 1: An Example tree of GO terms)

Given: an xml documents from Gene Ontology named '`go_obo.xml`'

Return:

an excel that contains: GO id, gene name, definition string, number of child nodes

Example:

`autophagosome.xlsx`

id	name	definition	childnodes
GO:0000045	autophagosome assembly	The formation of a dou	0
GO:0000421	autophagosome membrane	The lipid bilayer surr	1
GO:0016236	macroautophagy	The major inducible pa	10
GO:0016237	lysosomal microautopha	The transfer of cytoso	3
GO:0016240	autophagosome membrane	The initial attachment	0
GO:0016243	regulation of autophag	Any process that modul	2
GO:0030399	autophagosome membrane	The controlled breakdo	0
GO:0032258	protein localization t	A cytoplasm to vacuole	0
GO:0034423	autophagosome lumen	The volume enclosed wi	0
GO:0044753	amphisome	Intermediate organelle	0
GO:0044754	autolysosome	A type of secondary ly	0
GO:0045771	negative regulation of	Any process that reduc	0
GO:0045772	positive regulation of	Any process that incre	0
GO:0048102	autophagic cell death	A form of programmed c	1
GO:0061709	reticulophagy	The autophagic process	0
GO:0061739	protein lipidation inv	The protein lipidation	0

Tips:

- DOM
- pandas.DataFrame.to_excel