Coding Test (Applicants 2025)

In this task, you will process **Gene Ontology (GO)** annotation data. The aim is to read annotation files and compute how many genes are associated with each GO class, accounting for the class hierarchy. Some of the data has been preprocessed to make the task easier. The required files are available in the provided folder.

Background: What is Gene Ontology?

If you're unfamiliar with Gene Ontology (GO):
GO provides a structured vocabulary (a hierarchy) of gene functions used across many organisms. Each GO class (or "term") describes a biological process, molecular function, or cellular component. The structure is a **directed acyclic graph**: a GO class can have one or more parent classes, representing broader categories.

If a gene is annotated to a class X, it is **also considered annotated to all parent classes of X**.

You can find more information in the attached PDF slides (especially slides 18–24) or from GO website (https://geneontology.org/). You can also search GO tutorials from internet.

What You Should Do

Create a standalone Python script that takes the file names (for gene_association.mgi and mergeGO.out) as input arguments, from command line. When run, it should process the data and print the 50 largest GO classes.

Your solution will be evaluated based on **correctness** and **clarity of code**. Output formatting and code structure are flexible as long as the core functionality is correct.