

Course: Software Engineering

Project Title: Source Code Search Engine

Project Supervisor: Dr. David Lillis (david.lillis@ucd.ie)

Students Preallocated: Ciaran CONNOLLY (14445308); XINGYU Pan (14207156); XU Ke (14207177); ZHENG Mengya (14207169)

Information Retrieval, Source Code Analysis

Search engine technology has concentrated mostly on allowing users to search for text in natural language. Extensions of this approach have been applied to various types of multimedia retrieval (e.g. images, videos, etc.).

Some attempts have been made to extend search to source code. However, these have shown that this type of text offers different challenges to natural language. This project aims to build a search engine with source code as its focus.

The project will need to gather a large corpus of source code (e.g. from GitHub or similar sites), written in a variety of programming languages. It will be important to index this in an efficient manner so that it is easily searchable. A challenge that must be faced is the decision about how source code can be represented appropriately so as to facilitate different types of searches. For example, a user may wish to search for a particular function, or for an entire program. The search string supplied could be a description of code that the user wishes to see, or a user may wish to search for the origin of a particular piece of code.

A suitable user interface must also be developed. This may include other advanced features related to source code search. This may include features such as project visualisation and analysis, automatic language detection, etc.