Integrating Code Search into the Development Session

Mu-Woong Lee, Seung-won Hwang, and Sunghun Kim
Code Similarity Search / Clone Detection?

Usually, it is a "post-mortem" process

Code corpus
Why Not Online?

- Find clones ASAP!

- Query code segment

- Code corpus
  - Millions of code segments...

- Commercial search engines?
  - Source code ≠ Simple text

- Clones
Vector Abstraction of Code

- **Characteristic Vectors** [1]
  - Tree similarity $\rightarrow$ Vector similarity

Each sub-tree containing at least $T$ nodes

Indexing Structure [2]

- Based on the GEMINI framework
  - “Filtering-then-Ranking” approach

Application Scenario

- Integrating Code Search into the Development Session [3]
  - Given an internet scale software corpus, by posing every code segment as a query to this corpus at the time of generation, developers can instantly reference the relevant code segment to enhance productivity.

Demo

Development Session

```java
public void setDestinationBands(int[] destinationBands) {
    if (destinationBands == null) {
        this.destinationBands = null;
    } else {
        int numBands = destinationBands.length;
        for (int i = 0; i < numBands; i++) {
            int band = destinationBands[i];
            if (band < 0) {
                throw new IllegalArgumentException("Band value < 0!");
            }
        }
    }
}
```

Search Result

```java
public void setSourceBands(int[] sourceBands) {
    if (sourceBands == null) {
        this.sourceBands = null;
    } else {
        int numBands = sourceBands.length;
        for (int i = 0; i < numBands; i++) {
            int band = sourceBands[i];
            if (band < 0) {
                throw new IllegalArgumentException("Band value < 0!");
            }
        }
    }
}
```

Copy & Paste

```java
for (int j = i + 1; j < numBands; j++) {
    if (sourceBands[j] == sourceBands[i]) {
        throw new IllegalArgumentException("Duplicate band value!");
    }
}
this.sourceBands = (int[])(sourceBands.clone());
```
Thank You!