code2seq: Generating Sequences from Structured Representations of Code

Uri Alon, Shaked Brody, Omer Levy, Eran Yahav



Demo + code: http://www.code2seq.org

1. Task: Generate a Natural Language Sequence from a Given Snippet of Code

Code summarization in Java:

```
boolean ____(Set<String> set, String value) {
  for (String entry : set) {
      if (entry.equalsIgnoreCase(value)) {
           return true;
   return false;
```

contains ignore case

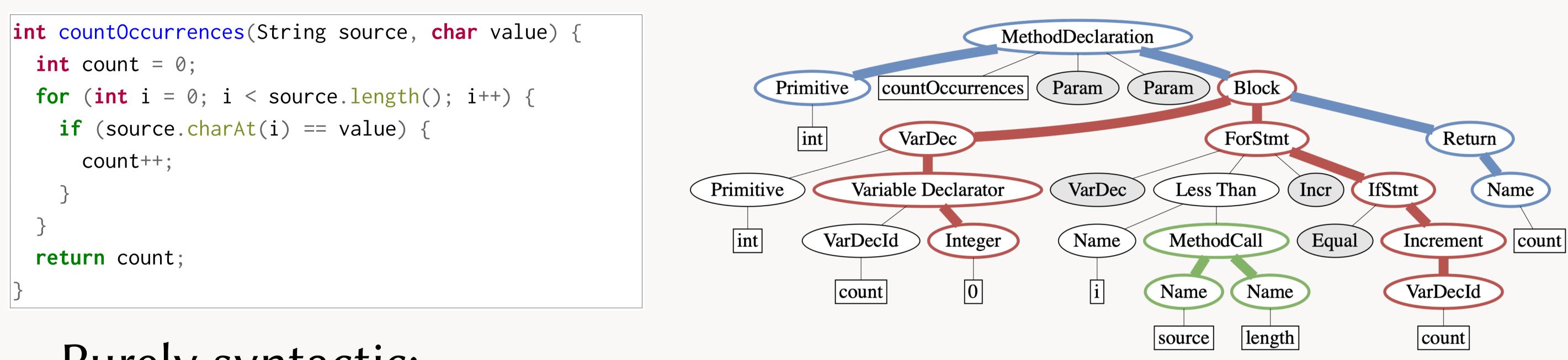
Code captioning in C#:

```
void Main()
string text = File.ReadAllText(@"T:\File1.txt");
 int num = 0;
text = (Regex.Replace(text, "map",
    delegate(Match m) {
  return "map" + num++;
 }));
 File.WriteAllText(@"T:\File1.txt", text);
```

replace a string in a text file

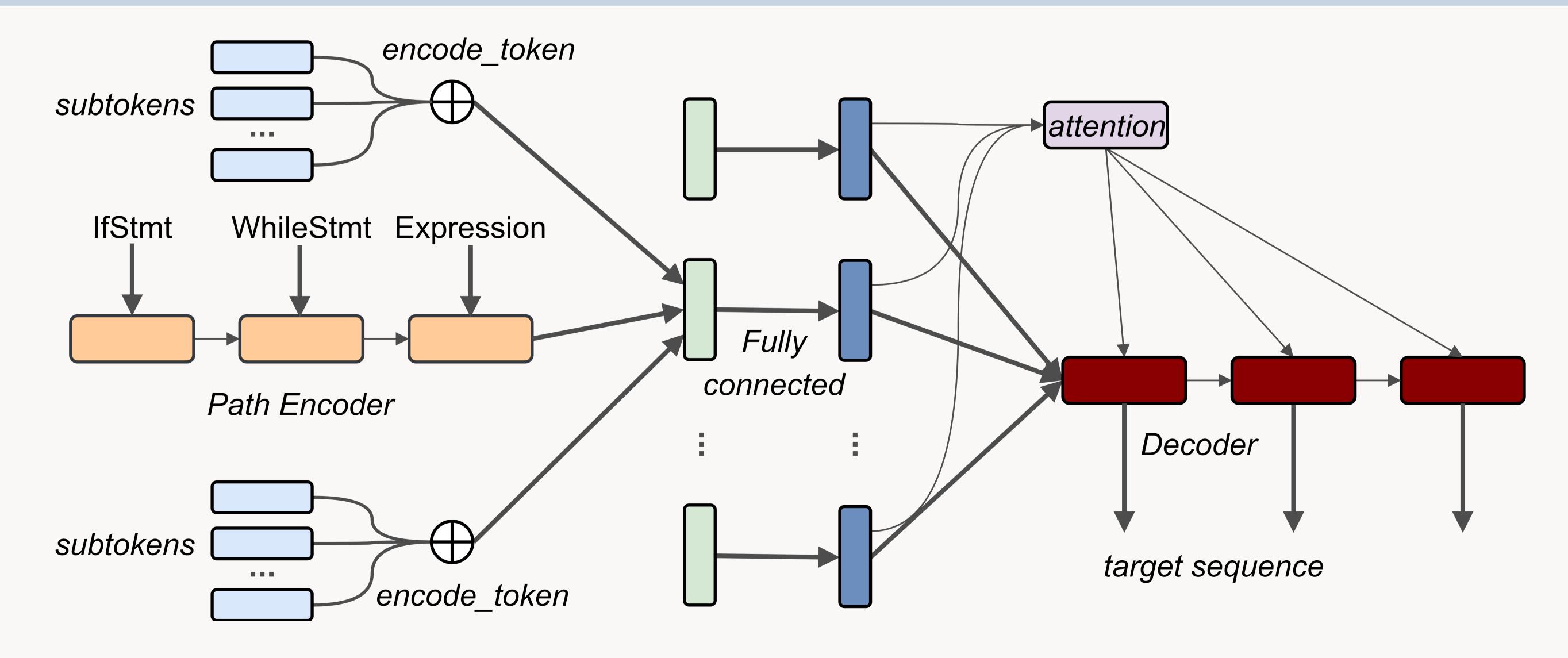
2. Abstract-Syntax-Tree (AST) Paths (Alon et al., PLDI'2018)

AST paths are a general approach for representing code in machine learning models.



- Purely syntactic:
- Do not require expert semantic knowledge about the language
- Do not require to hard code human knowledge into features
- The same approach works for any programming language

3. Key Idea: Represent Code as the Set of Paths in its AST



- A code snippet is represented as a set of its AST paths.
- An LSTM encodes each AST path as a sequence of nodes.
- While decoding, the model attends to the source paths instead of source tokens.

4. Main Results: More Accurate Than All Baselines, Across 5 Datasets

Tasks:

- Code Summarization generate a method name given a Java method body.
- Code Captioning generate an NL sentence describing a given code snippet in C#.
- Code Documentation generate a JavaDoc documentation given a Java method.

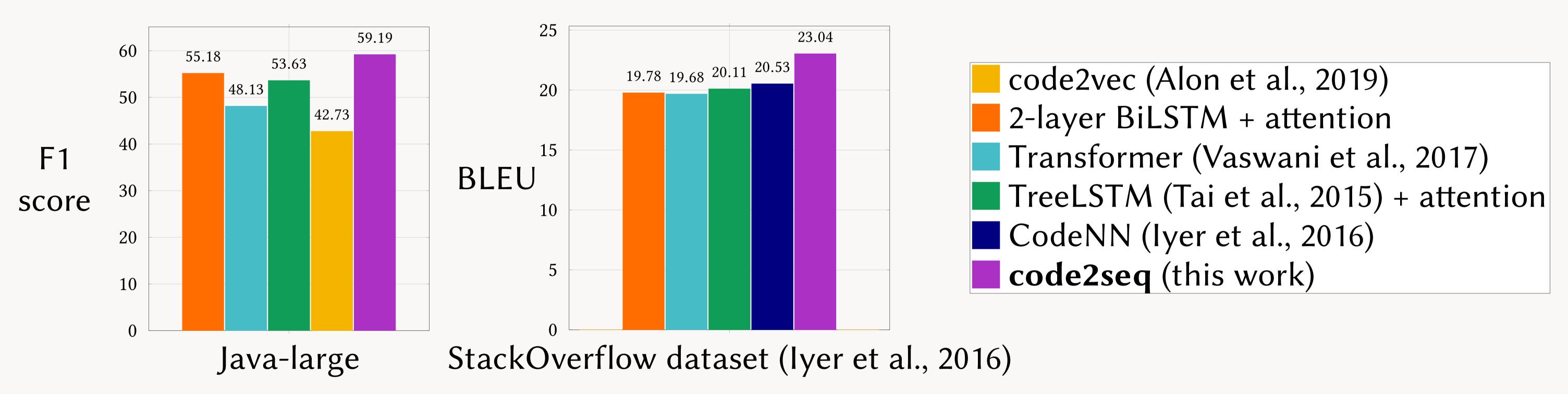
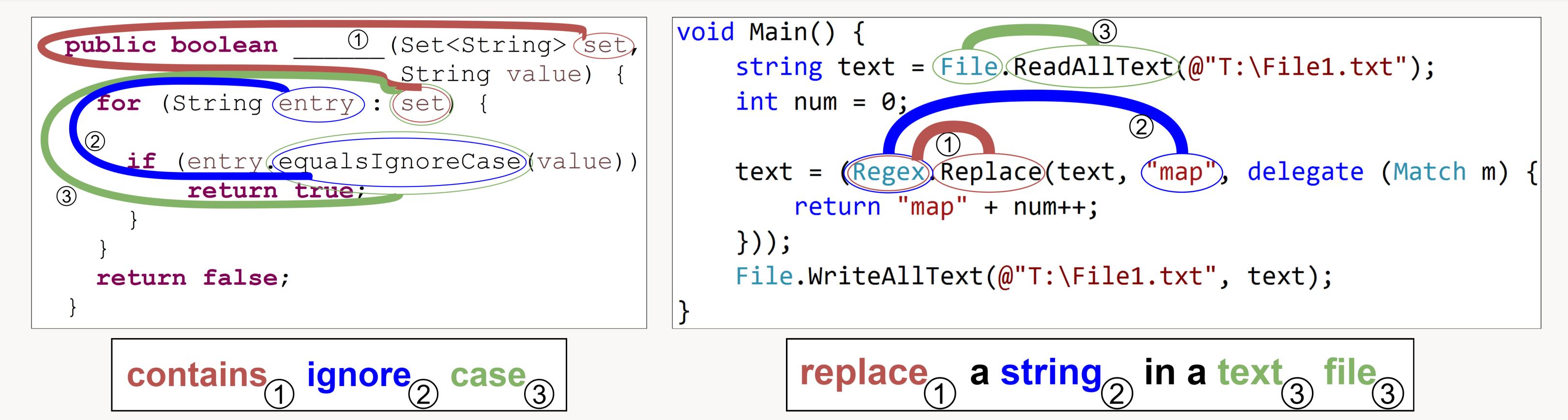


Figure: Code Summarization task (Java)

Figure: Code Captioning task (C#)

6. Examples



Model	Prediction	Model	Prediction
BiLSTMs	contains	BiLSTMs	how to read a text file from a text file
Transformer	contains	Transformer	how to write a . txt file in c #
TreeLSTM	contains ignore case	TreeLSTM	how to get the text from a text file in c #
code2vec	contains ignore case	CodeNN	how to read a text file in c #?
code2seq	contains ignore case	code2seq	replace a string in a text file