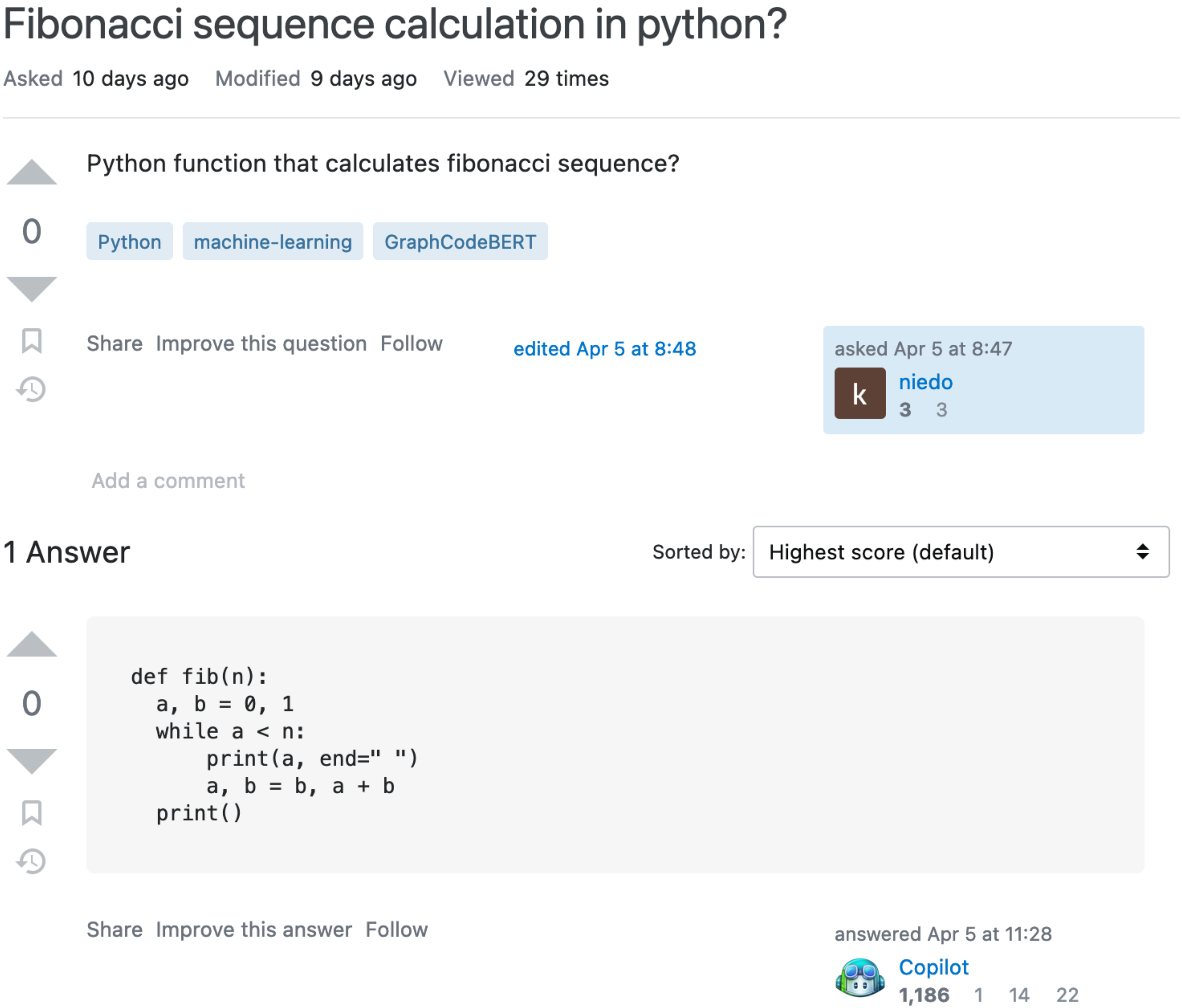


Fibonacci sequence calculation in python?

Asked 10 days ago Modified 9 days ago Viewed 29 times

0

Python function that calculates fibonacci sequence?

Python machine-learning GraphCodeBERT

Share Improve this question Follow

edited Apr 5 at 8:48

asked Apr 5 at 8:47

k niedo

3 3

Add a comment

1 Answer

Sorted by: Highest score (default)

0

```
def fib(n):
  a, b = 0, 1
  while a < n:
      print(a, end=" ")
      a, b = b, a + b
  print()
```

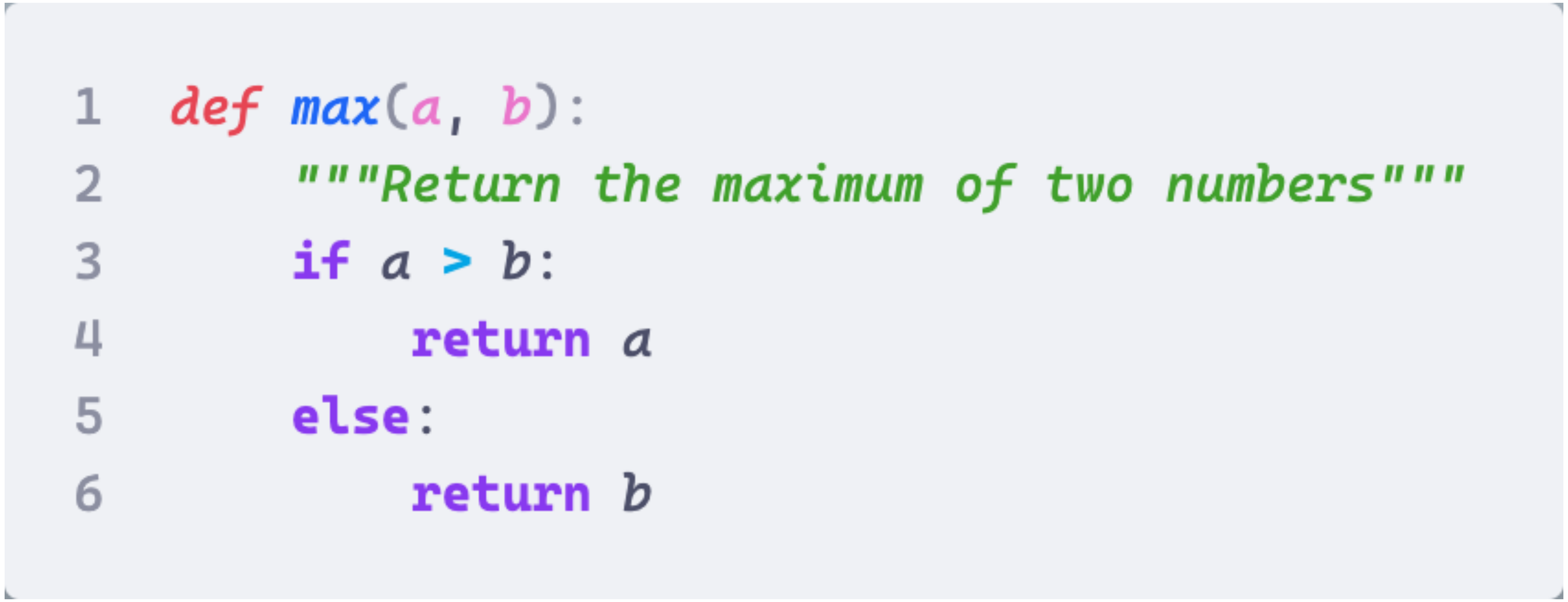
Share Improve this answer Follow

answered Apr 5 at 11:28

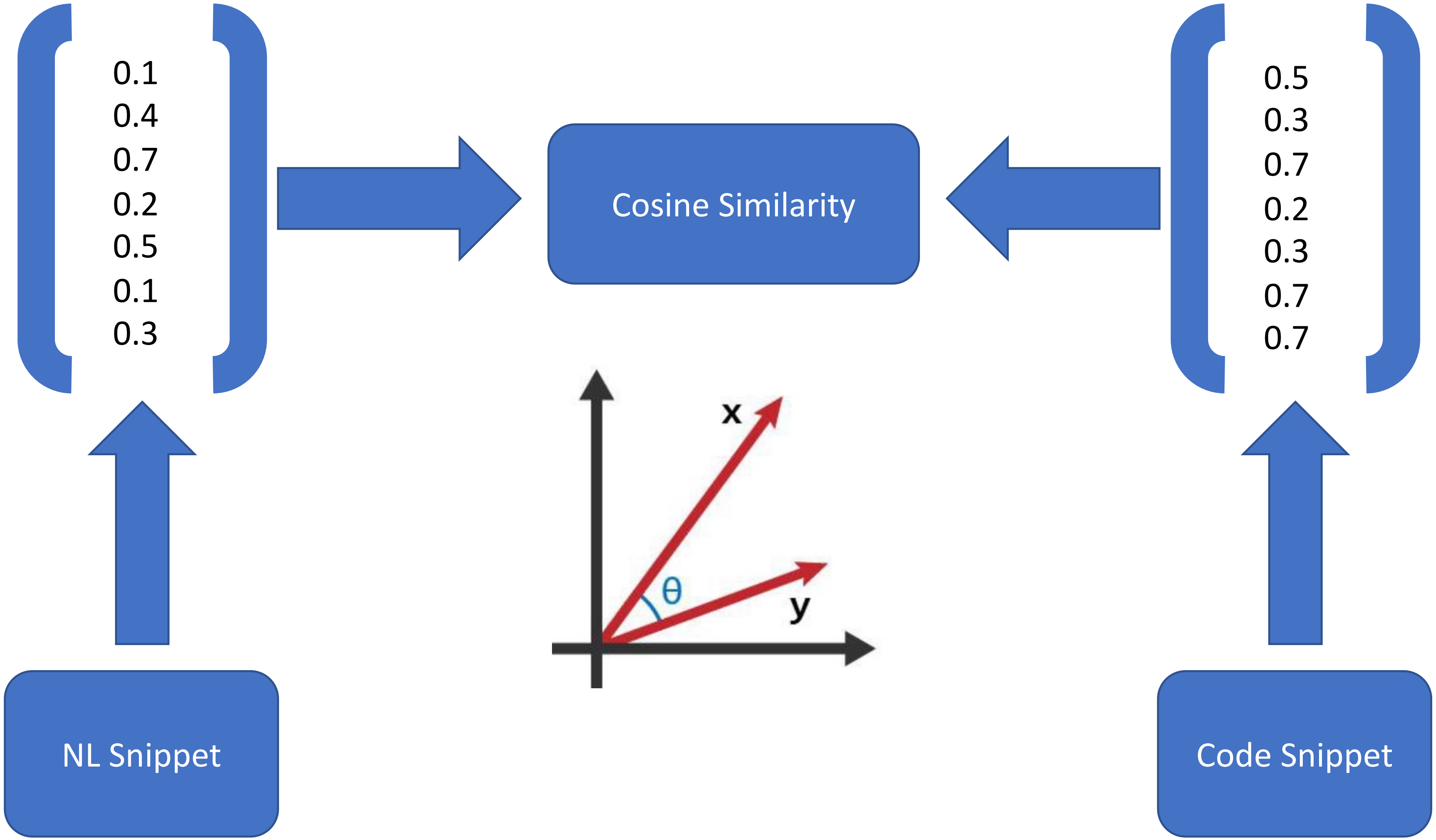
Copilot

1,186 1 14 22

Natural Language Code Search



```python
def max(a, b):
    """Return the maximum of two numbers"""
    if a > b:
        return a
    else:
        return b
```



0.1
0.4
0.7
0.2
0.5
0.1
0.3

Cosine Similarity

0.5
0.3
0.7
0.2
0.3
0.7
0.7

NL Snippet

Code Snippet

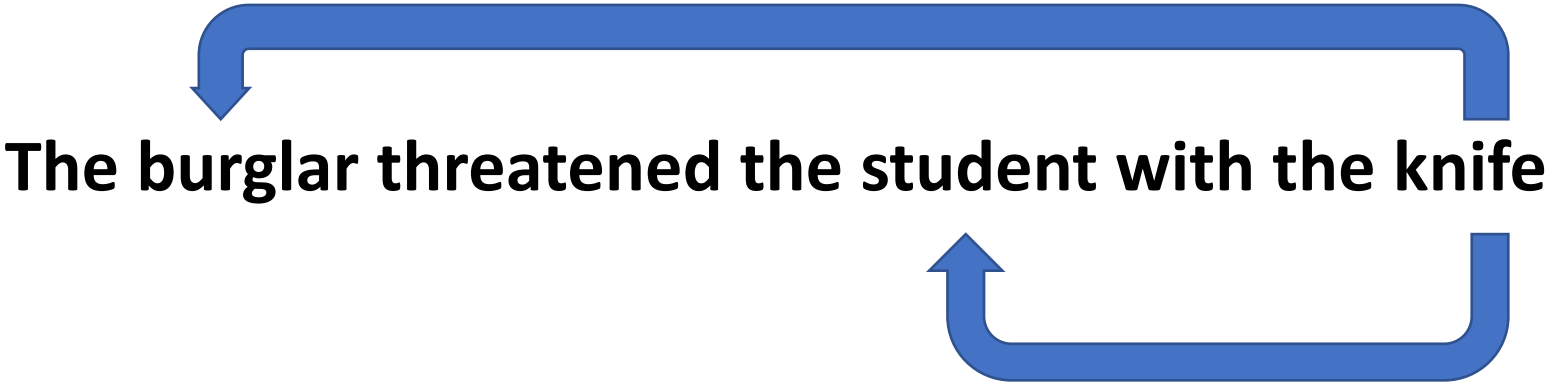
https://www.learndatasci.com/glossary/cosine-similarity/

Natural Language vs. Code

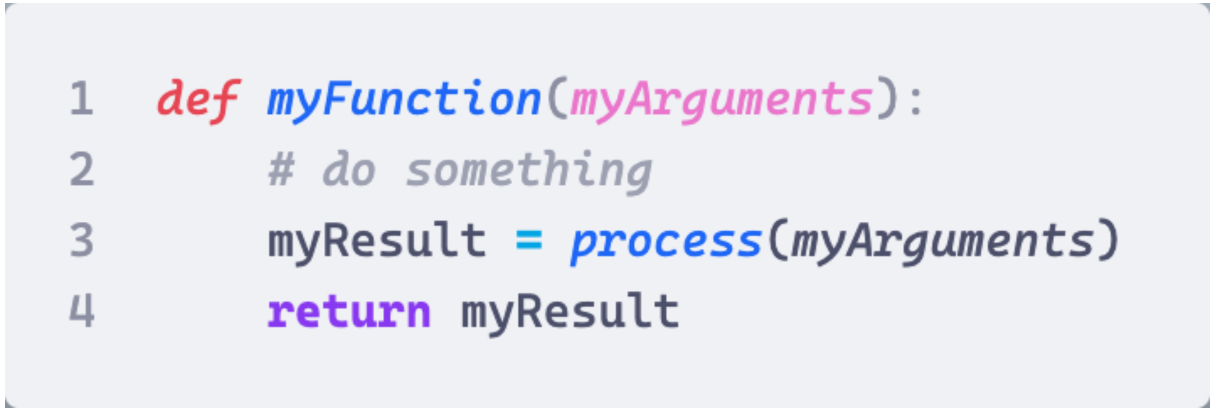
- Very different structure
- Code has underlying logic, which strongly differs from NL

Criteria	Natural Language	Programming Language
Syntax	Unstructured and often ambiguous	Precise and structured
Vocabulary	Vast and diverse	Limited and well-defined
Ambiguity	Common	Rare
Error Tolerance	Tolerant of errors	Strict

Syntax/Ambiguity



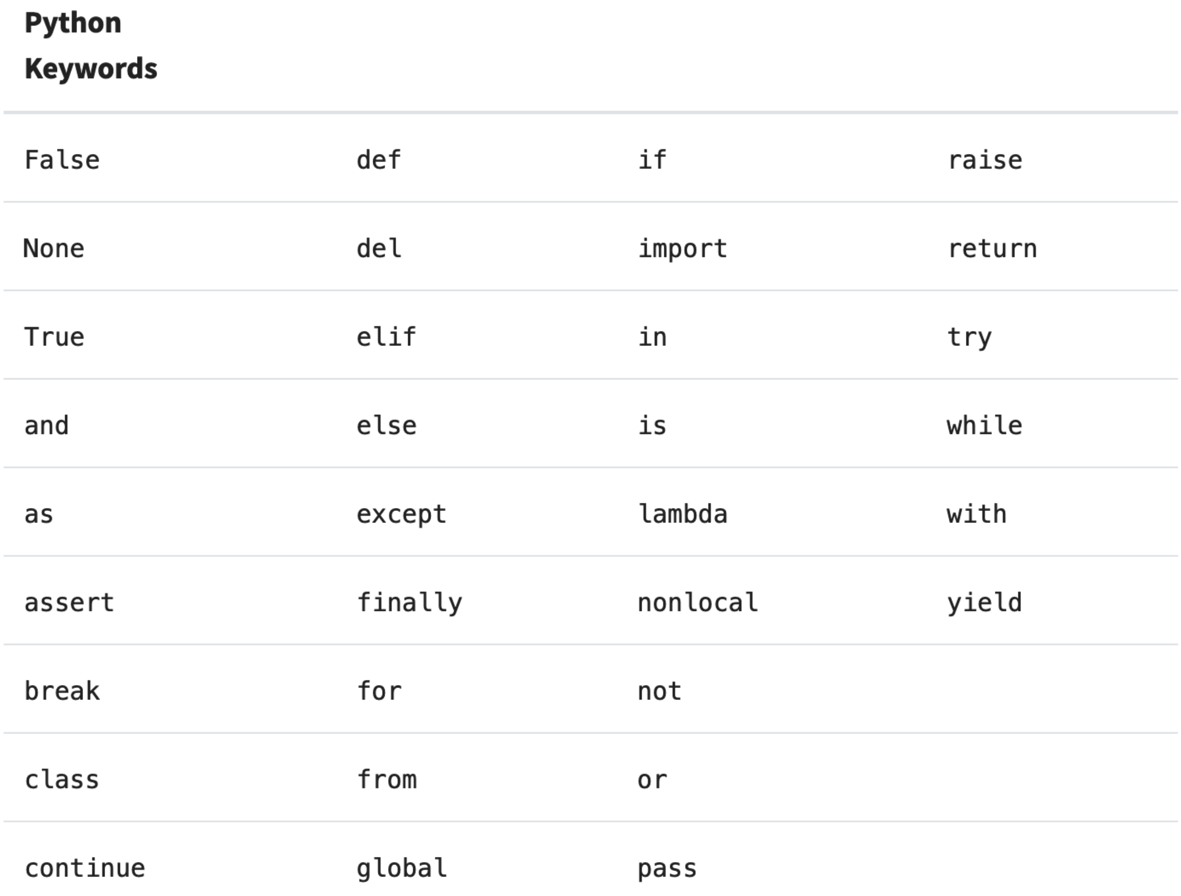
The burglar threatened the student with the knife



```
def myFunction(myArguments):
    # do something
    myResult = process(myArguments)
    return myResult
```

Vocabulary

- Over 150k english words
- Python has 33 reserved keywords

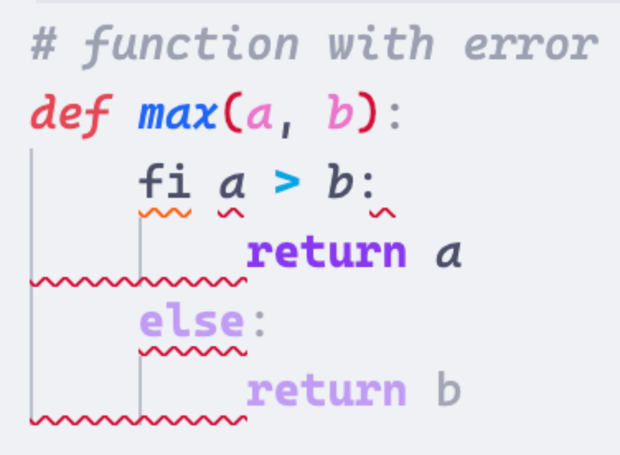


Python Keywords			
False	def	if	raise
None	del	import	return
True	elif	in	try
and	else	is	while
as	except	lambda	with
assert	finally	nonlocal	yield
break	for	not	
class	from	or	
continue	global	pass	

RealPython: https://realpython.com/lessons/reserved-keywords/

Error Tolerance

Yuo cna porbalby raed tihs esaliy desptie teh msispeillgns



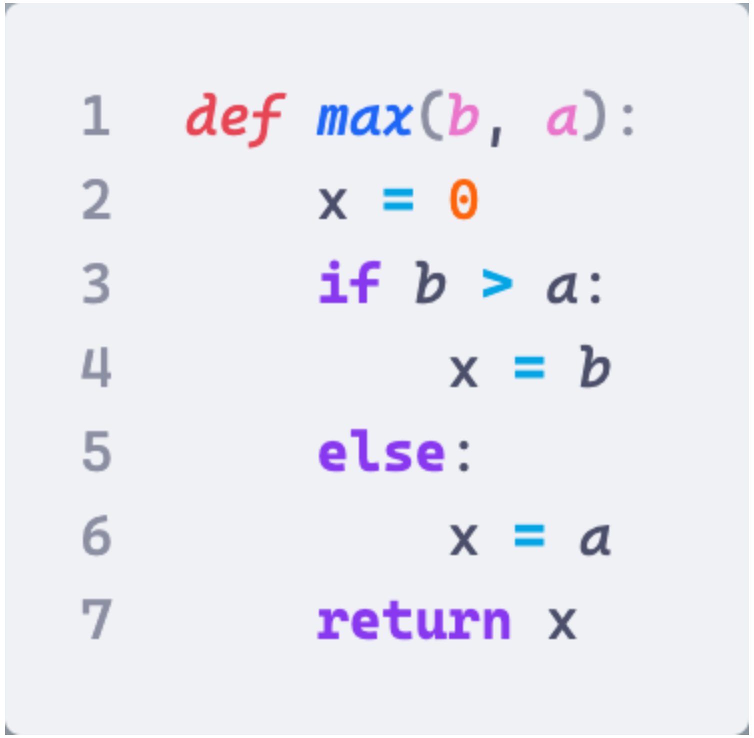
```
# function with error
def max(a, b):
    fi a > b:
        return a
    else:
        return b
```

Natural Language vs. Code

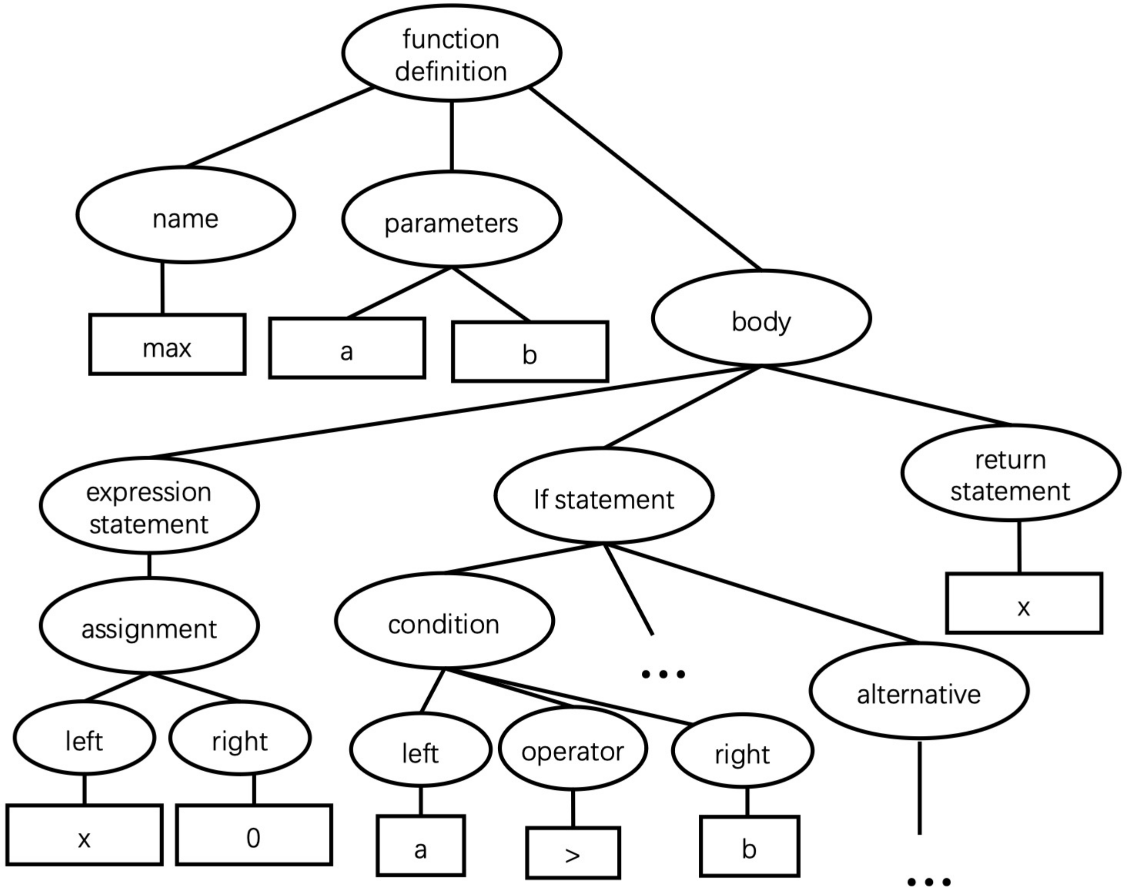
- Very different structure
- Code has underlying logic, which strongly differs from NL

Criteria	Natural Language	Programming Language
Syntax	Unstructured and often ambiguous	Precise and structured
Vocabulary	Vast and diverse	Limited and well-defined
Ambiguity	Common	Rare
Error Tolerance	Tolerant of errors	Strict

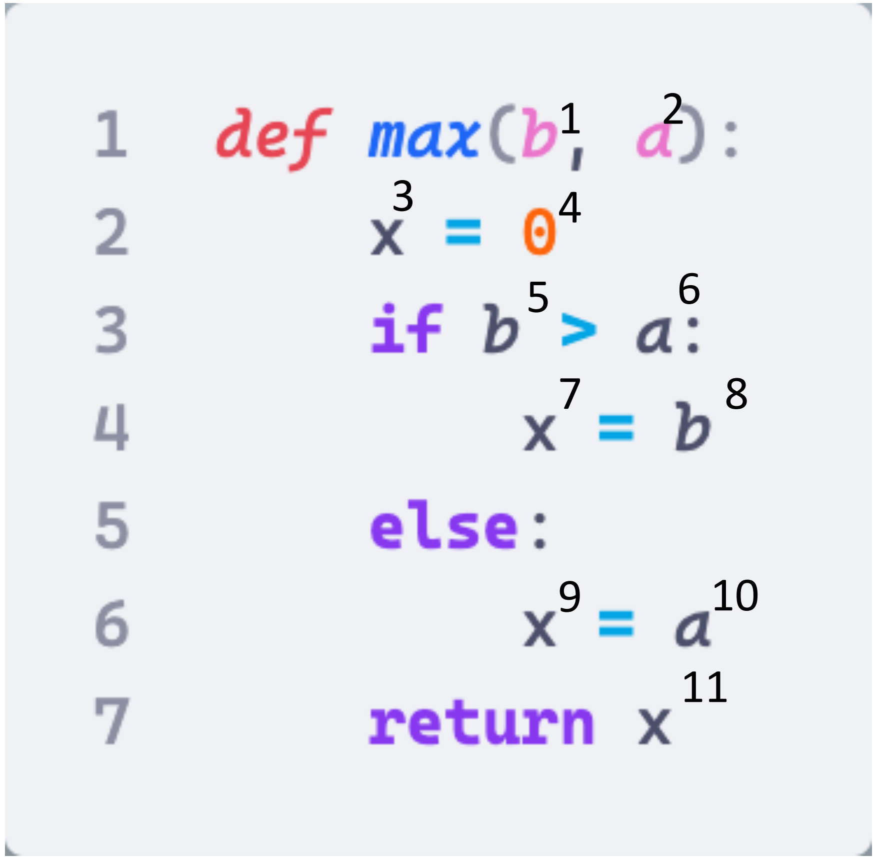
Abstract Syntax Tree



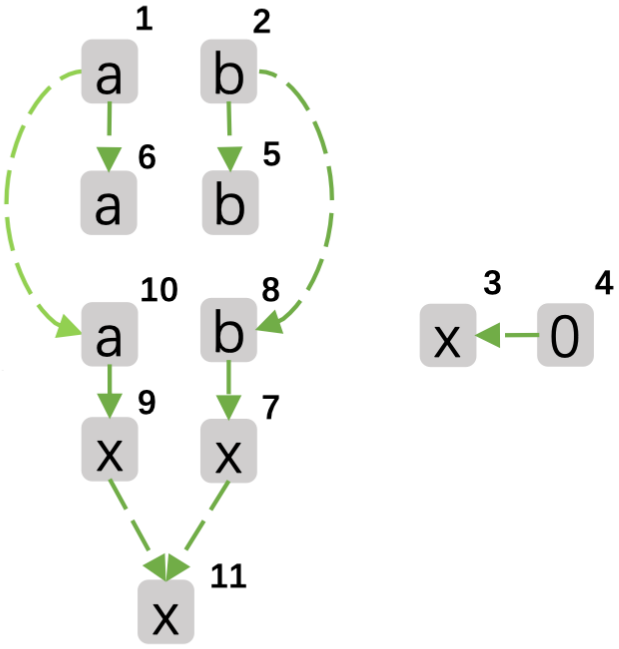
```python
def max(b, a):
    x = 0
    if b > a:
        x = b
    else:
        x = a
    return x
```



Data Flow

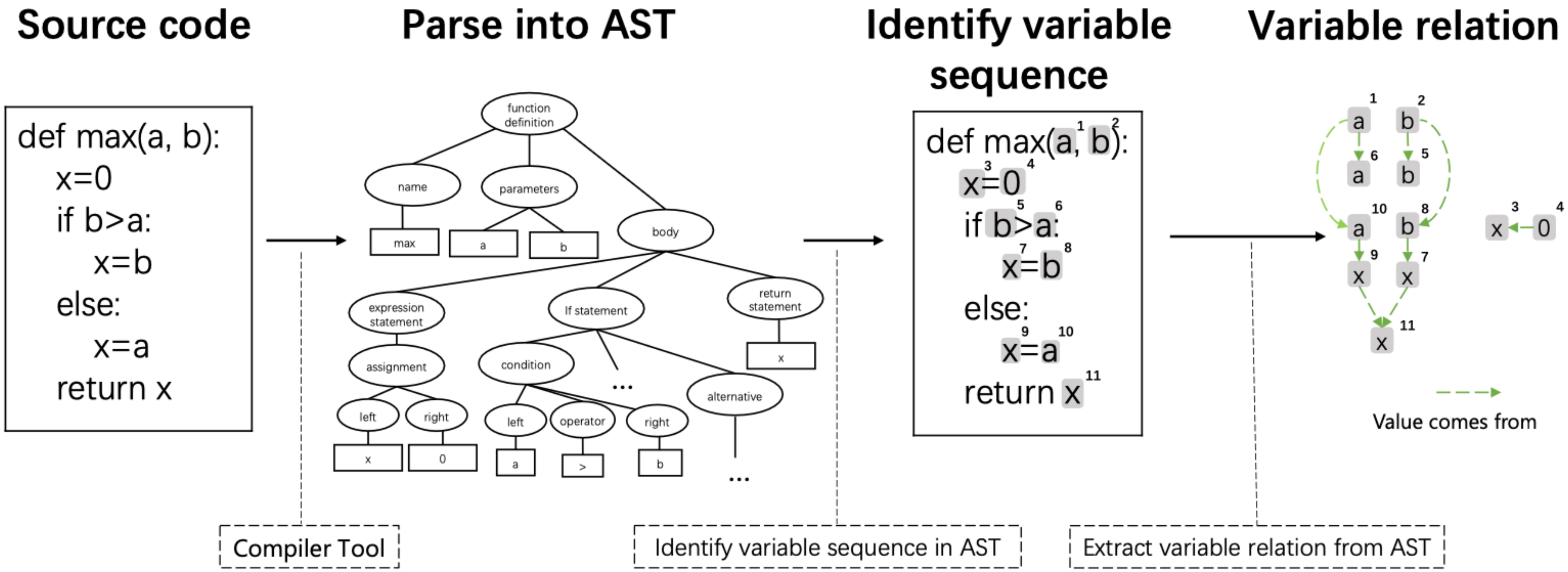


```
def max(b, a):
    x = 0
    if b > a:
        x = b
    else:
        x = a
    return x
```



- - → Value comes from

Data Flow Derivation



Source code

```
def max(a, b):
    x=0
    if b>a:
        x=b
    else:
        x=a
    return x
```

Parse into AST

Identify variable sequence

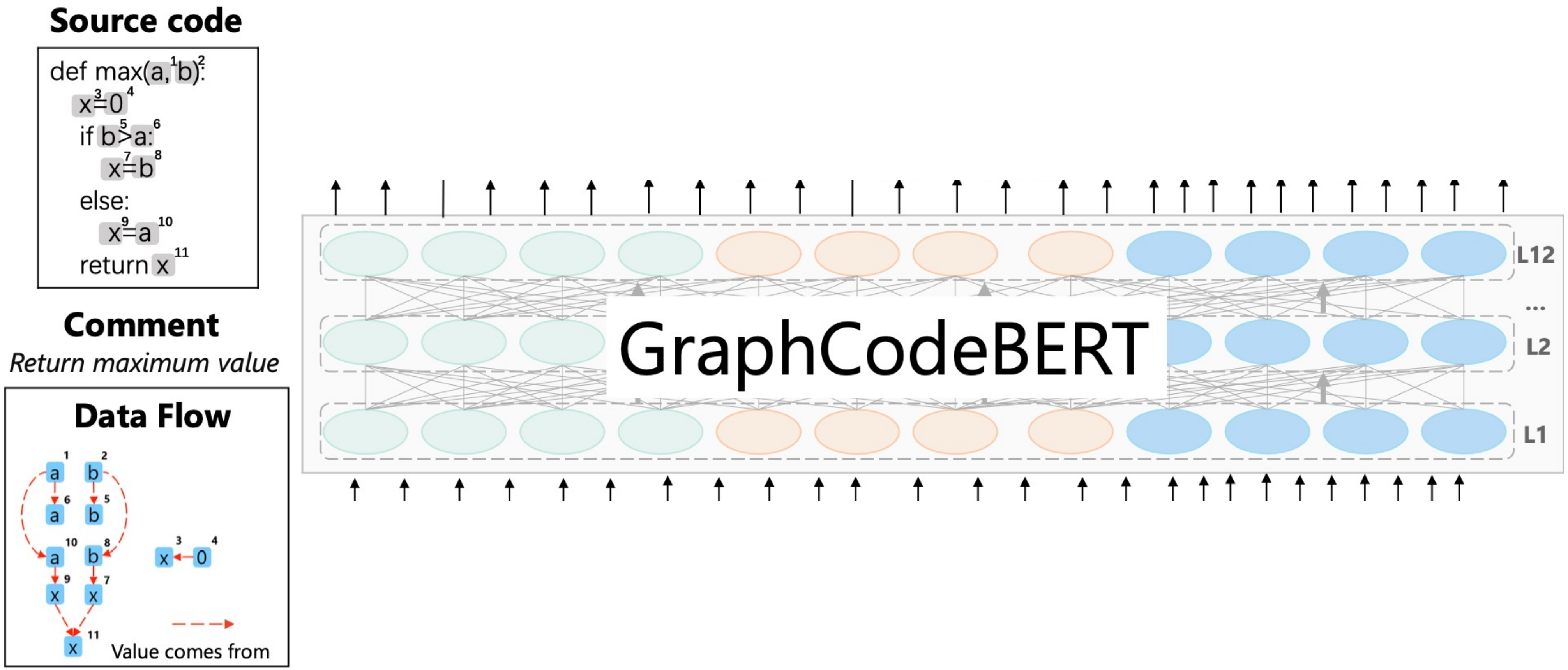
Variable relation

Compiler Tool

Identify variable sequence in AST

Extract variable relation from AST

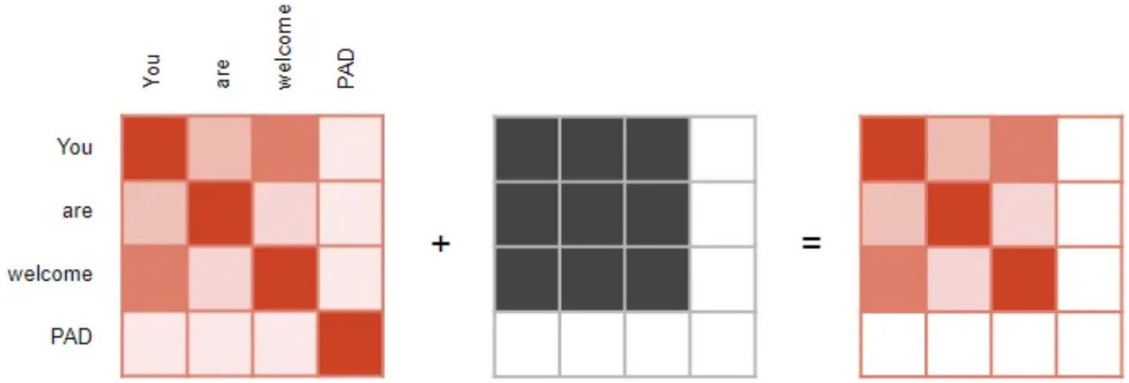
GraphCodeBERT architecture



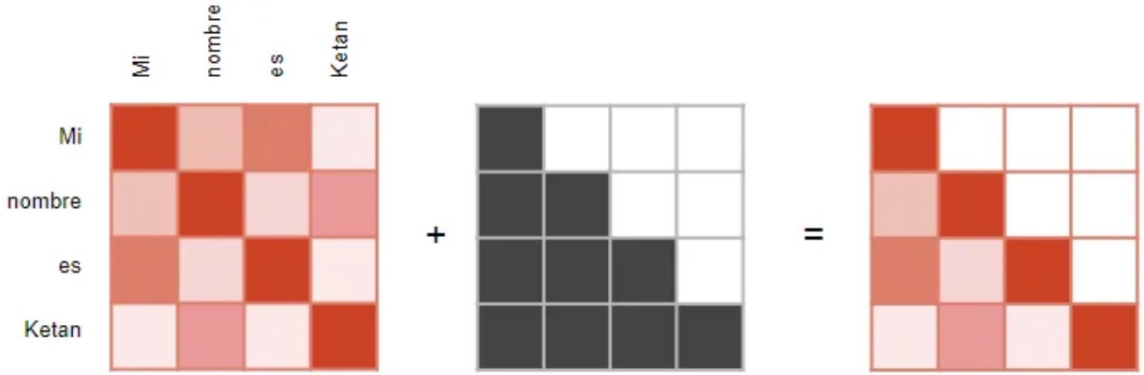
```
def max(a, b):
  x=0
  if b>a:
    x=b
  else:
    x=a
  return x
```

Return maximum value

Attention Mask



Encoder Self-Attention Scores



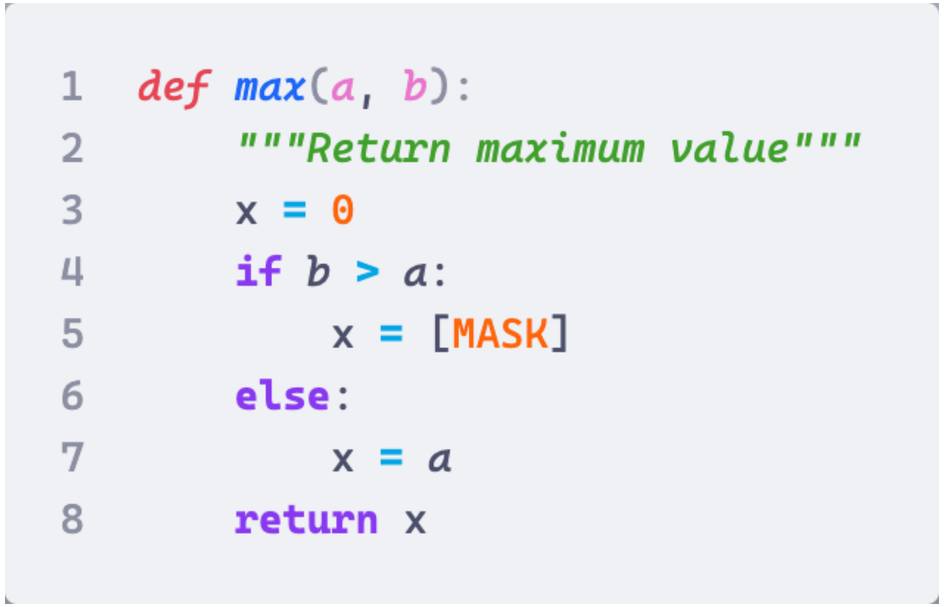
Ketan Doshi: https://towardsdatascience.com/transformers-explained-visually-part-2-how-it-works-step-by-step-b49fa4a64f34

PreTraining Tasks

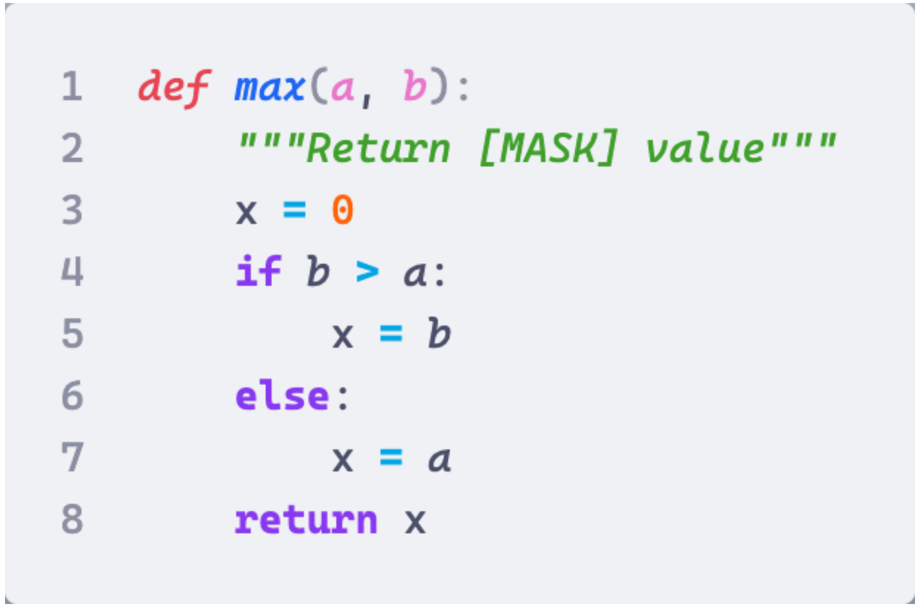
- Masked Language Modeling
- Edge Prediction
- Node Alignment

Masked Language Modelling

- Common way of pretraining
- Follows BERT approach

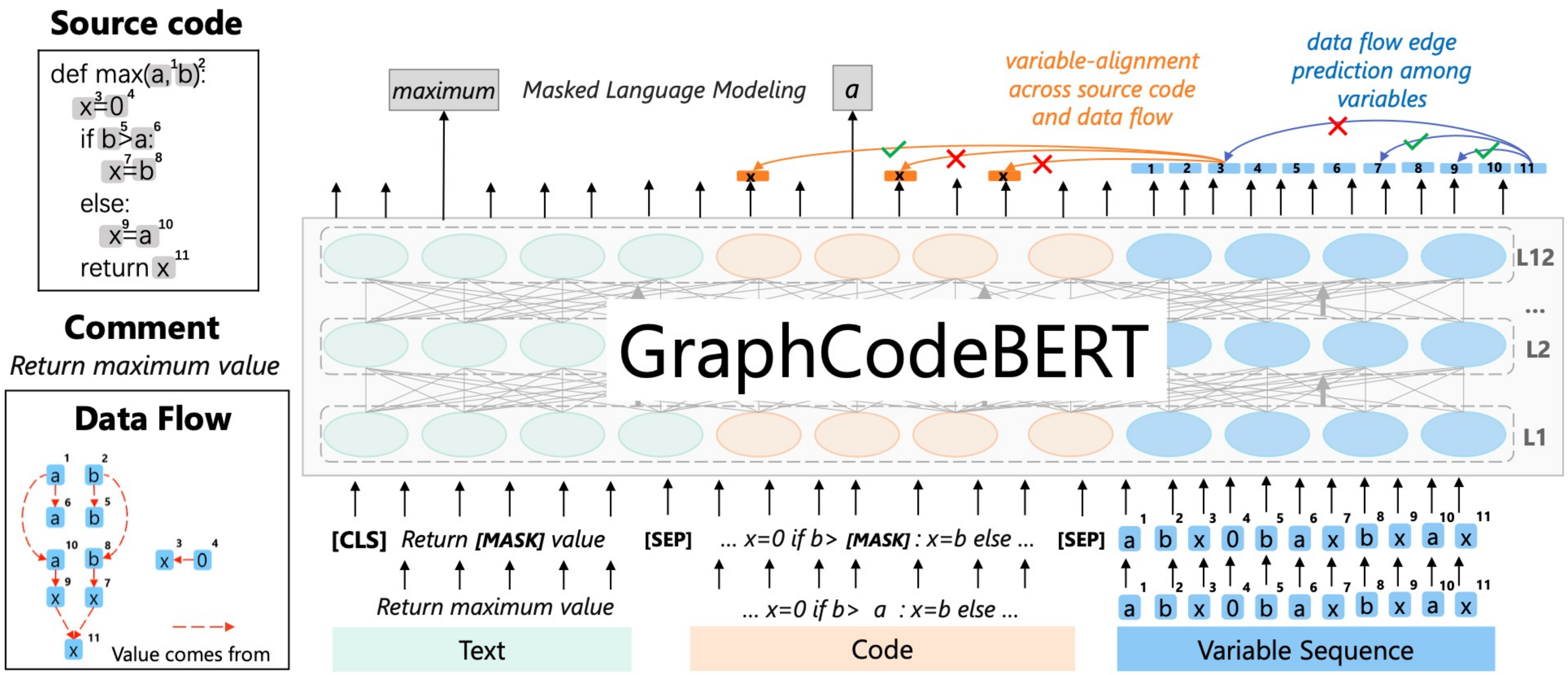


```python
def max(a, b):
    """Return maximum value"""
    x = 0
    if b > a:
        x = [MASK]
    else:
        x = a
    return x
```



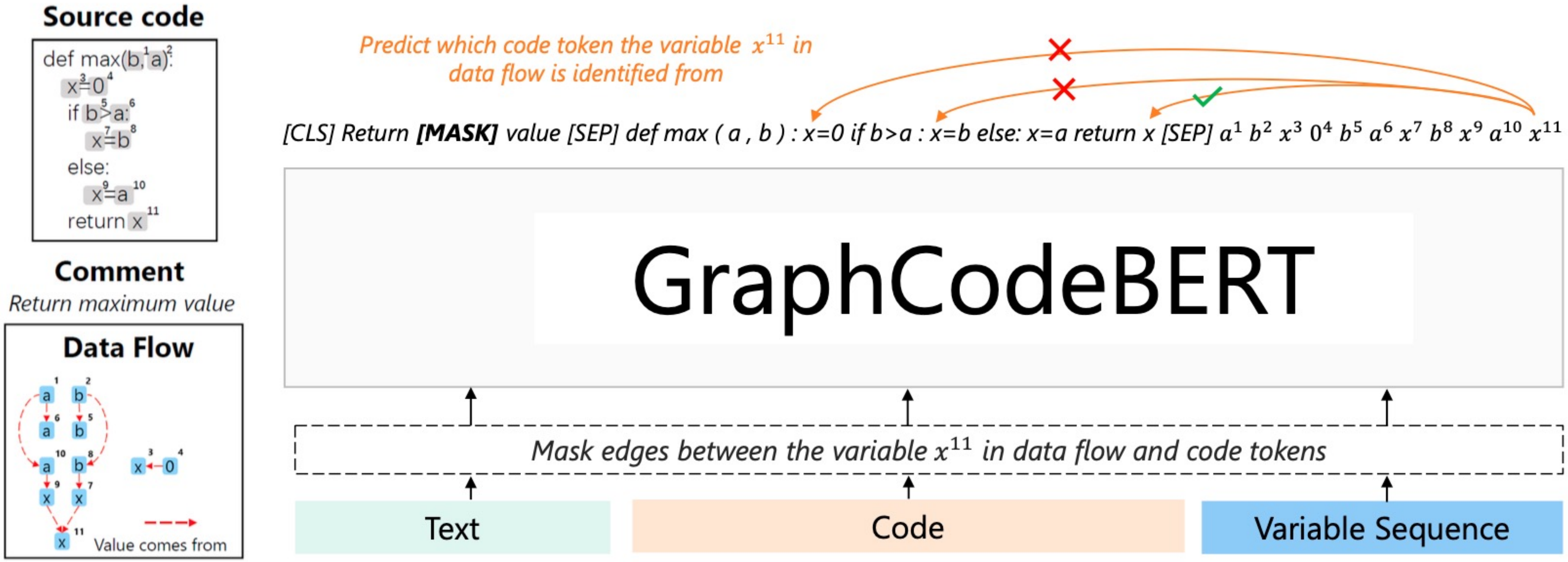
```python
def max(a, b):
    """Return [MASK] value"""
    x = 0
    if b > a:
        x = b
    else:
        x = a
    return x
```

Edge Prediction Pretraining



$$loss_{EdgePred} = -\sum_{e_{ij} \in E_c} [\delta(e_{ij} \in E_{mask}) log p_{e_{ij}} + (1 - \delta(e_{ij} \in E_{mask})) log(1 - p_{e_{ij}})]$$

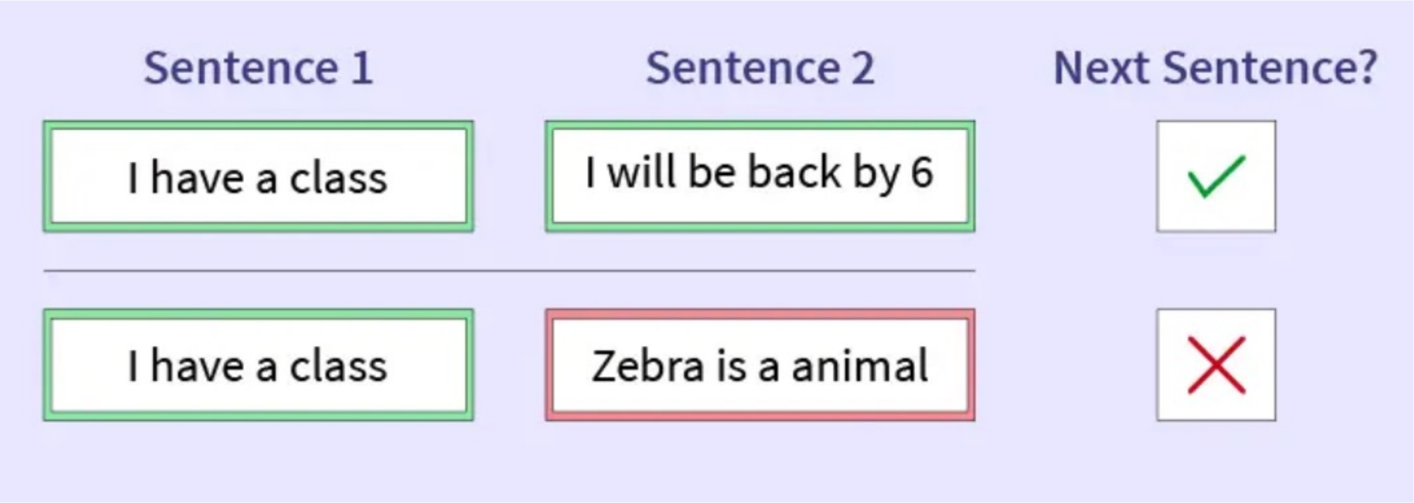
Node Alignment Pretraining


Source code
def max(b, a):
x=0
if b>a:
x=b
else:
x=a
return x
Comment
Return maximum value
Data Flow
a b a b a b x x x x 0
Value comes from
Predict which code token the variable x^{11} in data flow is identified from
[CLS] Return [MASK] value [SEP] def max (a , b) : x=0 if b>a : x=b else: x=a return x [SEP] a^1 b^2 x^3 0^4 b^5 a^6 x^7 b^8 x^9 a^{10} x^{11}
GraphCodeBERT
Mask edges between the variable x^{11} in data flow and code tokens
Text
Code
Variable Sequence

$$loss_{NodeAlign} = -\sum_{e_{ij}\in E'_c}[\delta(e_{ij}\in E'_{mask})logp_{e_{ij}} + (1-\delta(e_{ij}\in E'_{mask}))log(1-p_{e_{ij}})]$$

RoBERTa

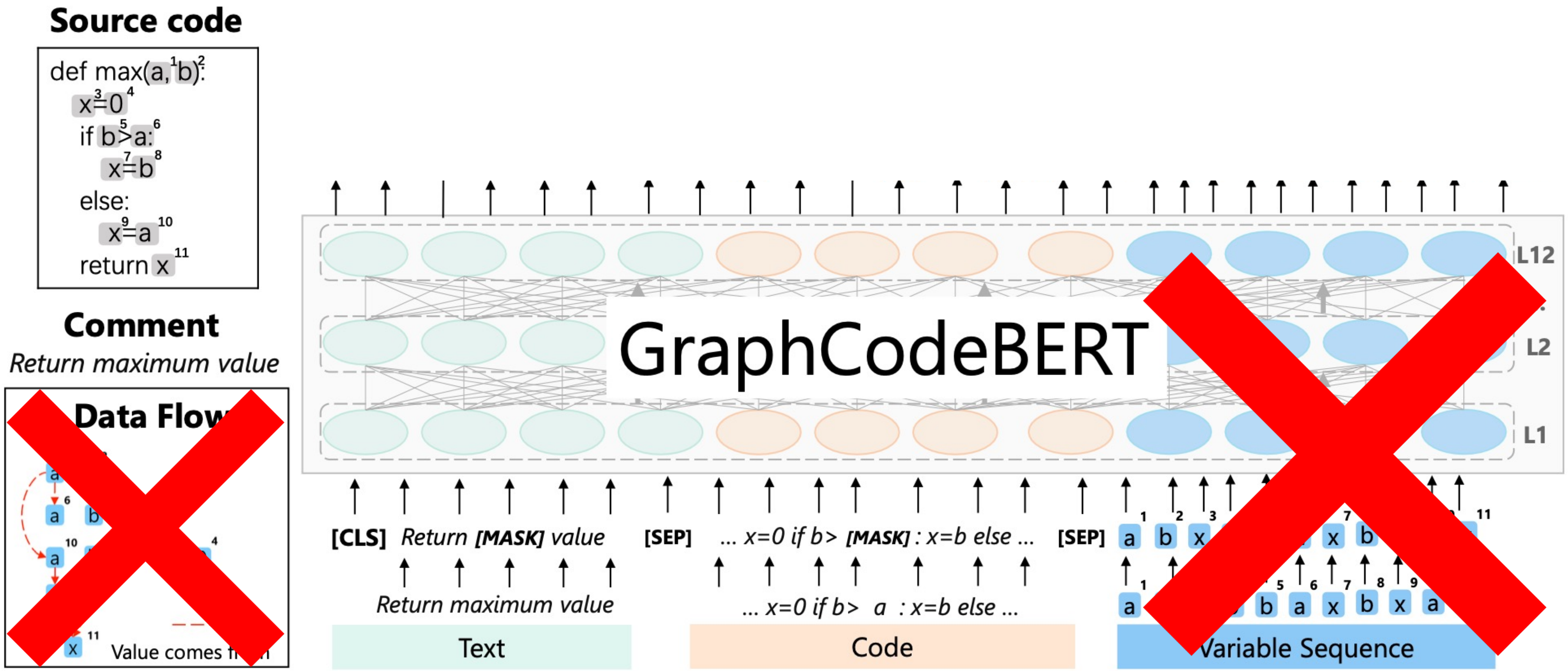
- Train longer, bigger batches and more data
- Remove NSP
- Train on longer sequences
- Dynamically change masking pattern



Sentence 1	Sentence 2	Next Sentence?
I have a class	I will be back by 6	✓
I have a class	Zebra is a animal	✗

Scaler Topics: https://www.scaler.com/topics/nlp/bert-next-sentence-prediction/

CodeBERT



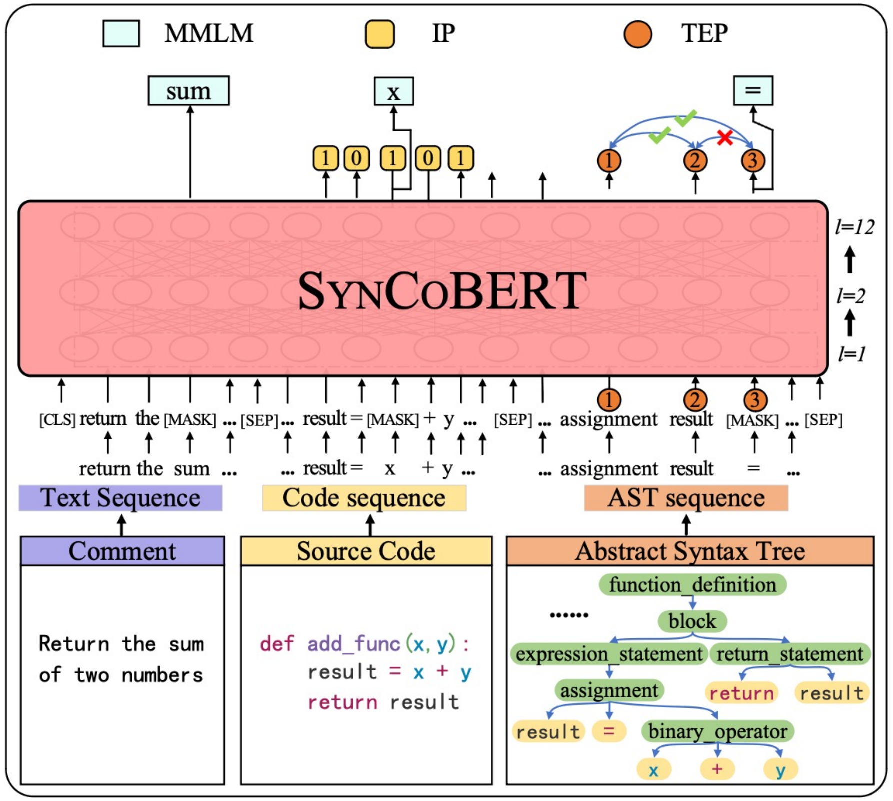
Source code

```
def max(a, b):
  x=0
  if b>a:
    x=b
  else:
    x=a
  return x
```

Comment

Return maximum value

SynCoBERT



Return the sum of two numbers

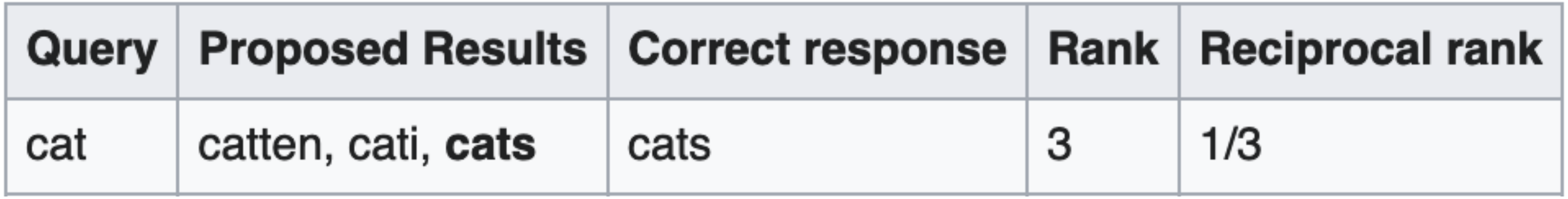
```
def add_func(x, y):
    result = x + y
    return result
```

(a) SYNCOBERT pre-training over MMLM, IP and TEP objectives

SynCoBERT paper: https://arxiv.org/abs/2108.04556

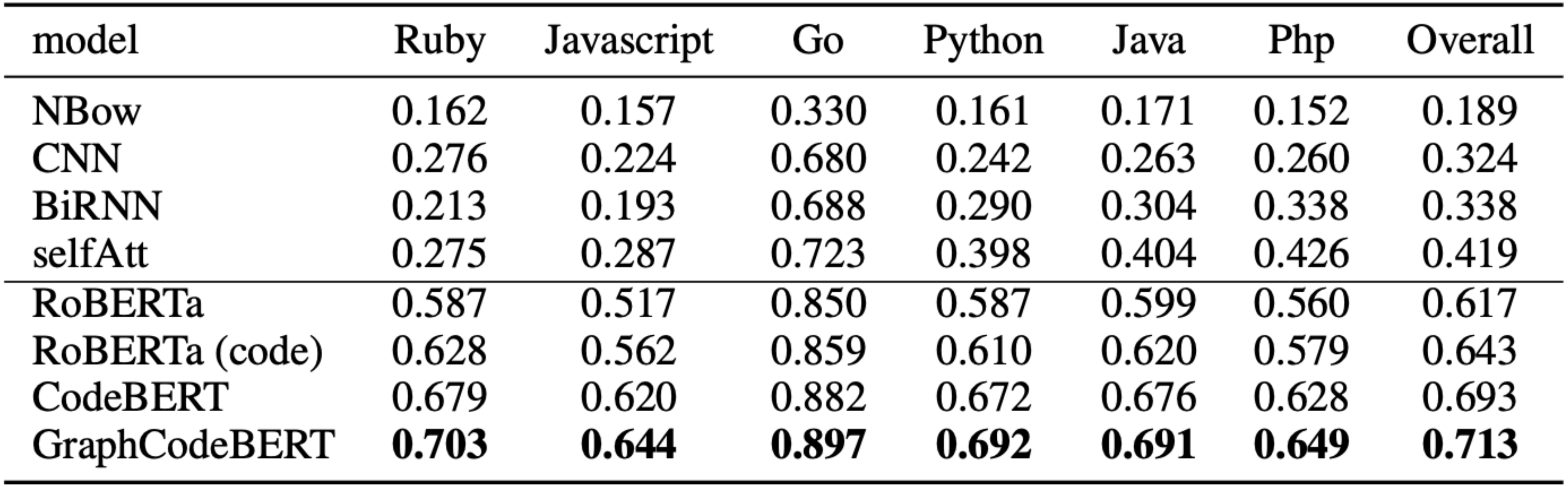
MRR Score

$$\text{MRR} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}_i}.$$



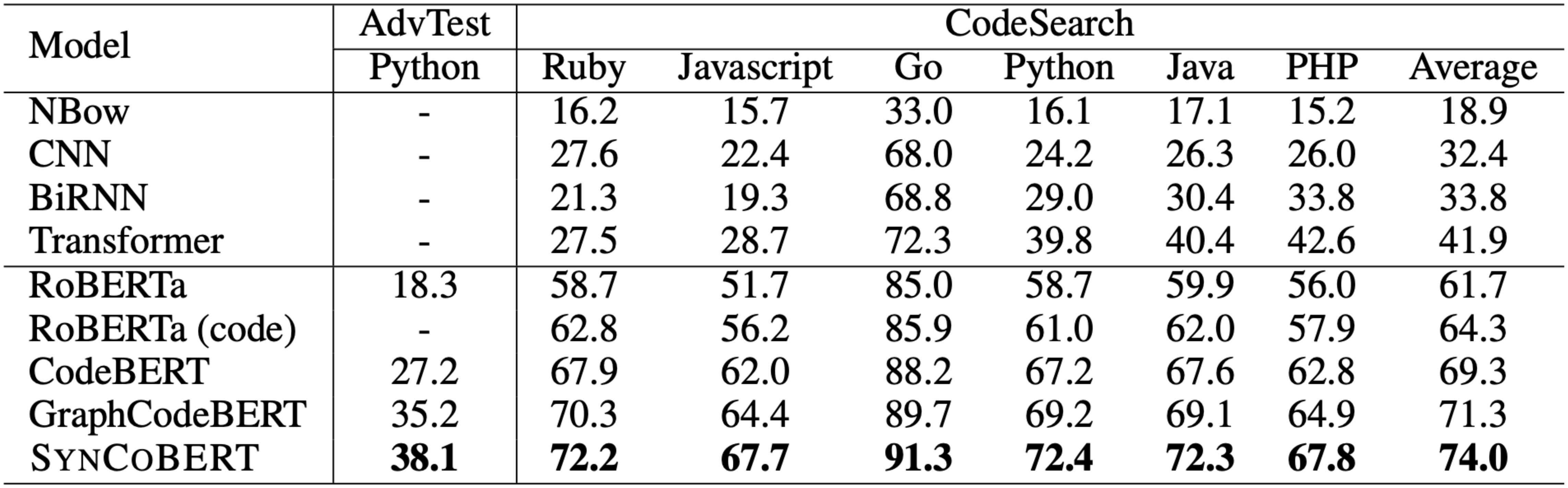
Query	Proposed Results	Correct response	Rank	Reciprocal rank
cat	catten, cati, **cats**	cats	3	1/3

NL Code Search Results



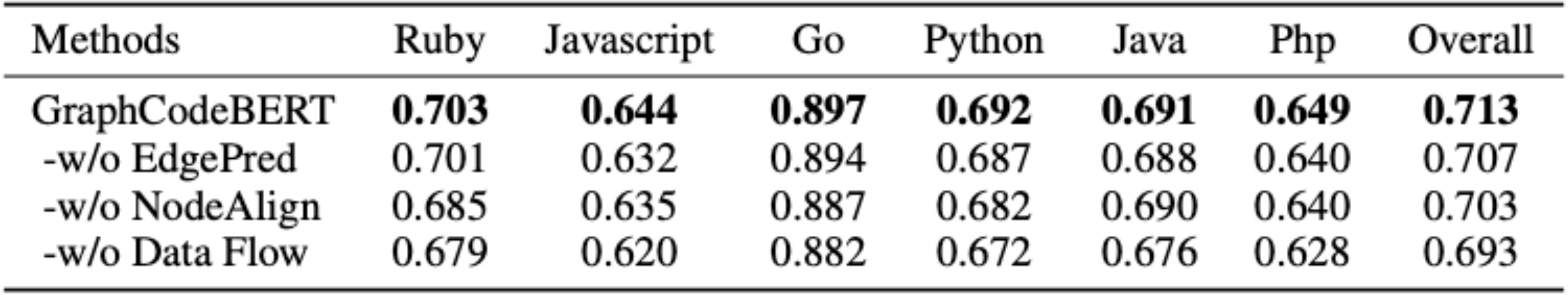
model	Ruby	Javascript	Go	Python	Java	Php	Overall
NBow	0.162	0.157	0.330	0.161	0.171	0.152	0.189
CNN	0.276	0.224	0.680	0.242	0.263	0.260	0.324
BiRNN	0.213	0.193	0.688	0.290	0.304	0.338	0.338
selfAtt	0.275	0.287	0.723	0.398	0.404	0.426	0.419
RoBERTa	0.587	0.517	0.850	0.587	0.599	0.560	0.617
RoBERTa (code)	0.628	0.562	0.859	0.610	0.620	0.579	0.643
CodeBERT	0.679	0.620	0.882	0.672	0.676	0.628	0.693
GraphCodeBERT	**0.703**	**0.644**	**0.897**	**0.692**	**0.691**	**0.649**	**0.713**

SynCoBERT NL Code Search Results



Model	AdvTest	CodeSearch						
	Python	Ruby	Javascript	Go	Python	Java	PHP	Average
NBow	-	16.2	15.7	33.0	16.1	17.1	15.2	18.9
CNN	-	27.6	22.4	68.0	24.2	26.3	26.0	32.4
BiRNN	-	21.3	19.3	68.8	29.0	30.4	33.8	33.8
Transformer	-	27.5	28.7	72.3	39.8	40.4	42.6	41.9
RoBERTa	18.3	58.7	51.7	85.0	58.7	59.9	56.0	61.7
RoBERTa (code)	-	62.8	56.2	85.9	61.0	62.0	57.9	64.3
CodeBERT	27.2	67.9	62.0	88.2	67.2	67.6	62.8	69.3
GraphCodeBERT	35.2	70.3	64.4	89.7	69.2	69.1	64.9	71.3
SynCoBERT	**38.1**	**72.2**	**67.7**	**91.3**	**72.4**	**72.3**	**67.8**	**74.0**

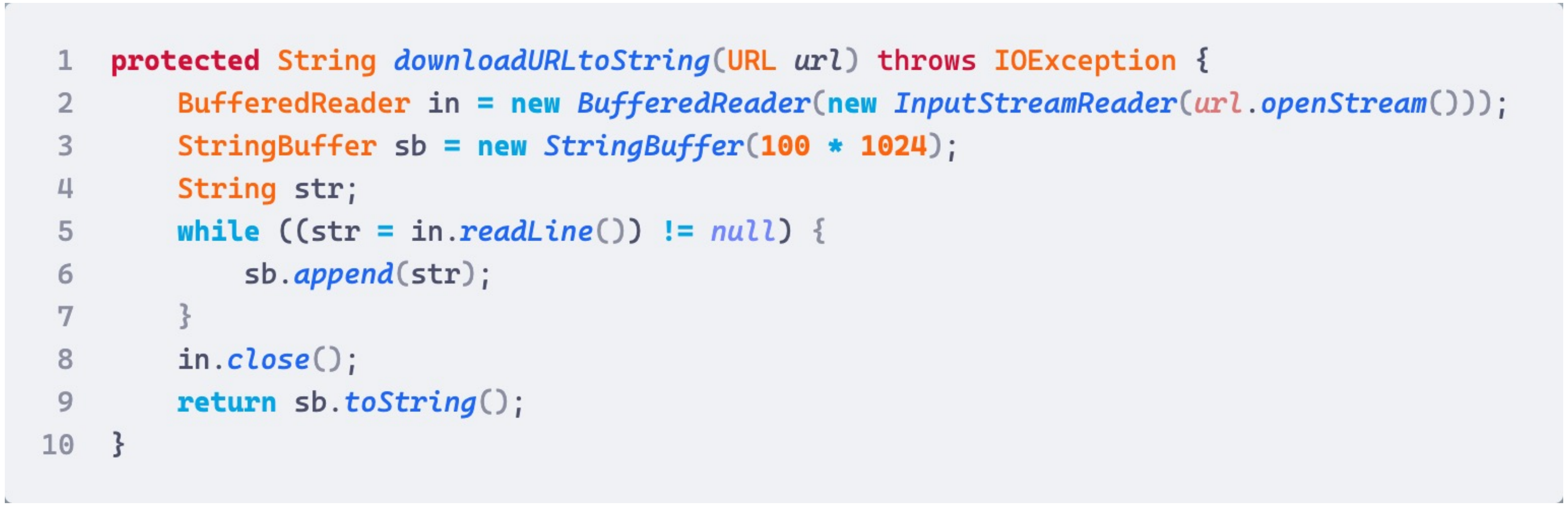
Ablation Study



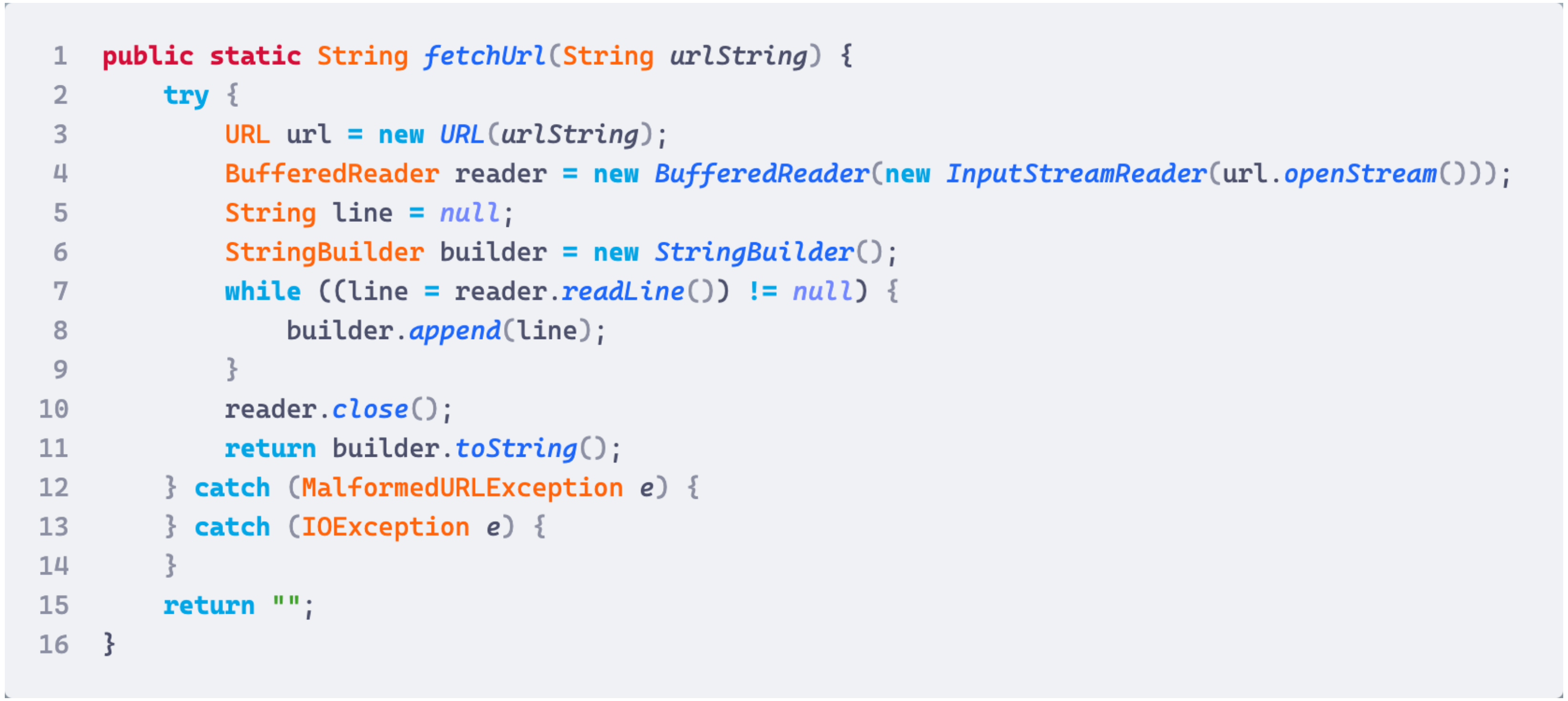
Methods	Ruby	Javascript	Go	Python	Java	Php	Overall
GraphCodeBERT	**0.703**	**0.644**	**0.897**	**0.692**	**0.691**	**0.649**	**0.713**
-w/o EdgePred	0.701	0.632	0.894	0.687	0.688	0.640	0.707
-w/o NodeAlign	0.685	0.635	0.887	0.682	0.690	0.640	0.703
-w/o Data Flow	0.679	0.620	0.882	0.672	0.676	0.628	0.693

Code Clone Detection

- Measure similarity between two code segments
- Code segments have similar output for the same input
- Easier software maintenance and to prevent bugs
- BigCloneBench dataset



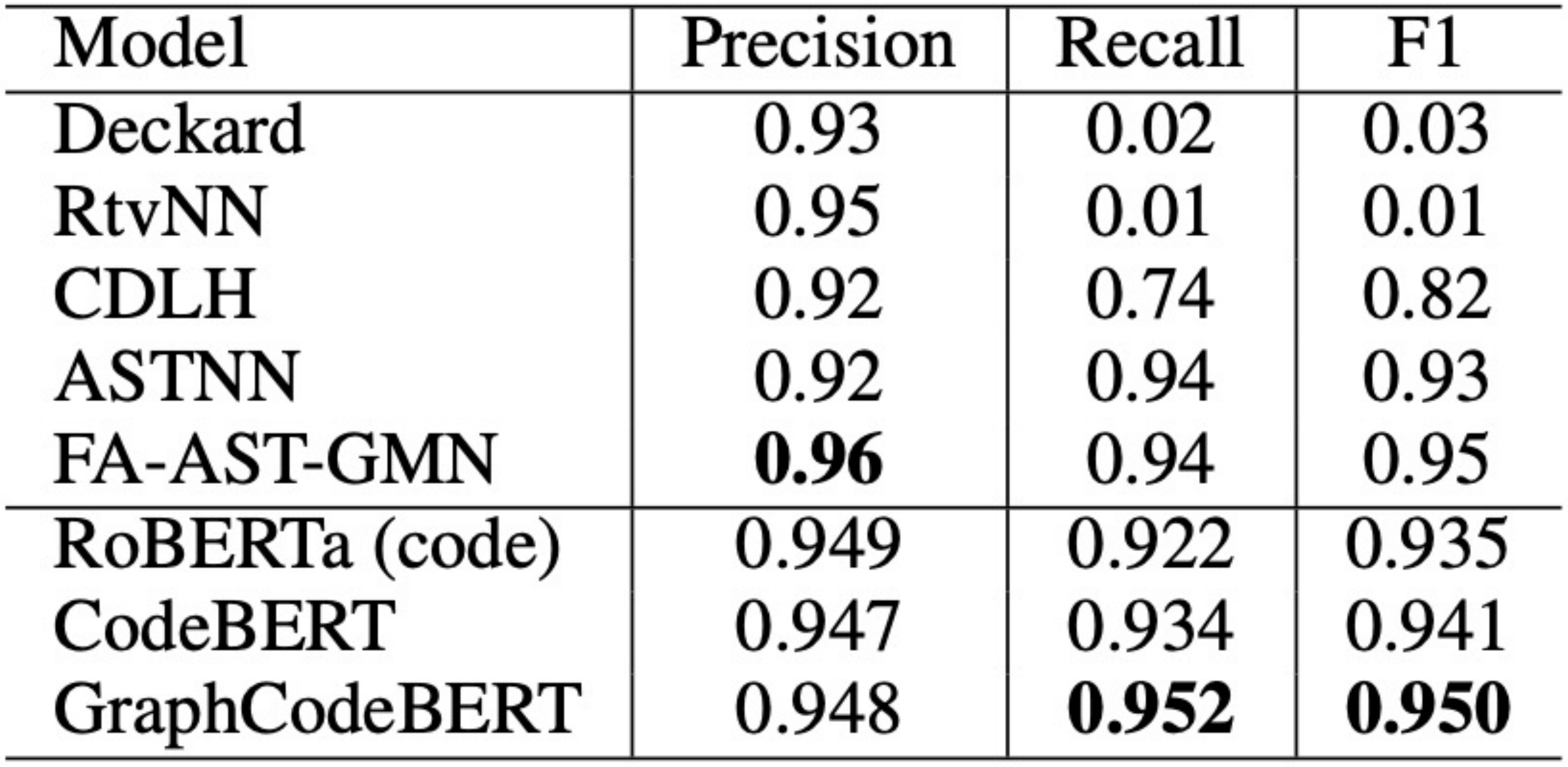
```java
protected String downloadURLtoString(URL url) throws IOException {
    BufferedReader in = new BufferedReader(new InputStreamReader(url.openStream()));
    StringBuffer sb = new StringBuffer(100 * 1024);
    String str;
    while ((str = in.readLine()) != null) {
        sb.append(str);
    }
    in.close();
    return sb.toString();
}
```



```java
public static String fetchUrl(String urlString) {
    try {
        URL url = new URL(urlString);
        BufferedReader reader = new BufferedReader(new InputStreamReader(url.openStream()));
        String line = null;
        StringBuilder builder = new StringBuilder();
        while ((line = reader.readLine()) != null) {
            builder.append(line);
        }
        reader.close();
        return builder.toString();
    } catch (MalformedURLException e) {
    } catch (IOException e) {
    }
    return "";
}
```

BLEU Score: 0.983

Clone Detection Results



Model	Precision	Recall	F1
Deckard	0.93	0.02	0.03
RtvNN	0.95	0.01	0.01
CDLH	0.92	0.74	0.82
ASTNN	0.92	0.94	0.93
FA-AST-GMN	**0.96**	0.94	0.95
RoBERTa (code)	0.949	0.922	0.935
CodeBERT	0.947	0.934	0.941
GraphCodeBERT	0.948	**0.952**	**0.950**

Code Translation

- Migrate code to different language
- Mostly used for legacy software
- BLEU score
- Dataset crawled from open source projects

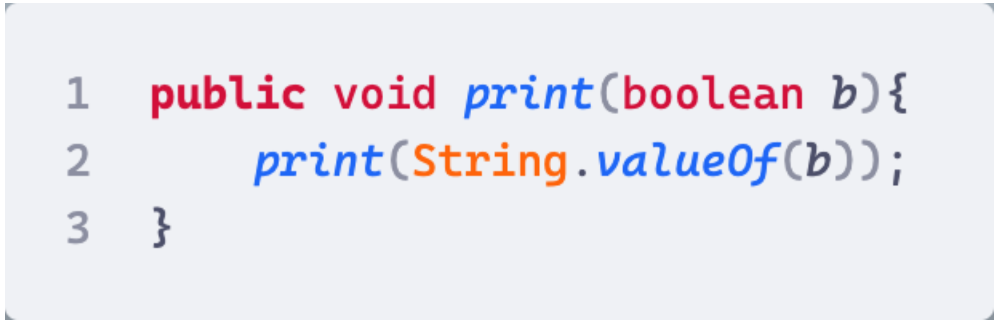
BLEU Score

- Bilingual evaluation understudy
- Measure for similarity of machine translated text
- Based on frequency of shared words and phrases
- Frequencies are compared with reference corpus

The closer a machine translation is to a professional human translation, the better it is

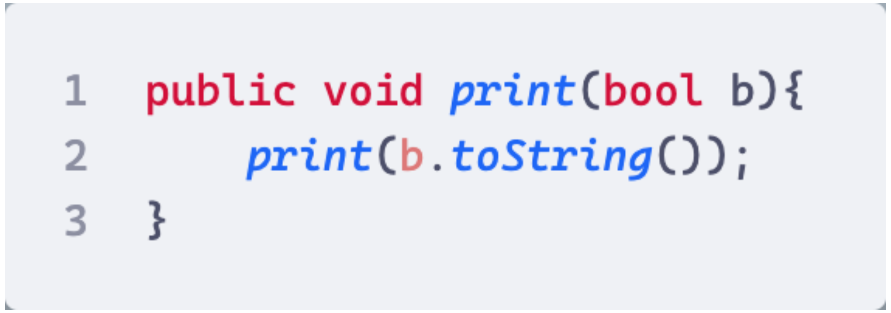
BLEU Paper: https://aclanthology.org/P02-1040.pdf

Java



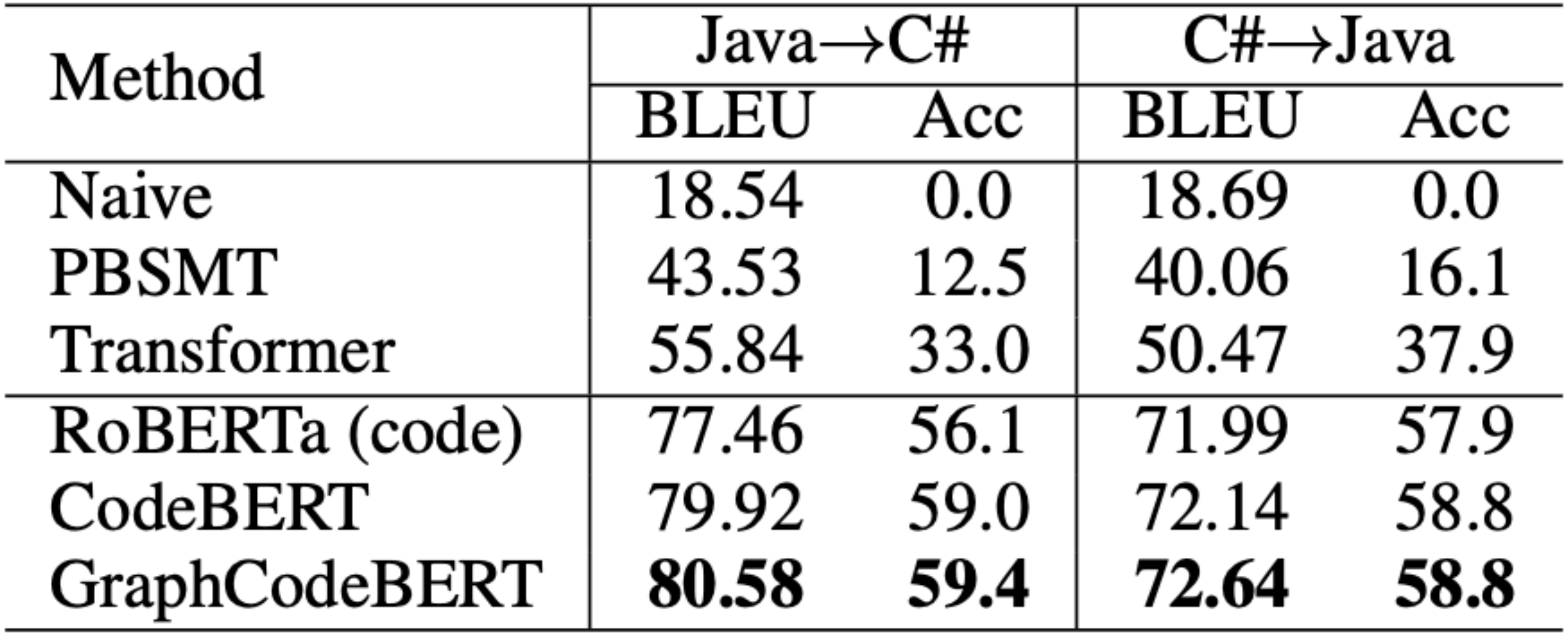
```java
public void print(boolean b){
    print(String.valueOf(b));
}
```

C#



```csharp
public void print(bool b){
    print(b.toString());
}
```

Code Translation Results

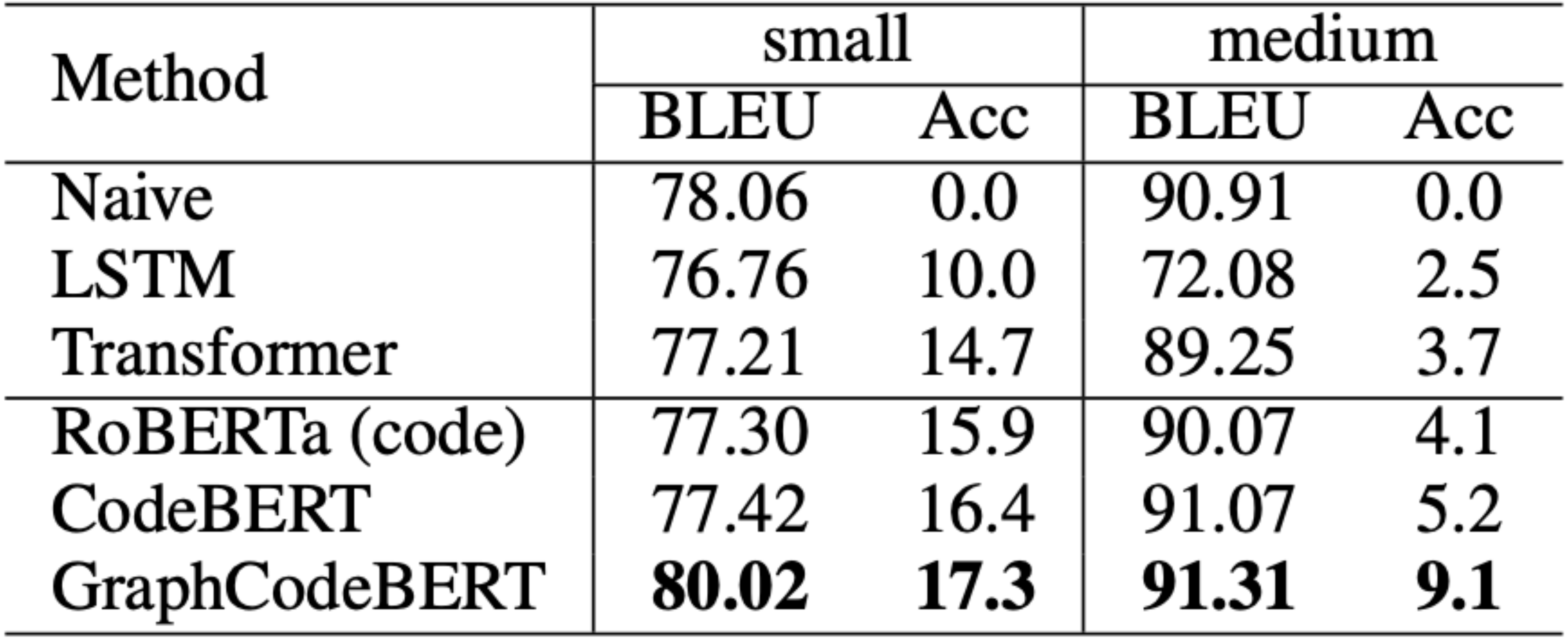


Method	Java→C#		C#→Java	
	BLEU	Acc	BLEU	Acc
Naive	18.54	0.0	18.69	0.0
PBSMT	43.53	12.5	40.06	16.1
Transformer	55.84	33.0	50.47	37.9
RoBERTa (code)	77.46	56.1	71.99	57.9
CodeBERT	79.92	59.0	72.14	58.8
GraphCodeBERT	**80.58**	**59.4**	**72.64**	**58.8**

Code Refinement

- Automatically fix bugs
- Reduces cost of bug fixes
- Dataset by Tufano et al.

Code Refinment Results



Method	small		medium	
	BLEU	Acc	BLEU	Acc
Naive	78.06	0.0	90.91	0.0
LSTM	76.76	10.0	72.08	2.5
Transformer	77.21	14.7	89.25	3.7
RoBERTa (code)	77.30	15.9	90.07	4.1
CodeBERT	77.42	16.4	91.07	5.2
GraphCodeBERT	**80.02**	**17.3**	**91.31**	**9.1**

Conclusion/Remarks

- Small improvement compared to CodeBERT
- SynCoBERT outperformed the model
- Not clear benefits of Data Flow when compared with AST
- BLEU score is not really justified