

Appendix B

Results Appendix

K	W	V	T	C	R	Mean Accuracy
1	10	TF	Standard	Source (SM)	Text	0.39
2	10	TF	Standard	Source (SM)	Text	0.53
3	10	TF	Standard	Source (SM)	Text	0.64
1	20	TF	Standard	Source (SM)	Text	0.30
1	50	TF	Standard	Source (SM)	Text	0.23
1	10	TF	Standard	Source (MD)	Text	0.48
2	10	TF	Standard	Source (MD)	Text	0.62
3	10	TF	Standard	Source (MD)	Text	0.73
1	20	TF	Standard	Source (MD)	Text	0.37
1	50	TF	Standard	Source (MD)	Text	0.28
1	10	TF	Standard	Source (LG)	Text	0.47
2	10	TF	Standard	Source (LG)	Text	0.61
3	10	TF	Standard	Source (LG)	Text	0.70
1	20	TF	Standard	Source (LG)	Text	0.35
1	50	TF	Standard	Source (LG)	Text	0.28
1	10	TF	Standard	Diff	Text	0.59
2	10	TF	Standard	Diff	Text	0.67
3	10	TF	Standard	Diff	Text	0.74
1	20	TF	Standard	Diff	Text	0.52
1	50	TF	Standard	Diff	Text	0.43
1	10	TF	Standard	Combined	Text	0.52

Continued on next page

Table B.1: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
2	10	TF	Standard	Combined	Text	0.69
3	10	TF	Standard	Combined	Text	0.78
1	20	TF	Standard	Combined	Text	0.41
1	50	TF	Standard	Combined	Text	0.32
1	10	TF	Subword	Source (SM)	Text	0.50
2	10	TF	Subword	Source (SM)	Text	0.64
3	10	TF	Subword	Source (SM)	Text	0.78
1	20	TF	Subword	Source (SM)	Text	0.45
1	50	TF	Subword	Source (SM)	Text	0.42
1	10	TF	Subword	Source (MD)	Text	0.56
2	10	TF	Subword	Source (MD)	Text	0.70
3	10	TF	Subword	Source (MD)	Text	0.82
1	20	TF	Subword	Source (MD)	Text	0.49
1	50	TF	Subword	Source (MD)	Text	0.41
1	10	TF	Subword	Source (LG)	Text	0.58
2	10	TF	Subword	Source (LG)	Text	0.72
3	10	TF	Subword	Source (LG)	Text	0.81
1	20	TF	Subword	Source (LG)	Text	0.49
1	50	TF	Subword	Source (LG)	Text	0.42
1	10	TF	Subword	Diff	Text	0.61
2	10	TF	Subword	Diff	Text	0.72
3	10	TF	Subword	Diff	Text	0.79
1	20	TF	Subword	Diff	Text	0.53
1	50	TF	Subword	Diff	Text	0.46
1	10	TF	Subword	Combined	Text	0.62
2	10	TF	Subword	Combined	Text	0.75
3	10	TF	Subword	Combined	Text	0.82
1	20	TF	Subword	Combined	Text	0.53
1	50	TF	Subword	Combined	Text	0.47

Continued on next page

Table B.1: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
1	10	TF-IDF	Standard	Source (SM)	Text	0.59
2	10	TF-IDF	Standard	Source (SM)	Text	0.67
3	10	TF-IDF	Standard	Source (SM)	Text	0.73
1	20	TF-IDF	Standard	Source (SM)	Text	0.54
1	50	TF-IDF	Standard	Source (SM)	Text	0.47
1	10	TF-IDF	Standard	Source (MD)	Text	0.71
2	10	TF-IDF	Standard	Source (MD)	Text	0.77
3	10	TF-IDF	Standard	Source (MD)	Text	0.83
1	20	TF-IDF	Standard	Source (MD)	Text	0.66
1	50	TF-IDF	Standard	Source (MD)	Text	0.57
1	10	TF-IDF	Standard	Source (LG)	Text	0.70
2	10	TF-IDF	Standard	Source (LG)	Text	0.80
3	10	TF-IDF	Standard	Source (LG)	Text	0.83
1	20	TF-IDF	Standard	Source (LG)	Text	0.64
1	50	TF-IDF	Standard	Source (LG)	Text	0.56
1	10	TF-IDF	Standard	Diff	Text	0.73
2	10	TF-IDF	Standard	Diff	Text	0.83
3	10	TF-IDF	Standard	Diff	Text	0.85
1	20	TF-IDF	Standard	Diff	Text	0.64
1	50	TF-IDF	Standard	Diff	Text	0.56
1	10	TF-IDF	Standard	Combined	Text	0.78
2	10	TF-IDF	Standard	Combined	Text	0.89
3	10	TF-IDF	Standard	Combined	Text	0.91
1	20	TF-IDF	Standard	Combined	Text	0.70
1	50	TF-IDF	Standard	Combined	Text	0.60
1	10	TF-IDF	Subword	Source (SM)	Text	0.75
2	10	TF-IDF	Subword	Source (SM)	Text	0.80
3	10	TF-IDF	Subword	Source (SM)	Text	0.82
1	20	TF-IDF	Subword	Source (SM)	Text	0.68

Continued on next page

Table B.1: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
1	50	TF-IDF	Subword	Source (SM)	Text	0.61
1	10	TF-IDF	Subword	Source (MD)	Text	0.77
2	10	TF-IDF	Subword	Source (MD)	Text	0.84
3	10	TF-IDF	Subword	Source (MD)	Text	0.87
1	20	TF-IDF	Subword	Source (MD)	Text	0.70
1	50	TF-IDF	Subword	Source (MD)	Text	0.65
1	10	TF-IDF	Subword	Source (LG)	Text	0.78
2	10	TF-IDF	Subword	Source (LG)	Text	0.85
3	10	TF-IDF	Subword	Source (LG)	Text	0.88
1	20	TF-IDF	Subword	Source (LG)	Text	0.70
1	50	TF-IDF	Subword	Source (LG)	Text	0.64
1	10	TF-IDF	Subword	Diff	Text	0.73
2	10	TF-IDF	Subword	Diff	Text	0.81
3	10	TF-IDF	Subword	Diff	Text	0.87
1	20	TF-IDF	Subword	Diff	Text	0.66
1	50	TF-IDF	Subword	Diff	Text	0.57
1	10	TF-IDF	Subword	Combined	Text	0.82
2	10	TF-IDF	Subword	Combined	Text	0.89
3	10	TF-IDF	Subword	Combined	Text	0.93
1	20	TF-IDF	Subword	Combined	Text	0.73
1	50	TF-IDF	Subword	Combined	Text	0.65
1	10	CodeBERT (Flattening)	-	Source (SM)	Code	0.17
2	10	CodeBERT (Flattening)	-	Source (SM)	Code	0.25
3	10	CodeBERT (Flattening)	-	Source (SM)	Code	0.36
1	20	CodeBERT (Flattening)	-	Source (SM)	Code	0.09
1	50	CodeBERT (Flattening)	-	Source (SM)	Code	0.04
1	10	CodeBERT (Flattening)	-	Source (MD)	Code	0.15
2	10	CodeBERT (Flattening)	-	Source (MD)	Code	0.21
3	10	CodeBERT (Flattening)	-	Source (MD)	Code	0.32

Continued on next page

Table B.1: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
1	20	CodeBERT (Flattening)	-	Source (MD)	Code	0.08
1	50	CodeBERT (Flattening)	-	Source (MD)	Code	0.04
1	10	CodeBERT (Flattening)	-	Source (LG)	Code	0.09
2	10	CodeBERT (Flattening)	-	Source (LG)	Code	0.20
3	10	CodeBERT (Flattening)	-	Source (LG)	Code	0.31
1	20	CodeBERT (Flattening)	-	Source (LG)	Code	0.04
1	50	CodeBERT (Flattening)	-	Source (LG)	Code	0.02
1	10	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.17
2	10	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.25
3	10	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.36
1	20	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.09
1	50	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.04
1	10	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.15
2	10	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.22
3	10	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.32
1	20	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.09
1	50	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.04
1	10	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.06
2	10	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.14
3	10	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.24
1	20	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.03
1	50	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.02

Table B.1: Result Runs for java-design-patterns by iluwatar

K	W	V	T	C	R	Mean Accuracy
1	10	TF	Standard	Source (SM)	Text	0.37
2	10	TF	Standard	Source (SM)	Text	0.53
3	10	TF	Standard	Source (SM)	Text	0.63
1	20	TF	Standard	Source (SM)	Text	0.25
1	50	TF	Standard	Source (SM)	Text	0.16
1	10	TF	Standard	Source (MD)	Text	0.43
2	10	TF	Standard	Source (MD)	Text	0.59
3	10	TF	Standard	Source (MD)	Text	0.70
1	20	TF	Standard	Source (MD)	Text	0.31
1	50	TF	Standard	Source (MD)	Text	0.18
1	10	TF	Standard	Source (LG)	Text	0.43
2	10	TF	Standard	Source (LG)	Text	0.59
3	10	TF	Standard	Source (LG)	Text	0.69
1	20	TF	Standard	Source (LG)	Text	0.30
1	50	TF	Standard	Source (LG)	Text	0.17
1	10	TF	Standard	Diff	Text	0.57
2	10	TF	Standard	Diff	Text	0.74
3	10	TF	Standard	Diff	Text	0.81
1	20	TF	Standard	Diff	Text	0.48
1	50	TF	Standard	Diff	Text	0.30
1	10	TF	Standard	Combined	Text	0.51
2	10	TF	Standard	Combined	Text	0.71
3	10	TF	Standard	Combined	Text	0.76
1	20	TF	Standard	Combined	Text	0.40
1	50	TF	Standard	Combined	Text	0.28
1	10	TF	Subword	Source (SM)	Text	0.59
2	10	TF	Subword	Source (SM)	Text	0.73
3	10	TF	Subword	Source (SM)	Text	0.79
1	20	TF	Subword	Source (SM)	Text	0.48
1	50	TF	Subword	Source (SM)	Text	0.36
1	10	TF	Subword	Source (MD)	Text	0.60

Continued on next page

Table B.2: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
2	10	TF	Subword	Source (MD)	Text	0.75
3	10	TF	Subword	Source (MD)	Text	0.80
1	20	TF	Subword	Source (MD)	Text	0.53
1	50	TF	Subword	Source (MD)	Text	0.39
1	10	TF	Subword	Source (LG)	Text	0.62
2	10	TF	Subword	Source (LG)	Text	0.75
3	10	TF	Subword	Source (LG)	Text	0.81
1	20	TF	Subword	Source (LG)	Text	0.53
1	50	TF	Subword	Source (LG)	Text	0.42
1	10	TF	Subword	Diff	Text	0.74
2	10	TF	Subword	Diff	Text	0.81
3	10	TF	Subword	Diff	Text	0.85
1	20	TF	Subword	Diff	Text	0.67
1	50	TF	Subword	Diff	Text	0.56
1	10	TF	Subword	Combined	Text	0.71
2	10	TF	Subword	Combined	Text	0.81
3	10	TF	Subword	Combined	Text	0.86
1	20	TF	Subword	Combined	Text	0.63
1	50	TF	Subword	Combined	Text	0.52
1	10	TF-IDF	Standard	Source (SM)	Text	0.50
2	10	TF-IDF	Standard	Source (SM)	Text	0.63
3	10	TF-IDF	Standard	Source (SM)	Text	0.70
1	20	TF-IDF	Standard	Source (SM)	Text	0.39
1	50	TF-IDF	Standard	Source (SM)	Text	0.28
1	10	TF-IDF	Standard	Source (MD)	Text	0.59
2	10	TF-IDF	Standard	Source (MD)	Text	0.72
3	10	TF-IDF	Standard	Source (MD)	Text	0.78
1	20	TF-IDF	Standard	Source (MD)	Text	0.47
1	50	TF-IDF	Standard	Source (MD)	Text	0.33

Continued on next page

Table B.2: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
1	10	TF-IDF	Standard	Source (LG)	Text	0.55
2	10	TF-IDF	Standard	Source (LG)	Text	0.71
3	10	TF-IDF	Standard	Source (LG)	Text	0.78
1	20	TF-IDF	Standard	Source (LG)	Text	0.42
1	50	TF-IDF	Standard	Source (LG)	Text	0.31
1	10	TF-IDF	Standard	Diff	Text	0.65
2	10	TF-IDF	Standard	Diff	Text	0.77
3	10	TF-IDF	Standard	Diff	Text	0.83
1	20	TF-IDF	Standard	Diff	Text	0.57
1	50	TF-IDF	Standard	Diff	Text	0.43
1	10	TF-IDF	Standard	Combined	Text	0.61
2	10	TF-IDF	Standard	Combined	Text	0.77
3	10	TF-IDF	Standard	Combined	Text	0.84
1	20	TF-IDF	Standard	Combined	Text	0.51
1	50	TF-IDF	Standard	Combined	Text	0.37
1	10	TF-IDF	Subword	Source (SM)	Text	0.65
2	10	TF-IDF	Subword	Source (SM)	Text	0.75
3	10	TF-IDF	Subword	Source (SM)	Text	0.79
1	20	TF-IDF	Subword	Source (SM)	Text	0.59
1	50	TF-IDF	Subword	Source (SM)	Text	0.49
1	10	TF-IDF	Subword	Source (MD)	Text	0.65
2	10	TF-IDF	Subword	Source (MD)	Text	0.76
3	10	TF-IDF	Subword	Source (MD)	Text	0.81
1	20	TF-IDF	Subword	Source (MD)	Text	0.58
1	50	TF-IDF	Subword	Source (MD)	Text	0.49
1	10	TF-IDF	Subword	Source (LG)	Text	0.66
2	10	TF-IDF	Subword	Source (LG)	Text	0.78
3	10	TF-IDF	Subword	Source (LG)	Text	0.83
1	20	TF-IDF	Subword	Source (LG)	Text	0.57

Continued on next page

Table B.2: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
1	50	TF-IDF	Subword	Source (LG)	Text	0.47
1	10	TF-IDF	Subword	Diff	Text	0.77
2	10	TF-IDF	Subword	Diff	Text	0.84
3	10	TF-IDF	Subword	Diff	Text	0.86
1	20	TF-IDF	Subword	Diff	Text	0.70
1	50	TF-IDF	Subword	Diff	Text	0.60
1	10	TF-IDF	Subword	Combined	Text	0.75
2	10	TF-IDF	Subword	Combined	Text	0.83
3	10	TF-IDF	Subword	Combined	Text	0.86
1	20	TF-IDF	Subword	Combined	Text	0.66
1	50	TF-IDF	Subword	Combined	Text	0.54
1	10	CodeBERT (Flattening)	-	Source (SM)	Code	0.09
2	10	CodeBERT (Flattening)	-	Source (SM)	Code	0.19
3	10	CodeBERT (Flattening)	-	Source (SM)	Code	0.29
1	20	CodeBERT (Flattening)	-	Source (SM)	Code	0.05
1	50	CodeBERT (Flattening)	-	Source (SM)	Code	0.02
1	10	CodeBERT (Flattening)	-	Source (MD)	Code	0.10
2	10	CodeBERT (Flattening)	-	Source (MD)	Code	0.20
3	10	CodeBERT (Flattening)	-	Source (MD)	Code	0.30
1	20	CodeBERT (Flattening)	-	Source (MD)	Code	0.05
1	50	CodeBERT (Flattening)	-	Source (MD)	Code	0.02
1	10	CodeBERT (Flattening)	-	Source (LG)	Code	0.10
2	10	CodeBERT (Flattening)	-	Source (LG)	Code	0.21
3	10	CodeBERT (Flattening)	-	Source (LG)	Code	0.30
1	20	CodeBERT (Flattening)	-	Source (LG)	Code	0.05
1	50	CodeBERT (Flattening)	-	Source (LG)	Code	0.02
1	10	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.09
2	10	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.20
3	10	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.30

Continued on next page

Table B.2: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
1	20	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.05
1	50	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.02
1	10	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.10
2	10	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.20
3	10	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.30
1	20	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.05
1	50	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.02
1	10	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.10
2	10	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.21
3	10	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.30
1	20	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.05
1	50	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.02

Table B.2: Result Runs for RxJava by ReactiveX

K	W	V	T	C	R	Mean Accuracy
1	10	TF	Standard	Source (SM)	Text	0.22
2	10	TF	Standard	Source (SM)	Text	0.39
3	10	TF	Standard	Source (SM)	Text	0.49
1	20	TF	Standard	Source (SM)	Text	0.11
1	50	TF	Standard	Source (SM)	Text	0.07
1	10	TF	Standard	Source (MD)	Text	0.23
2	10	TF	Standard	Source (MD)	Text	0.39
3	10	TF	Standard	Source (MD)	Text	0.51
1	20	TF	Standard	Source (MD)	Text	0.14
1	50	TF	Standard	Source (MD)	Text	0.09
1	10	TF	Standard	Source (LG)	Text	0.21
2	10	TF	Standard	Source (LG)	Text	0.38
3	10	TF	Standard	Source (LG)	Text	0.48
1	20	TF	Standard	Source (LG)	Text	0.11
1	50	TF	Standard	Source (LG)	Text	0.07
1	10	TF	Standard	Diff	Text	0.47
2	10	TF	Standard	Diff	Text	0.65
3	10	TF	Standard	Diff	Text	0.75
1	20	TF	Standard	Diff	Text	0.35
1	50	TF	Standard	Diff	Text	0.27
1	10	TF	Standard	Combined	Text	0.36
2	10	TF	Standard	Combined	Text	0.58
3	10	TF	Standard	Combined	Text	0.66
1	20	TF	Standard	Combined	Text	0.24
1	50	TF	Standard	Combined	Text	0.16
1	10	TF	Subword	Source (SM)	Text	0.48
2	10	TF	Subword	Source (SM)	Text	0.65
3	10	TF	Subword	Source (SM)	Text	0.73
1	20	TF	Subword	Source (SM)	Text	0.39
1	50	TF	Subword	Source (SM)	Text	0.28
1	10	TF	Subword	Source (MD)	Text	0.51

Continued on next page

Table B.3: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
2	10	TF	Subword	Source (MD)	Text	0.68
3	10	TF	Subword	Source (MD)	Text	0.75
1	20	TF	Subword	Source (MD)	Text	0.41
1	50	TF	Subword	Source (MD)	Text	0.29
1	10	TF	Subword	Source (LG)	Text	0.48
2	10	TF	Subword	Source (LG)	Text	0.69
3	10	TF	Subword	Source (LG)	Text	0.75
1	20	TF	Subword	Source (LG)	Text	0.35
1	50	TF	Subword	Source (LG)	Text	0.26
1	10	TF	Subword	Diff	Text	0.61
2	10	TF	Subword	Diff	Text	0.76
3	10	TF	Subword	Diff	Text	0.84
1	20	TF	Subword	Diff	Text	0.50
1	50	TF	Subword	Diff	Text	0.39
1	10	TF	Subword	Combined	Text	0.58
2	10	TF	Subword	Combined	Text	0.78
3	10	TF	Subword	Combined	Text	0.85
1	20	TF	Subword	Combined	Text	0.49
1	50	TF	Subword	Combined	Text	0.41
1	10	TF-IDF	Standard	Source (SM)	Text	0.53
2	10	TF-IDF	Standard	Source (SM)	Text	0.61
3	10	TF-IDF	Standard	Source (SM)	Text	0.68
1	20	TF-IDF	Standard	Source (SM)	Text	0.44
1	50	TF-IDF	Standard	Source (SM)	Text	0.37
1	10	TF-IDF	Standard	Source (MD)	Text	0.53
2	10	TF-IDF	Standard	Source (MD)	Text	0.65
3	10	TF-IDF	Standard	Source (MD)	Text	0.72
1	20	TF-IDF	Standard	Source (MD)	Text	0.47
1	50	TF-IDF	Standard	Source (MD)	Text	0.42

Continued on next page

Table B.3: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
1	10	TF-IDF	Standard	Source (LG)	Text	0.53
2	10	TF-IDF	Standard	Source (LG)	Text	0.65
3	10	TF-IDF	Standard	Source (LG)	Text	0.71
1	20	TF-IDF	Standard	Source (LG)	Text	0.46
1	50	TF-IDF	Standard	Source (LG)	Text	0.40
1	10	TF-IDF	Standard	Diff	Text	0.65
2	10	TF-IDF	Standard	Diff	Text	0.76
3	10	TF-IDF	Standard	Diff	Text	0.79
1	20	TF-IDF	Standard	Diff	Text	0.58
1	50	TF-IDF	Standard	Diff	Text	0.49
1	10	TF-IDF	Standard	Combined	Text	0.66
2	10	TF-IDF	Standard	Combined	Text	0.77
3	10	TF-IDF	Standard	Combined	Text	0.81
1	20	TF-IDF	Standard	Combined	Text	0.58
1	50	TF-IDF	Standard	Combined	Text	0.53
1	10	TF-IDF	Subword	Source (SM)	Text	0.62
2	10	TF-IDF	Subword	Source (SM)	Text	0.70
3	10	TF-IDF	Subword	Source (SM)	Text	0.76
1	20	TF-IDF	Subword	Source (SM)	Text	0.55
1	50	TF-IDF	Subword	Source (SM)	Text	0.47
1	10	TF-IDF	Subword	Source (MD)	Text	0.63
2	10	TF-IDF	Subword	Source (MD)	Text	0.71
3	10	TF-IDF	Subword	Source (MD)	Text	0.77
1	20	TF-IDF	Subword	Source (MD)	Text	0.57
1	50	TF-IDF	Subword	Source (MD)	Text	0.49
1	10	TF-IDF	Subword	Source (LG)	Text	0.64
2	10	TF-IDF	Subword	Source (LG)	Text	0.74
3	10	TF-IDF	Subword	Source (LG)	Text	0.79
1	20	TF-IDF	Subword	Source (LG)	Text	0.56

Continued on next page

Table B.3: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
1	50	TF-IDF	Subword	Source (LG)	Text	0.46
1	10	TF-IDF	Subword	Diff	Text	0.70
2	10	TF-IDF	Subword	Diff	Text	0.82
3	10	TF-IDF	Subword	Diff	Text	0.87
1	20	TF-IDF	Subword	Diff	Text	0.62
1	50	TF-IDF	Subword	Diff	Text	0.55
1	10	TF-IDF	Subword	Combined	Text	0.70
2	10	TF-IDF	Subword	Combined	Text	0.83
3	10	TF-IDF	Subword	Combined	Text	0.89
1	20	TF-IDF	Subword	Combined	Text	0.62
1	50	TF-IDF	Subword	Combined	Text	0.52
1	10	CodeBERT (Flattening)	-	Source (SM)	Code	0.11
2	10	CodeBERT (Flattening)	-	Source (SM)	Code	0.24
3	10	CodeBERT (Flattening)	-	Source (SM)	Code	0.35
1	20	CodeBERT (Flattening)	-	Source (SM)	Code	0.06
1	50	CodeBERT (Flattening)	-	Source (SM)	Code	0.02
1	10	CodeBERT (Flattening)	-	Source (MD)	Code	0.11
2	10	CodeBERT (Flattening)	-	Source (MD)	Code	0.24
3	10	CodeBERT (Flattening)	-	Source (MD)	Code	0.35
1	20	CodeBERT (Flattening)	-	Source (MD)	Code	0.06
1	50	CodeBERT (Flattening)	-	Source (MD)	Code	0.02
1	10	CodeBERT (Flattening)	-	Source (LG)	Code	0.10
2	10	CodeBERT (Flattening)	-	Source (LG)	Code	0.19
3	10	CodeBERT (Flattening)	-	Source (LG)	Code	0.29
1	20	CodeBERT (Flattening)	-	Source (LG)	Code	0.07
1	50	CodeBERT (Flattening)	-	Source (LG)	Code	0.04
1	10	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.13
2	10	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.25
3	10	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.35

Continued on next page

Table B.3: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
1	20	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.07
1	50	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.04
1	10	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.13
2	10	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.25
3	10	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.35
1	20	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.07
1	50	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.04
1	10	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.11
2	10	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.19
3	10	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.29
1	20	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.07
1	50	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.05

Table B.3: Result Runs for dubbo by apache

K	W	V	T	C	R	Mean Accuracy
1	10	TF	Standard	Source (SM)	Text	0.29
2	10	TF	Standard	Source (SM)	Text	0.40
3	10	TF	Standard	Source (SM)	Text	0.52
1	20	TF	Standard	Source (SM)	Text	0.24
1	50	TF	Standard	Source (SM)	Text	0.20
1	10	TF	Standard	Source (MD)	Text	0.31
2	10	TF	Standard	Source (MD)	Text	0.42
3	10	TF	Standard	Source (MD)	Text	0.52
1	20	TF	Standard	Source (MD)	Text	0.25
1	50	TF	Standard	Source (MD)	Text	0.21
1	10	TF	Standard	Source (LG)	Text	0.33
2	10	TF	Standard	Source (LG)	Text	0.47
3	10	TF	Standard	Source (LG)	Text	0.56
1	20	TF	Standard	Source (LG)	Text	0.27
1	50	TF	Standard	Source (LG)	Text	0.21
1	10	TF	Standard	Diff	Text	0.41
2	10	TF	Standard	Diff	Text	0.53
3	10	TF	Standard	Diff	Text	0.63
1	20	TF	Standard	Diff	Text	0.39
1	50	TF	Standard	Diff	Text	0.34
1	10	TF	Standard	Combined	Text	0.36
2	10	TF	Standard	Combined	Text	0.53
3	10	TF	Standard	Combined	Text	0.61
1	20	TF	Standard	Combined	Text	0.28
1	50	TF	Standard	Combined	Text	0.22
1	10	TF	Subword	Source (SM)	Text	0.42
2	10	TF	Subword	Source (SM)	Text	0.49
3	10	TF	Subword	Source (SM)	Text	0.60
1	20	TF	Subword	Source (SM)	Text	0.37
1	50	TF	Subword	Source (SM)	Text	0.29
1	10	TF	Subword	Source (MD)	Text	0.43

Continued on next page

Table B.4: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
2	10	TF	Subword	Source (MD)	Text	0.52
3	10	TF	Subword	Source (MD)	Text	0.60
1	20	TF	Subword	Source (MD)	Text	0.37
1	50	TF	Subword	Source (MD)	Text	0.29
1	10	TF	Subword	Source (LG)	Text	0.47
2	10	TF	Subword	Source (LG)	Text	0.55
3	10	TF	Subword	Source (LG)	Text	0.63
1	20	TF	Subword	Source (LG)	Text	0.43
1	50	TF	Subword	Source (LG)	Text	0.35
1	10	TF	Subword	Diff	Text	0.48
2	10	TF	Subword	Diff	Text	0.59
3	10	TF	Subword	Diff	Text	0.67
1	20	TF	Subword	Diff	Text	0.45
1	50	TF	Subword	Diff	Text	0.37
1	10	TF	Subword	Combined	Text	0.47
2	10	TF	Subword	Combined	Text	0.58
3	10	TF	Subword	Combined	Text	0.65
1	20	TF	Subword	Combined	Text	0.43
1	50	TF	Subword	Combined	Text	0.37
1	10	TF-IDF	Standard	Source (SM)	Text	0.41
2	10	TF-IDF	Standard	Source (SM)	Text	0.51
3	10	TF-IDF	Standard	Source (SM)	Text	0.56
1	20	TF-IDF	Standard	Source (SM)	Text	0.40
1	50	TF-IDF	Standard	Source (SM)	Text	0.39
1	10	TF-IDF	Standard	Source (MD)	Text	0.42
2	10	TF-IDF	Standard	Source (MD)	Text	0.50
3	10	TF-IDF	Standard	Source (MD)	Text	0.55
1	20	TF-IDF	Standard	Source (MD)	Text	0.42
1	50	TF-IDF	Standard	Source (MD)	Text	0.41

Continued on next page

Table B.4: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
1	10	TF-IDF	Standard	Source (LG)	Text	0.47
2	10	TF-IDF	Standard	Source (LG)	Text	0.53
3	10	TF-IDF	Standard	Source (LG)	Text	0.58
1	20	TF-IDF	Standard	Source (LG)	Text	0.47
1	50	TF-IDF	Standard	Source (LG)	Text	0.45
1	10	TF-IDF	Standard	Diff	Text	0.50
2	10	TF-IDF	Standard	Diff	Text	0.62
3	10	TF-IDF	Standard	Diff	Text	0.68
1	20	TF-IDF	Standard	Diff	Text	0.46
1	50	TF-IDF	Standard	Diff	Text	0.43
1	10	TF-IDF	Standard	Combined	Text	0.50
2	10	TF-IDF	Standard	Combined	Text	0.57
3	10	TF-IDF	Standard	Combined	Text	0.64
1	20	TF-IDF	Standard	Combined	Text	0.49
1	50	TF-IDF	Standard	Combined	Text	0.47
1	10	TF-IDF	Subword	Source (SM)	Text	0.52
2	10	TF-IDF	Subword	Source (SM)	Text	0.60
3	10	TF-IDF	Subword	Source (SM)	Text	0.65
1	20	TF-IDF	Subword	Source (SM)	Text	0.49
1	50	TF-IDF	Subword	Source (SM)	Text	0.43
1	10	TF-IDF	Subword	Source (MD)	Text	0.52
2	10	TF-IDF	Subword	Source (MD)	Text	0.61
3	10	TF-IDF	Subword	Source (MD)	Text	0.66
1	20	TF-IDF	Subword	Source (MD)	Text	0.49
1	50	TF-IDF	Subword	Source (MD)	Text	0.43
1	10	TF-IDF	Subword	Source (LG)	Text	0.52
2	10	TF-IDF	Subword	Source (LG)	Text	0.62
3	10	TF-IDF	Subword	Source (LG)	Text	0.67
1	20	TF-IDF	Subword	Source (LG)	Text	0.46

Continued on next page

Table B.4: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
1	50	TF-IDF	Subword	Source (LG)	Text	0.40
1	10	TF-IDF	Subword	Diff	Text	0.53
2	10	TF-IDF	Subword	Diff	Text	0.65
3	10	TF-IDF	Subword	Diff	Text	0.71
1	20	TF-IDF	Subword	Diff	Text	0.49
1	50	TF-IDF	Subword	Diff	Text	0.47
1	10	TF-IDF	Subword	Combined	Text	0.54
2	10	TF-IDF	Subword	Combined	Text	0.64
3	10	TF-IDF	Subword	Combined	Text	0.68
1	20	TF-IDF	Subword	Combined	Text	0.51
1	50	TF-IDF	Subword	Combined	Text	0.45
1	10	CodeBERT (Flattening)	-	Source (SM)	Code	0.12
2	10	CodeBERT (Flattening)	-	Source (SM)	Code	0.20
3	10	CodeBERT (Flattening)	-	Source (SM)	Code	0.30
1	20	CodeBERT (Flattening)	-	Source (SM)	Code	0.06
1	50	CodeBERT (Flattening)	-	Source (SM)	Code	0.02
1	10	CodeBERT (Flattening)	-	Source (MD)	Code	0.12
2	10	CodeBERT (Flattening)	-	Source (MD)	Code	0.20
3	10	CodeBERT (Flattening)	-	Source (MD)	Code	0.30
1	20	CodeBERT (Flattening)	-	Source (MD)	Code	0.06
1	50	CodeBERT (Flattening)	-	Source (MD)	Code	0.02
1	10	CodeBERT (Flattening)	-	Source (LG)	Code	0.09
2	10	CodeBERT (Flattening)	-	Source (LG)	Code	0.19
3	10	CodeBERT (Flattening)	-	Source (LG)	Code	0.30
1	20	CodeBERT (Flattening)	-	Source (LG)	Code	0.05
1	50	CodeBERT (Flattening)	-	Source (LG)	Code	0.02
1	10	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.11
2	10	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.20
3	10	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.30

Continued on next page

Table B.4: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
1	20	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.05
1	50	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.02
1	10	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.11
2	10	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.20
3	10	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.30
1	20	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.05
1	50	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.02
1	10	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.10
2	10	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.19
3	10	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.30
1	20	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.05
1	50	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.02

Table B.4: Result Runs for tutorials by eugenp

K	W	V	T	C	R	Mean Accuracy
1	10	TF	Standard	Source (SM)	Text	0.22
2	10	TF	Standard	Source (SM)	Text	0.33
3	10	TF	Standard	Source (SM)	Text	0.44
1	20	TF	Standard	Source (SM)	Text	0.18
1	50	TF	Standard	Source (SM)	Text	0.13
1	10	TF	Standard	Source (MD)	Text	0.33
2	10	TF	Standard	Source (MD)	Text	0.44
3	10	TF	Standard	Source (MD)	Text	0.52
1	20	TF	Standard	Source (MD)	Text	0.24
1	50	TF	Standard	Source (MD)	Text	0.14
1	10	TF	Standard	Source (LG)	Text	0.32
2	10	TF	Standard	Source (LG)	Text	0.43
3	10	TF	Standard	Source (LG)	Text	0.50
1	20	TF	Standard	Source (LG)	Text	0.20
1	50	TF	Standard	Source (LG)	Text	0.13
1	10	TF	Standard	Diff	Text	0.58
2	10	TF	Standard	Diff	Text	0.73
3	10	TF	Standard	Diff	Text	0.79
1	20	TF	Standard	Diff	Text	0.51
1	50	TF	Standard	Diff	Text	0.41
1	10	TF	Standard	Combined	Text	0.45
2	10	TF	Standard	Combined	Text	0.60
3	10	TF	Standard	Combined	Text	0.70
1	20	TF	Standard	Combined	Text	0.36
1	50	TF	Standard	Combined	Text	0.27
1	10	TF	Subword	Source (SM)	Text	0.38
2	10	TF	Subword	Source (SM)	Text	0.52
3	10	TF	Subword	Source (SM)	Text	0.63
1	20	TF	Subword	Source (SM)	Text	0.31
1	50	TF	Subword	Source (SM)	Text	0.19
1	10	TF	Subword	Source (MD)	Text	0.39

Continued on next page

Table B.5: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
2	10	TF	Subword	Source (MD)	Text	0.55
3	10	TF	Subword	Source (MD)	Text	0.66
1	20	TF	Subword	Source (MD)	Text	0.30
1	50	TF	Subword	Source (MD)	Text	0.19
1	10	TF	Subword	Source (LG)	Text	0.36
2	10	TF	Subword	Source (LG)	Text	0.54
3	10	TF	Subword	Source (LG)	Text	0.66
1	20	TF	Subword	Source (LG)	Text	0.28
1	50	TF	Subword	Source (LG)	Text	0.18
1	10	TF	Subword	Diff	Text	0.63
2	10	TF	Subword	Diff	Text	0.76
3	10	TF	Subword	Diff	Text	0.82
1	20	TF	Subword	Diff	Text	0.54
1	50	TF	Subword	Diff	Text	0.42
1	10	TF	Subword	Combined	Text	0.52
2	10	TF	Subword	Combined	Text	0.66
3	10	TF	Subword	Combined	Text	0.76
1	20	TF	Subword	Combined	Text	0.44
1	50	TF	Subword	Combined	Text	0.35
1	10	TF-IDF	Standard	Source (SM)	Text	0.30
2	10	TF-IDF	Standard	Source (SM)	Text	0.40
3	10	TF-IDF	Standard	Source (SM)	Text	0.51
1	20	TF-IDF	Standard	Source (SM)	Text	0.24
1	50	TF-IDF	Standard	Source (SM)	Text	0.16
1	10	TF-IDF	Standard	Source (MD)	Text	0.48
2	10	TF-IDF	Standard	Source (MD)	Text	0.58
3	10	TF-IDF	Standard	Source (MD)	Text	0.65
1	20	TF-IDF	Standard	Source (MD)	Text	0.40
1	50	TF-IDF	Standard	Source (MD)	Text	0.30

Continued on next page

Table B.5: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
1	10	TF-IDF	Standard	Source (LG)	Text	0.49
2	10	TF-IDF	Standard	Source (LG)	Text	0.60
3	10	TF-IDF	Standard	Source (LG)	Text	0.66
1	20	TF-IDF	Standard	Source (LG)	Text	0.41
1	50	TF-IDF	Standard	Source (LG)	Text	0.28
1	10	TF-IDF	Standard	Diff	Text	0.66
2	10	TF-IDF	Standard	Diff	Text	0.78
3	10	TF-IDF	Standard	Diff	Text	0.81
1	20	TF-IDF	Standard	Diff	Text	0.58
1	50	TF-IDF	Standard	Diff	Text	0.51
1	10	TF-IDF	Standard	Combined	Text	0.63
2	10	TF-IDF	Standard	Combined	Text	0.74
3	10	TF-IDF	Standard	Combined	Text	0.78
1	20	TF-IDF	Standard	Combined	Text	0.57
1	50	TF-IDF	Standard	Combined	Text	0.50
1	10	TF-IDF	Subword	Source (SM)	Text	0.51
2	10	TF-IDF	Subword	Source (SM)	Text	0.64
3	10	TF-IDF	Subword	Source (SM)	Text	0.72
1	20	TF-IDF	Subword	Source (SM)	Text	0.43
1	50	TF-IDF	Subword	Source (SM)	Text	0.30
1	10	TF-IDF	Subword	Source (MD)	Text	0.58
2	10	TF-IDF	Subword	Source (MD)	Text	0.70
3	10	TF-IDF	Subword	Source (MD)	Text	0.76
1	20	TF-IDF	Subword	Source (MD)	Text	0.46
1	50	TF-IDF	Subword	Source (MD)	Text	0.32
1	10	TF-IDF	Subword	Source (LG)	Text	0.59
2	10	TF-IDF	Subword	Source (LG)	Text	0.71
3	10	TF-IDF	Subword	Source (LG)	Text	0.76
1	20	TF-IDF	Subword	Source (LG)	Text	0.46

Continued on next page

Table B.5: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
1	50	TF-IDF	Subword	Source (LG)	Text	0.34
1	10	TF-IDF	Subword	Diff	Text	0.73
2	10	TF-IDF	Subword	Diff	Text	0.82
3	10	TF-IDF	Subword	Diff	Text	0.86
1	20	TF-IDF	Subword	Diff	Text	0.67
1	50	TF-IDF	Subword	Diff	Text	0.56
1	10	TF-IDF	Subword	Combined	Text	0.71
2	10	TF-IDF	Subword	Combined	Text	0.80
3	10	TF-IDF	Subword	Combined	Text	0.84
1	20	TF-IDF	Subword	Combined	Text	0.63
1	50	TF-IDF	Subword	Combined	Text	0.51
1	10	CodeBERT (Flattening)	-	Source (SM)	Code	0.10
2	10	CodeBERT (Flattening)	-	Source (SM)	Code	0.19
3	10	CodeBERT (Flattening)	-	Source (SM)	Code	0.29
1	20	CodeBERT (Flattening)	-	Source (SM)	Code	0.06
1	50	CodeBERT (Flattening)	-	Source (SM)	Code	0.02
1	10	CodeBERT (Flattening)	-	Source (MD)	Code	0.10
2	10	CodeBERT (Flattening)	-	Source (MD)	Code	0.19
3	10	CodeBERT (Flattening)	-	Source (MD)	Code	0.29
1	20	CodeBERT (Flattening)	-	Source (MD)	Code	0.06
1	50	CodeBERT (Flattening)	-	Source (MD)	Code	0.02
1	10	CodeBERT (Flattening)	-	Source (LG)	Code	0.10
2	10	CodeBERT (Flattening)	-	Source (LG)	Code	0.19
3	10	CodeBERT (Flattening)	-	Source (LG)	Code	0.29
1	20	CodeBERT (Flattening)	-	Source (LG)	Code	0.05
1	50	CodeBERT (Flattening)	-	Source (LG)	Code	0.02
1	10	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.10
2	10	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.19
3	10	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.30

Continued on next page

Table B.5: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
1	20	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.06
1	50	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.02
1	10	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.10
2	10	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.19
3	10	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.30
1	20	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.06
1	50	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.02
1	10	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.10
2	10	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.19
3	10	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.29
1	20	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.05
1	50	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.02

Table B.5: Result Runs for lottie-android by airbnb

K	W	V	T	C	R	Mean Accuracy
1	10	TF	Standard	Source (SM)	Text	0.35
2	10	TF	Standard	Source (SM)	Text	0.46
3	10	TF	Standard	Source (SM)	Text	0.54
1	20	TF	Standard	Source (SM)	Text	0.29
1	50	TF	Standard	Source (SM)	Text	0.22
1	10	TF	Standard	Source (MD)	Text	0.42
2	10	TF	Standard	Source (MD)	Text	0.54
3	10	TF	Standard	Source (MD)	Text	0.65
1	20	TF	Standard	Source (MD)	Text	0.38
1	50	TF	Standard	Source (MD)	Text	0.30
1	10	TF	Standard	Source (LG)	Text	0.44
2	10	TF	Standard	Source (LG)	Text	0.55
3	10	TF	Standard	Source (LG)	Text	0.65
1	20	TF	Standard	Source (LG)	Text	0.38
1	50	TF	Standard	Source (LG)	Text	0.28
1	10	TF	Standard	Diff	Text	0.51
2	10	TF	Standard	Diff	Text	0.64
3	10	TF	Standard	Diff	Text	0.71
1	20	TF	Standard	Diff	Text	0.45
1	50	TF	Standard	Diff	Text	0.37
1	10	TF	Standard	Combined	Text	0.49
2	10	TF	Standard	Combined	Text	0.64
3	10	TF	Standard	Combined	Text	0.74
1	20	TF	Standard	Combined	Text	0.42
1	50	TF	Standard	Combined	Text	0.37
1	10	TF	Subword	Source (SM)	Text	0.48
2	10	TF	Subword	Source (SM)	Text	0.63
3	10	TF	Subword	Source (SM)	Text	0.71
1	20	TF	Subword	Source (SM)	Text	0.39
1	50	TF	Subword	Source (SM)	Text	0.33
1	10	TF	Subword	Source (MD)	Text	0.52

Continued on next page

Table B.6: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
2	10	TF	Subword	Source (MD)	Text	0.67
3	10	TF	Subword	Source (MD)	Text	0.75
1	20	TF	Subword	Source (MD)	Text	0.43
1	50	TF	Subword	Source (MD)	Text	0.36
1	10	TF	Subword	Source (LG)	Text	0.55
2	10	TF	Subword	Source (LG)	Text	0.70
3	10	TF	Subword	Source (LG)	Text	0.77
1	20	TF	Subword	Source (LG)	Text	0.46
1	50	TF	Subword	Source (LG)	Text	0.36
1	10	TF	Subword	Diff	Text	0.58
2	10	TF	Subword	Diff	Text	0.66
3	10	TF	Subword	Diff	Text	0.73
1	20	TF	Subword	Diff	Text	0.52
1	50	TF	Subword	Diff	Text	0.44
1	10	TF	Subword	Combined	Text	0.57
2	10	TF	Subword	Combined	Text	0.71
3	10	TF	Subword	Combined	Text	0.78
1	20	TF	Subword	Combined	Text	0.49
1	50	TF	Subword	Combined	Text	0.42
1	10	TF-IDF	Standard	Source (SM)	Text	0.44
2	10	TF-IDF	Standard	Source (SM)	Text	0.56
3	10	TF-IDF	Standard	Source (SM)	Text	0.62
1	20	TF-IDF	Standard	Source (SM)	Text	0.38
1	50	TF-IDF	Standard	Source (SM)	Text	0.32
1	10	TF-IDF	Standard	Source (MD)	Text	0.59
2	10	TF-IDF	Standard	Source (MD)	Text	0.64
3	10	TF-IDF	Standard	Source (MD)	Text	0.68
1	20	TF-IDF	Standard	Source (MD)	Text	0.54
1	50	TF-IDF	Standard	Source (MD)	Text	0.45

Continued on next page

Table B.6: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
1	10	TF-IDF	Standard	Source (LG)	Text	0.58
2	10	TF-IDF	Standard	Source (LG)	Text	0.68
3	10	TF-IDF	Standard	Source (LG)	Text	0.72
1	20	TF-IDF	Standard	Source (LG)	Text	0.54
1	50	TF-IDF	Standard	Source (LG)	Text	0.44
1	10	TF-IDF	Standard	Diff	Text	0.60
2	10	TF-IDF	Standard	Diff	Text	0.68
3	10	TF-IDF	Standard	Diff	Text	0.75
1	20	TF-IDF	Standard	Diff	Text	0.58
1	50	TF-IDF	Standard	Diff	Text	0.49
1	10	TF-IDF	Standard	Combined	Text	0.63
2	10	TF-IDF	Standard	Combined	Text	0.76
3	10	TF-IDF	Standard	Combined	Text	0.81
1	20	TF-IDF	Standard	Combined	Text	0.58
1	50	TF-IDF	Standard	Combined	Text	0.50
1	10	TF-IDF	Subword	Source (SM)	Text	0.65
2	10	TF-IDF	Subword	Source (SM)	Text	0.76
3	10	TF-IDF	Subword	Source (SM)	Text	0.79
1	20	TF-IDF	Subword	Source (SM)	Text	0.54
1	50	TF-IDF	Subword	Source (SM)	Text	0.40
1	10	TF-IDF	Subword	Source (MD)	Text	0.69
2	10	TF-IDF	Subword	Source (MD)	Text	0.80
3	10	TF-IDF	Subword	Source (MD)	Text	0.83
1	20	TF-IDF	Subword	Source (MD)	Text	0.58
1	50	TF-IDF	Subword	Source (MD)	Text	0.46
1	10	TF-IDF	Subword	Source (LG)	Text	0.72
2	10	TF-IDF	Subword	Source (LG)	Text	0.82
3	10	TF-IDF	Subword	Source (LG)	Text	0.86
1	20	TF-IDF	Subword	Source (LG)	Text	0.62

Continued on next page

Table B.6: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
1	50	TF-IDF	Subword	Source (LG)	Text	0.45
1	10	TF-IDF	Subword	Diff	Text	0.63
2	10	TF-IDF	Subword	Diff	Text	0.70
3	10	TF-IDF	Subword	Diff	Text	0.74
1	20	TF-IDF	Subword	Diff	Text	0.56
1	50	TF-IDF	Subword	Diff	Text	0.49
1	10	TF-IDF	Subword	Combined	Text	0.67
2	10	TF-IDF	Subword	Combined	Text	0.80
3	10	TF-IDF	Subword	Combined	Text	0.82
1	20	TF-IDF	Subword	Combined	Text	0.58
1	50	TF-IDF	Subword	Combined	Text	0.48
1	10	CodeBERT (Flattening)	-	Source (SM)	Code	0.12
2	10	CodeBERT (Flattening)	-	Source (SM)	Code	0.22
3	10	CodeBERT (Flattening)	-	Source (SM)	Code	0.32
1	20	CodeBERT (Flattening)	-	Source (SM)	Code	0.07
1	50	CodeBERT (Flattening)	-	Source (SM)	Code	0.05
1	10	CodeBERT (Flattening)	-	Source (MD)	Code	0.11
2	10	CodeBERT (Flattening)	-	Source (MD)	Code	0.20
3	10	CodeBERT (Flattening)	-	Source (MD)	Code	0.31
1	20	CodeBERT (Flattening)	-	Source (MD)	Code	0.07
1	50	CodeBERT (Flattening)	-	Source (MD)	Code	0.05
1	10	CodeBERT (Flattening)	-	Source (LG)	Code	0.08
2	10	CodeBERT (Flattening)	-	Source (LG)	Code	0.21
3	10	CodeBERT (Flattening)	-	Source (LG)	Code	0.31
1	20	CodeBERT (Flattening)	-	Source (LG)	Code	0.04
1	50	CodeBERT (Flattening)	-	Source (LG)	Code	0.02
1	10	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.12
2	10	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.21
3	10	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.31

Continued on next page

Table B.6: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
1	20	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.08
1	50	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.06
1	10	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.12
2	10	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.20
3	10	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.31
1	20	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.08
1	50	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.06
1	10	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.09
2	10	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.19
3	10	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.29
1	20	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.04
1	50	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.02

Table B.6: Result Runs for glide by bumptech

K	W	V	T	C	R	Mean Accuracy
1	10	TF	Standard	Source (SM)	Text	0.40
2	10	TF	Standard	Source (SM)	Text	0.58
3	10	TF	Standard	Source (SM)	Text	0.77
1	20	TF	Standard	Source (SM)	Text	0.29
1	50	TF	Standard	Source (SM)	Text	0.15
1	10	TF	Standard	Source (MD)	Text	0.51
2	10	TF	Standard	Source (MD)	Text	0.69
3	10	TF	Standard	Source (MD)	Text	0.80
1	20	TF	Standard	Source (MD)	Text	0.37
1	50	TF	Standard	Source (MD)	Text	0.20
1	10	TF	Standard	Source (LG)	Text	0.50
2	10	TF	Standard	Source (LG)	Text	0.68
3	10	TF	Standard	Source (LG)	Text	0.79
1	20	TF	Standard	Source (LG)	Text	0.36
1	50	TF	Standard	Source (LG)	Text	0.17
1	10	TF	Standard	Diff	Text	0.75
2	10	TF	Standard	Diff	Text	0.89
3	10	TF	Standard	Diff	Text	0.93
1	20	TF	Standard	Diff	Text	0.63
1	50	TF	Standard	Diff	Text	0.53
1	10	TF	Standard	Combined	Text	0.65
2	10	TF	Standard	Combined	Text	0.82
3	10	TF	Standard	Combined	Text	0.91
1	20	TF	Standard	Combined	Text	0.54
1	50	TF	Standard	Combined	Text	0.43
1	10	TF	Subword	Source (SM)	Text	0.67
2	10	TF	Subword	Source (SM)	Text	0.87
3	10	TF	Subword	Source (SM)	Text	0.91
1	20	TF	Subword	Source (SM)	Text	0.54
1	50	TF	Subword	Source (SM)	Text	0.43
1	10	TF	Subword	Source (MD)	Text	0.70

Continued on next page

Table B.7: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
2	10	TF	Subword	Source (MD)	Text	0.85
3	10	TF	Subword	Source (MD)	Text	0.91
1	20	TF	Subword	Source (MD)	Text	0.59
1	50	TF	Subword	Source (MD)	Text	0.46
1	10	TF	Subword	Source (LG)	Text	0.71
2	10	TF	Subword	Source (LG)	Text	0.87
3	10	TF	Subword	Source (LG)	Text	0.92
1	20	TF	Subword	Source (LG)	Text	0.59
1	50	TF	Subword	Source (LG)	Text	0.46
1	10	TF	Subword	Diff	Text	0.85
2	10	TF	Subword	Diff	Text	0.96
3	10	TF	Subword	Diff	Text	0.99
1	20	TF	Subword	Diff	Text	0.75
1	50	TF	Subword	Diff	Text	0.67
1	10	TF	Subword	Combined	Text	0.84
2	10	TF	Subword	Combined	Text	0.95
3	10	TF	Subword	Combined	Text	1.00
1	20	TF	Subword	Combined	Text	0.75
1	50	TF	Subword	Combined	Text	0.65
1	10	TF-IDF	Standard	Source (SM)	Text	0.65
2	10	TF-IDF	Standard	Source (SM)	Text	0.76
3	10	TF-IDF	Standard	Source (SM)	Text	0.83
1	20	TF-IDF	Standard	Source (SM)	Text	0.56
1	50	TF-IDF	Standard	Source (SM)	Text	0.45
1	10	TF-IDF	Standard	Source (MD)	Text	0.76
2	10	TF-IDF	Standard	Source (MD)	Text	0.84
3	10	TF-IDF	Standard	Source (MD)	Text	0.89
1	20	TF-IDF	Standard	Source (MD)	Text	0.69
1	50	TF-IDF	Standard	Source (MD)	Text	0.59

Continued on next page

Table B.7: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
1	10	TF-IDF	Standard	Source (LG)	Text	0.72
2	10	TF-IDF	Standard	Source (LG)	Text	0.82
3	10	TF-IDF	Standard	Source (LG)	Text	0.88
1	20	TF-IDF	Standard	Source (LG)	Text	0.64
1	50	TF-IDF	Standard	Source (LG)	Text	0.54
1	10	TF-IDF	Standard	Diff	Text	0.89
2	10	TF-IDF	Standard	Diff	Text	0.94
3	10	TF-IDF	Standard	Diff	Text	0.95
1	20	TF-IDF	Standard	Diff	Text	0.82
1	50	TF-IDF	Standard	Diff	Text	0.74
1	10	TF-IDF	Standard	Combined	Text	0.89
2	10	TF-IDF	Standard	Combined	Text	0.94
3	10	TF-IDF	Standard	Combined	Text	0.98
1	20	TF-IDF	Standard	Combined	Text	0.81
1	50	TF-IDF	Standard	Combined	Text	0.74
1	10	TF-IDF	Subword	Source (SM)	Text	0.75
2	10	TF-IDF	Subword	Source (SM)	Text	0.90
3	10	TF-IDF	Subword	Source (SM)	Text	0.94
1	20	TF-IDF	Subword	Source (SM)	Text	0.63
1	50	TF-IDF	Subword	Source (SM)	Text	0.49
1	10	TF-IDF	Subword	Source (MD)	Text	0.76
2	10	TF-IDF	Subword	Source (MD)	Text	0.89
3	10	TF-IDF	Subword	Source (MD)	Text	0.94
1	20	TF-IDF	Subword	Source (MD)	Text	0.66
1	50	TF-IDF	Subword	Source (MD)	Text	0.57
1	10	TF-IDF	Subword	Source (LG)	Text	0.78
2	10	TF-IDF	Subword	Source (LG)	Text	0.90
3	10	TF-IDF	Subword	Source (LG)	Text	0.94
1	20	TF-IDF	Subword	Source (LG)	Text	0.67

Continued on next page

Table B.7: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
1	50	TF-IDF	Subword	Source (LG)	Text	0.56
1	10	TF-IDF	Subword	Diff	Text	0.91
2	10	TF-IDF	Subword	Diff	Text	0.98
3	10	TF-IDF	Subword	Diff	Text	0.99
1	20	TF-IDF	Subword	Diff	Text	0.81
1	50	TF-IDF	Subword	Diff	Text	0.72
1	10	TF-IDF	Subword	Combined	Text	0.88
2	10	TF-IDF	Subword	Combined	Text	0.98
3	10	TF-IDF	Subword	Combined	Text	0.99
1	20	TF-IDF	Subword	Combined	Text	0.81
1	50	TF-IDF	Subword	Combined	Text	0.70
1	10	CodeBERT (Flattening)	-	Source (SM)	Code	0.12
2	10	CodeBERT (Flattening)	-	Source (SM)	Code	0.23
3	10	CodeBERT (Flattening)	-	Source (SM)	Code	0.34
1	20	CodeBERT (Flattening)	-	Source (SM)	Code	0.06
1	50	CodeBERT (Flattening)	-	Source (SM)	Code	0.02
1	10	CodeBERT (Flattening)	-	Source (MD)	Code	0.12
2	10	CodeBERT (Flattening)	-	Source (MD)	Code	0.23
3	10	CodeBERT (Flattening)	-	Source (MD)	Code	0.34
1	20	CodeBERT (Flattening)	-	Source (MD)	Code	0.06
1	50	CodeBERT (Flattening)	-	Source (MD)	Code	0.02
1	10	CodeBERT (Flattening)	-	Source (LG)	Code	0.12
2	10	CodeBERT (Flattening)	-	Source (LG)	Code	0.25
3	10	CodeBERT (Flattening)	-	Source (LG)	Code	0.35
1	20	CodeBERT (Flattening)	-	Source (LG)	Code	0.06
1	50	CodeBERT (Flattening)	-	Source (LG)	Code	0.02
1	10	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.12
2	10	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.24
3	10	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.35

Continued on next page

Table B.7: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
1	20	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.06
1	50	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.03
1	10	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.12
2	10	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.24
3	10	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.35
1	20	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.06
1	50	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.03
1	10	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.12
2	10	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.25
3	10	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.35
1	20	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.05
1	50	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.01

Table B.7: Result Runs for netty by netty

K	W	V	T	C	R	Mean Accuracy
1	10	TF	Standard	Source (SM)	Text	0.23
2	10	TF	Standard	Source (SM)	Text	0.37
3	10	TF	Standard	Source (SM)	Text	0.44
1	20	TF	Standard	Source (SM)	Text	0.14
1	50	TF	Standard	Source (SM)	Text	0.07
1	10	TF	Standard	Source (MD)	Text	0.26
2	10	TF	Standard	Source (MD)	Text	0.37
3	10	TF	Standard	Source (MD)	Text	0.45
1	20	TF	Standard	Source (MD)	Text	0.15
1	50	TF	Standard	Source (MD)	Text	0.07
1	10	TF	Standard	Source (LG)	Text	0.25
2	10	TF	Standard	Source (LG)	Text	0.40
3	10	TF	Standard	Source (LG)	Text	0.49
1	20	TF	Standard	Source (LG)	Text	0.15
1	50	TF	Standard	Source (LG)	Text	0.08
1	10	TF	Standard	Diff	Text	0.60
2	10	TF	Standard	Diff	Text	0.75
3	10	TF	Standard	Diff	Text	0.80
1	20	TF	Standard	Diff	Text	0.48
1	50	TF	Standard	Diff	Text	0.34
1	10	TF	Standard	Combined	Text	0.46
2	10	TF	Standard	Combined	Text	0.61
3	10	TF	Standard	Combined	Text	0.69
1	20	TF	Standard	Combined	Text	0.32
1	50	TF	Standard	Combined	Text	0.22
1	10	TF	Subword	Source (SM)	Text	0.50
2	10	TF	Subword	Source (SM)	Text	0.60
3	10	TF	Subword	Source (SM)	Text	0.69
1	20	TF	Subword	Source (SM)	Text	0.39
1	50	TF	Subword	Source (SM)	Text	0.30
1	10	TF	Subword	Source (MD)	Text	0.53

Continued on next page

Table B.8: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
2	10	TF	Subword	Source (MD)	Text	0.63
3	10	TF	Subword	Source (MD)	Text	0.70
1	20	TF	Subword	Source (MD)	Text	0.42
1	50	TF	Subword	Source (MD)	Text	0.31
1	10	TF	Subword	Source (LG)	Text	0.54
2	10	TF	Subword	Source (LG)	Text	0.66
3	10	TF	Subword	Source (LG)	Text	0.72
1	20	TF	Subword	Source (LG)	Text	0.42
1	50	TF	Subword	Source (LG)	Text	0.28
1	10	TF	Subword	Diff	Text	0.67
2	10	TF	Subword	Diff	Text	0.79
3	10	TF	Subword	Diff	Text	0.84
1	20	TF	Subword	Diff	Text	0.59
1	50	TF	Subword	Diff	Text	0.49
1	10	TF	Subword	Combined	Text	0.67
2	10	TF	Subword	Combined	Text	0.77
3	10	TF	Subword	Combined	Text	0.79
1	20	TF	Subword	Combined	Text	0.58
1	50	TF	Subword	Combined	Text	0.43
1	10	TF-IDF	Standard	Source (SM)	Text	0.34
2	10	TF-IDF	Standard	Source (SM)	Text	0.42
3	10	TF-IDF	Standard	Source (SM)	Text	0.49
1	20	TF-IDF	Standard	Source (SM)	Text	0.32
1	50	TF-IDF	Standard	Source (SM)	Text	0.27
1	10	TF-IDF	Standard	Source (MD)	Text	0.40
2	10	TF-IDF	Standard	Source (MD)	Text	0.47
3	10	TF-IDF	Standard	Source (MD)	Text	0.54
1	20	TF-IDF	Standard	Source (MD)	Text	0.37
1	50	TF-IDF	Standard	Source (MD)	Text	0.33

Continued on next page

Table B.8: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
1	10	TF-IDF	Standard	Source (LG)	Text	0.47
2	10	TF-IDF	Standard	Source (LG)	Text	0.52
3	10	TF-IDF	Standard	Source (LG)	Text	0.59
1	20	TF-IDF	Standard	Source (LG)	Text	0.40
1	50	TF-IDF	Standard	Source (LG)	Text	0.34
1	10	TF-IDF	Standard	Diff	Text	0.75
2	10	TF-IDF	Standard	Diff	Text	0.80
3	10	TF-IDF	Standard	Diff	Text	0.83
1	20	TF-IDF	Standard	Diff	Text	0.66
1	50	TF-IDF	Standard	Diff	Text	0.51
1	10	TF-IDF	Standard	Combined	Text	0.66
2	10	TF-IDF	Standard	Combined	Text	0.76
3	10	TF-IDF	Standard	Combined	Text	0.82
1	20	TF-IDF	Standard	Combined	Text	0.59
1	50	TF-IDF	Standard	Combined	Text	0.49
1	10	TF-IDF	Subword	Source (SM)	Text	0.67
2	10	TF-IDF	Subword	Source (SM)	Text	0.73
3	10	TF-IDF	Subword	Source (SM)	Text	0.77
1	20	TF-IDF	Subword	Source (SM)	Text	0.63
1	50	TF-IDF	Subword	Source (SM)	Text	0.50
1	10	TF-IDF	Subword	Source (MD)	Text	0.67
2	10	TF-IDF	Subword	Source (MD)	Text	0.74
3	10	TF-IDF	Subword	Source (MD)	Text	0.78
1	20	TF-IDF	Subword	Source (MD)	Text	0.62
1	50	TF-IDF	Subword	Source (MD)	Text	0.52
1	10	TF-IDF	Subword	Source (LG)	Text	0.67
2	10	TF-IDF	Subword	Source (LG)	Text	0.75
3	10	TF-IDF	Subword	Source (LG)	Text	0.81
1	20	TF-IDF	Subword	Source (LG)	Text	0.60

Continued on next page

Table B.8: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
1	50	TF-IDF	Subword	Source (LG)	Text	0.48
1	10	TF-IDF	Subword	Diff	Text	0.77
2	10	TF-IDF	Subword	Diff	Text	0.83
3	10	TF-IDF	Subword	Diff	Text	0.86
1	20	TF-IDF	Subword	Diff	Text	0.70
1	50	TF-IDF	Subword	Diff	Text	0.61
1	10	TF-IDF	Subword	Combined	Text	0.75
2	10	TF-IDF	Subword	Combined	Text	0.83
3	10	TF-IDF	Subword	Combined	Text	0.85
1	20	TF-IDF	Subword	Combined	Text	0.69
1	50	TF-IDF	Subword	Combined	Text	0.60
1	10	CodeBERT (Flattening)	-	Source (SM)	Code	0.12
2	10	CodeBERT (Flattening)	-	Source (SM)	Code	0.26
3	10	CodeBERT (Flattening)	-	Source (SM)	Code	0.35
1	20	CodeBERT (Flattening)	-	Source (SM)	Code	0.05
1	50	CodeBERT (Flattening)	-	Source (SM)	Code	0.02
1	10	CodeBERT (Flattening)	-	Source (MD)	Code	0.12
2	10	CodeBERT (Flattening)	-	Source (MD)	Code	0.26
3	10	CodeBERT (Flattening)	-	Source (MD)	Code	0.35
1	20	CodeBERT (Flattening)	-	Source (MD)	Code	0.05
1	50	CodeBERT (Flattening)	-	Source (MD)	Code	0.02
1	10	CodeBERT (Flattening)	-	Source (LG)	Code	0.11
2	10	CodeBERT (Flattening)	-	Source (LG)	Code	0.23
3	10	CodeBERT (Flattening)	-	Source (LG)	Code	0.32
1	20	CodeBERT (Flattening)	-	Source (LG)	Code	0.06
1	50	CodeBERT (Flattening)	-	Source (LG)	Code	0.03
1	10	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.14
2	10	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.27
3	10	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.35

Continued on next page

Table B.8: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
1	20	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.06
1	50	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.01
1	10	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.14
2	10	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.27
3	10	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.35
1	20	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.06
1	50	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.01
1	10	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.12
2	10	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.22
3	10	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.33
1	20	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.06
1	50	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.02

Table B.8: Result Runs for apollo by apolloconfig

K	W	V	T	C	R	Mean Accuracy
1	10	TF	Standard	Source (SM)	Text	0.47
2	10	TF	Standard	Source (SM)	Text	0.59
3	10	TF	Standard	Source (SM)	Text	0.66
1	20	TF	Standard	Source (SM)	Text	0.38
1	50	TF	Standard	Source (SM)	Text	0.24
1	10	TF	Standard	Source (MD)	Text	0.53
2	10	TF	Standard	Source (MD)	Text	0.61
3	10	TF	Standard	Source (MD)	Text	0.68
1	20	TF	Standard	Source (MD)	Text	0.46
1	50	TF	Standard	Source (MD)	Text	0.33
1	10	TF	Standard	Source (LG)	Text	0.49
2	10	TF	Standard	Source (LG)	Text	0.63
3	10	TF	Standard	Source (LG)	Text	0.70
1	20	TF	Standard	Source (LG)	Text	0.39
1	50	TF	Standard	Source (LG)	Text	0.25
1	10	TF	Standard	Diff	Text	0.59
2	10	TF	Standard	Diff	Text	0.80
3	10	TF	Standard	Diff	Text	0.85
1	20	TF	Standard	Diff	Text	0.48
1	50	TF	Standard	Diff	Text	0.33
1	10	TF	Standard	Combined	Text	0.57
2	10	TF	Standard	Combined	Text	0.78
3	10	TF	Standard	Combined	Text	0.84
1	20	TF	Standard	Combined	Text	0.44
1	50	TF	Standard	Combined	Text	0.25
1	10	TF	Subword	Source (SM)	Text	0.68
2	10	TF	Subword	Source (SM)	Text	0.79
3	10	TF	Subword	Source (SM)	Text	0.85
1	20	TF	Subword	Source (SM)	Text	0.61
1	50	TF	Subword	Source (SM)	Text	0.48
1	10	TF	Subword	Source (MD)	Text	0.70

Continued on next page

Table B.9: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
2	10	TF	Subword	Source (MD)	Text	0.83
3	10	TF	Subword	Source (MD)	Text	0.87
1	20	TF	Subword	Source (MD)	Text	0.61
1	50	TF	Subword	Source (MD)	Text	0.46
1	10	TF	Subword	Source (LG)	Text	0.66
2	10	TF	Subword	Source (LG)	Text	0.81
3	10	TF	Subword	Source (LG)	Text	0.86
1	20	TF	Subword	Source (LG)	Text	0.56
1	50	TF	Subword	Source (LG)	Text	0.43
1	10	TF	Subword	Diff	Text	0.72
2	10	TF	Subword	Diff	Text	0.88
3	10	TF	Subword	Diff	Text	0.93
1	20	TF	Subword	Diff	Text	0.61
1	50	TF	Subword	Diff	Text	0.49
1	10	TF	Subword	Combined	Text	0.73
2	10	TF	Subword	Combined	Text	0.89
3	10	TF	Subword	Combined	Text	0.91
1	20	TF	Subword	Combined	Text	0.61
1	50	TF	Subword	Combined	Text	0.47
1	10	TF-IDF	Standard	Source (SM)	Text	0.63
2	10	TF-IDF	Standard	Source (SM)	Text	0.74
3	10	TF-IDF	Standard	Source (SM)	Text	0.76
1	20	TF-IDF	Standard	Source (SM)	Text	0.58
1	50	TF-IDF	Standard	Source (SM)	Text	0.44
1	10	TF-IDF	Standard	Source (MD)	Text	0.68
2	10	TF-IDF	Standard	Source (MD)	Text	0.77
3	10	TF-IDF	Standard	Source (MD)	Text	0.80
1	20	TF-IDF	Standard	Source (MD)	Text	0.62
1	50	TF-IDF	Standard	Source (MD)	Text	0.48

Continued on next page

Table B.9: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
1	10	TF-IDF	Standard	Source (LG)	Text	0.71
2	10	TF-IDF	Standard	Source (LG)	Text	0.81
3	10	TF-IDF	Standard	Source (LG)	Text	0.85
1	20	TF-IDF	Standard	Source (LG)	Text	0.63
1	50	TF-IDF	Standard	Source (LG)	Text	0.48
1	10	TF-IDF	Standard	Diff	Text	0.78
2	10	TF-IDF	Standard	Diff	Text	0.89
3	10	TF-IDF	Standard	Diff	Text	0.92
1	20	TF-IDF	Standard	Diff	Text	0.68
1	50	TF-IDF	Standard	Diff	Text	0.58
1	10	TF-IDF	Standard	Combined	Text	0.80
2	10	TF-IDF	Standard	Combined	Text	0.91
3	10	TF-IDF	Standard	Combined	Text	0.94
1	20	TF-IDF	Standard	Combined	Text	0.72
1	50	TF-IDF	Standard	Combined	Text	0.61
1	10	TF-IDF	Subword	Source (SM)	Text	0.79
2	10	TF-IDF	Subword	Source (SM)	Text	0.87
3	10	TF-IDF	Subword	Source (SM)	Text	0.90
1	20	TF-IDF	Subword	Source (SM)	Text	0.73
1	50	TF-IDF	Subword	Source (SM)	Text	0.53
1	10	TF-IDF	Subword	Source (MD)	Text	0.77
2	10	TF-IDF	Subword	Source (MD)	Text	0.88
3	10	TF-IDF	Subword	Source (MD)	Text	0.90
1	20	TF-IDF	Subword	Source (MD)	Text	0.70
1	50	TF-IDF	Subword	Source (MD)	Text	0.50
1	10	TF-IDF	Subword	Source (LG)	Text	0.75
2	10	TF-IDF	Subword	Source (LG)	Text	0.88
3	10	TF-IDF	Subword	Source (LG)	Text	0.90
1	20	TF-IDF	Subword	Source (LG)	Text	0.68

Continued on next page

Table B.9: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
1	50	TF-IDF	Subword	Source (LG)	Text	0.52
1	10	TF-IDF	Subword	Diff	Text	0.80
2	10	TF-IDF	Subword	Diff	Text	0.91
3	10	TF-IDF	Subword	Diff	Text	0.95
1	20	TF-IDF	Subword	Diff	Text	0.72
1	50	TF-IDF	Subword	Diff	Text	0.64
1	10	TF-IDF	Subword	Combined	Text	0.84
2	10	TF-IDF	Subword	Combined	Text	0.92
3	10	TF-IDF	Subword	Combined	Text	0.95
1	20	TF-IDF	Subword	Combined	Text	0.77
1	50	TF-IDF	Subword	Combined	Text	0.65
1	10	CodeBERT (Flattening)	-	Source (SM)	Code	0.08
2	10	CodeBERT (Flattening)	-	Source (SM)	Code	0.17
3	10	CodeBERT (Flattening)	-	Source (SM)	Code	0.26
1	20	CodeBERT (Flattening)	-	Source (SM)	Code	0.02
1	50	CodeBERT (Flattening)	-	Source (SM)	Code	0.00
1	10	CodeBERT (Flattening)	-	Source (MD)	Code	0.08
2	10	CodeBERT (Flattening)	-	Source (MD)	Code	0.16
3	10	CodeBERT (Flattening)	-	Source (MD)	Code	0.27
1	20	CodeBERT (Flattening)	-	Source (MD)	Code	0.02
1	50	CodeBERT (Flattening)	-	Source (MD)	Code	0.00
1	10	CodeBERT (Flattening)	-	Source (LG)	Code	0.09
2	10	CodeBERT (Flattening)	-	Source (LG)	Code	0.19
3	10	CodeBERT (Flattening)	-	Source (LG)	Code	0.28
1	20	CodeBERT (Flattening)	-	Source (LG)	Code	0.03
1	50	CodeBERT (Flattening)	-	Source (LG)	Code	0.00
1	10	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.08
2	10	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.17
3	10	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.26

Continued on next page

Table B.9: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
1	20	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.03
1	50	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.00
1	10	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.09
2	10	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.18
3	10	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.26
1	20	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.03
1	50	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.00
1	10	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.10
2	10	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.21
3	10	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.29
1	20	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.04
1	50	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.01

Table B.9: Result Runs for selenium by SeleniumHQ

K	W	V	T	C	R	Mean Accuracy
1	10	TF	Standard	Source (SM)	Text	0.22
2	10	TF	Standard	Source (SM)	Text	0.43
3	10	TF	Standard	Source (SM)	Text	0.56
1	20	TF	Standard	Source (SM)	Text	0.13
1	50	TF	Standard	Source (SM)	Text	0.04
1	10	TF	Standard	Source (MD)	Text	0.22
2	10	TF	Standard	Source (MD)	Text	0.45
3	10	TF	Standard	Source (MD)	Text	0.58
1	20	TF	Standard	Source (MD)	Text	0.12
1	50	TF	Standard	Source (MD)	Text	0.04
1	10	TF	Standard	Source (LG)	Text	0.22
2	10	TF	Standard	Source (LG)	Text	0.45
3	10	TF	Standard	Source (LG)	Text	0.58
1	20	TF	Standard	Source (LG)	Text	0.11
1	50	TF	Standard	Source (LG)	Text	0.05
1	10	TF	Standard	Diff	Text	0.38
2	10	TF	Standard	Diff	Text	0.57
3	10	TF	Standard	Diff	Text	0.69
1	20	TF	Standard	Diff	Text	0.29
1	50	TF	Standard	Diff	Text	0.16
1	10	TF	Standard	Combined	Text	0.27
2	10	TF	Standard	Combined	Text	0.49
3	10	TF	Standard	Combined	Text	0.64
1	20	TF	Standard	Combined	Text	0.17
1	50	TF	Standard	Combined	Text	0.08
1	10	TF	Subword	Source (SM)	Text	0.39
2	10	TF	Subword	Source (SM)	Text	0.64
3	10	TF	Subword	Source (SM)	Text	0.72
1	20	TF	Subword	Source (SM)	Text	0.25
1	50	TF	Subword	Source (SM)	Text	0.10
1	10	TF	Subword	Source (MD)	Text	0.41

Continued on next page

Table B.10: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
2	10	TF	Subword	Source (MD)	Text	0.64
3	10	TF	Subword	Source (MD)	Text	0.73
1	20	TF	Subword	Source (MD)	Text	0.26
1	50	TF	Subword	Source (MD)	Text	0.10
1	10	TF	Subword	Source (LG)	Text	0.39
2	10	TF	Subword	Source (LG)	Text	0.63
3	10	TF	Subword	Source (LG)	Text	0.74
1	20	TF	Subword	Source (LG)	Text	0.24
1	50	TF	Subword	Source (LG)	Text	0.11
1	10	TF	Subword	Diff	Text	0.46
2	10	TF	Subword	Diff	Text	0.64
3	10	TF	Subword	Diff	Text	0.76
1	20	TF	Subword	Diff	Text	0.36
1	50	TF	Subword	Diff	Text	0.23
1	10	TF	Subword	Combined	Text	0.45
2	10	TF	Subword	Combined	Text	0.67
3	10	TF	Subword	Combined	Text	0.81
1	20	TF	Subword	Combined	Text	0.30
1	50	TF	Subword	Combined	Text	0.18
1	10	TF-IDF	Standard	Source (SM)	Text	0.48
2	10	TF-IDF	Standard	Source (SM)	Text	0.62
3	10	TF-IDF	Standard	Source (SM)	Text	0.69
1	20	TF-IDF	Standard	Source (SM)	Text	0.40
1	50	TF-IDF	Standard	Source (SM)	Text	0.28
1	10	TF-IDF	Standard	Source (MD)	Text	0.51
2	10	TF-IDF	Standard	Source (MD)	Text	0.66
3	10	TF-IDF	Standard	Source (MD)	Text	0.70
1	20	TF-IDF	Standard	Source (MD)	Text	0.42
1	50	TF-IDF	Standard	Source (MD)	Text	0.28

Continued on next page

Table B.10: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
1	10	TF-IDF	Standard	Source (LG)	Text	0.49
2	10	TF-IDF	Standard	Source (LG)	Text	0.66
3	10	TF-IDF	Standard	Source (LG)	Text	0.72
1	20	TF-IDF	Standard	Source (LG)	Text	0.39
1	50	TF-IDF	Standard	Source (LG)	Text	0.26
1	10	TF-IDF	Standard	Diff	Text	0.53
2	10	TF-IDF	Standard	Diff	Text	0.66
3	10	TF-IDF	Standard	Diff	Text	0.77
1	20	TF-IDF	Standard	Diff	Text	0.46
1	50	TF-IDF	Standard	Diff	Text	0.37
1	10	TF-IDF	Standard	Combined	Text	0.55
2	10	TF-IDF	Standard	Combined	Text	0.73
3	10	TF-IDF	Standard	Combined	Text	0.80
1	20	TF-IDF	Standard	Combined	Text	0.46
1	50	TF-IDF	Standard	Combined	Text	0.31
1	10	TF-IDF	Subword	Source (SM)	Text	0.68
2	10	TF-IDF	Subword	Source (SM)	Text	0.81
3	10	TF-IDF	Subword	Source (SM)	Text	0.86
1	20	TF-IDF	Subword	Source (SM)	Text	0.58
1	50	TF-IDF	Subword	Source (SM)	Text	0.38
1	10	TF-IDF	Subword	Source (MD)	Text	0.68
2	10	TF-IDF	Subword	Source (MD)	Text	0.80
3	10	TF-IDF	Subword	Source (MD)	Text	0.85
1	20	TF-IDF	Subword	Source (MD)	Text	0.59
1	50	TF-IDF	Subword	Source (MD)	Text	0.39
1	10	TF-IDF	Subword	Source (LG)	Text	0.66
2	10	TF-IDF	Subword	Source (LG)	Text	0.80
3	10	TF-IDF	Subword	Source (LG)	Text	0.85
1	20	TF-IDF	Subword	Source (LG)	Text	0.56

Continued on next page

Table B.10: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
1	50	TF-IDF	Subword	Source (LG)	Text	0.36
1	10	TF-IDF	Subword	Diff	Text	0.63
2	10	TF-IDF	Subword	Diff	Text	0.77
3	10	TF-IDF	Subword	Diff	Text	0.84
1	20	TF-IDF	Subword	Diff	Text	0.54
1	50	TF-IDF	Subword	Diff	Text	0.40
1	10	TF-IDF	Subword	Combined	Text	0.63
2	10	TF-IDF	Subword	Combined	Text	0.81
3	10	TF-IDF	Subword	Combined	Text	0.88
1	20	TF-IDF	Subword	Combined	Text	0.53
1	50	TF-IDF	Subword	Combined	Text	0.39
1	10	CodeBERT (Flattening)	-	Source (SM)	Code	0.08
2	10	CodeBERT (Flattening)	-	Source (SM)	Code	0.18
3	10	CodeBERT (Flattening)	-	Source (SM)	Code	0.29
1	20	CodeBERT (Flattening)	-	Source (SM)	Code	0.06
1	50	CodeBERT (Flattening)	-	Source (SM)	Code	0.02
1	10	CodeBERT (Flattening)	-	Source (MD)	Code	0.08
2	10	CodeBERT (Flattening)	-	Source (MD)	Code	0.17
3	10	CodeBERT (Flattening)	-	Source (MD)	Code	0.28
1	20	CodeBERT (Flattening)	-	Source (MD)	Code	0.06
1	50	CodeBERT (Flattening)	-	Source (MD)	Code	0.02
1	10	CodeBERT (Flattening)	-	Source (LG)	Code	0.09
2	10	CodeBERT (Flattening)	-	Source (LG)	Code	0.19
3	10	CodeBERT (Flattening)	-	Source (LG)	Code	0.28
1	20	CodeBERT (Flattening)	-	Source (LG)	Code	0.05
1	50	CodeBERT (Flattening)	-	Source (LG)	Code	0.02
1	10	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.08
2	10	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.17
3	10	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.28

Continued on next page

Table B.10: Continued from previous page

K	W	V	T	C	R	Mean Accuracy
1	20	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.05
1	50	CodeBERT (Sum Pooling)	-	Source (SM)	Code	0.02
1	10	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.08
2	10	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.16
3	10	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.27
1	20	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.05
1	50	CodeBERT (Sum Pooling)	-	Source (MD)	Code	0.02
1	10	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.10
2	10	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.18
3	10	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.29
1	20	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.06
1	50	CodeBERT (Sum Pooling)	-	Source (LG)	Code	0.03

Table B.10: Result Runs for nacos by alibaba