> library(textreuse)

> dir <- ("corpus")

> corpus <- TextReuseCorpus(dir = dir, meta = list(title = "Collosal Cave Adventure"), tokenizer = tokenize\_ngrams, n = 7)

Loading, tokenizing, and hashing 13 documents.

|=====================================================================| 100%

There were 50 or more warnings (use warnings() to see the first 50)

> comparisons <- pairwise\_compare(corpus, jaccard\_similarity)

Making 78 comparisons.

|=====================================================================| 100%

> comparisons

arna0440-dos-code\_text arna0440-source\_text

arna0440-dos-code\_text NA 0

arna0440-source\_text NA NA

blac0350-code\_text NA NA

dove0550-code\_text NA NA

jame0551-code\_text NA NA

long0500-code\_text NA NA

muno0370-code\_text NA NA

olss0551-code\_text NA NA

oska0551-code\_text NA NA

supn0350-code\_text NA NA

vane0560-code\_text NA NA

wood0350-code\_text NA NA

wood043b-code\_text NA NA

blac0350-code\_text dove0550-code\_text

arna0440-dos-code\_text 0 0

arna0440-source\_text 0 0

blac0350-code\_text NA 0

dove0550-code\_text NA NA

jame0551-code\_text NA NA

long0500-code\_text NA NA

muno0370-code\_text NA NA

olss0551-code\_text NA NA

oska0551-code\_text NA NA

supn0350-code\_text NA NA

vane0560-code\_text NA NA

wood0350-code\_text NA NA

wood043b-code\_text NA NA

jame0551-code\_text long0500-code\_text

arna0440-dos-code\_text 0 0

arna0440-source\_text 0 0

blac0350-code\_text 0 0

dove0550-code\_text 0 0

jame0551-code\_text NA 0

long0500-code\_text NA NA

muno0370-code\_text NA NA

olss0551-code\_text NA NA

oska0551-code\_text NA NA

supn0350-code\_text NA NA

vane0560-code\_text NA NA

wood0350-code\_text NA NA

wood043b-code\_text NA NA

muno0370-code\_text olss0551-code\_text

arna0440-dos-code\_text 0.00000000 0.0003876971

arna0440-source\_text 0.09239156 0.0000000000

blac0350-code\_text 0.00000000 0.0000000000

dove0550-code\_text 0.00000000 0.0000000000

jame0551-code\_text 0.00000000 0.0001653302

long0500-code\_text 0.00000000 0.0000000000

muno0370-code\_text NA 0.0001121076

olss0551-code\_text NA NA

oska0551-code\_text NA NA

supn0350-code\_text NA NA

vane0560-code\_text NA NA

wood0350-code\_text NA NA

wood043b-code\_text NA NA

oska0551-code\_text supn0350-code\_text

arna0440-dos-code\_text 0.0000000 0

arna0440-source\_text 0.0000000 0

blac0350-code\_text 0.0000000 1

dove0550-code\_text 0.0000000 0

jame0551-code\_text 0.0000000 0

long0500-code\_text 0.6969697 0

muno0370-code\_text 0.0000000 0

olss0551-code\_text 0.0000000 0

oska0551-code\_text NA 0

supn0350-code\_text NA NA

vane0560-code\_text NA NA

wood0350-code\_text NA NA

wood043b-code\_text NA NA

vane0560-code\_text wood0350-code\_text

arna0440-dos-code\_text 0.000000e+00 0.00000000

arna0440-source\_text 1.434444e-02 0.16697222

blac0350-code\_text 0.000000e+00 0.00000000

dove0550-code\_text 0.000000e+00 0.00000000

jame0551-code\_text 3.028468e-05 0.00000000

long0500-code\_text 0.000000e+00 0.01831453

muno0370-code\_text 4.135749e-02 0.42620968

olss0551-code\_text 1.124290e-04 0.00000000

oska0551-code\_text 0.000000e+00 0.01761330

supn0350-code\_text 0.000000e+00 0.00000000

vane0560-code\_text NA 0.04067647

wood0350-code\_text NA NA

wood043b-code\_text NA NA

wood043b-code\_text

arna0440-dos-code\_text 0.00000000

arna0440-source\_text 0.00986704

blac0350-code\_text 0.00000000

dove0550-code\_text 0.00000000

jame0551-code\_text 0.00000000

long0500-code\_text 0.00000000

muno0370-code\_text 0.05947801

olss0551-code\_text 0.00000000

oska0551-code\_text 0.00000000

supn0350-code\_text 0.00000000

vane0560-code\_text 0.02433006

wood0350-code\_text 0.05696615

wood043b-code\_text NA

> pairwise\_candidates(comparisons)

a b score

1 arna0440-dos-code\_text arna0440-source\_text 0.000000e+00

2 arna0440-dos-code\_text blac0350-code\_text 0.000000e+00

4 arna0440-dos-code\_text dove0550-code\_text 0.000000e+00

7 arna0440-dos-code\_text jame0551-code\_text 0.000000e+00

11 arna0440-dos-code\_text long0500-code\_text 0.000000e+00

16 arna0440-dos-code\_text muno0370-code\_text 0.000000e+00

22 arna0440-dos-code\_text olss0551-code\_text 3.876971e-04

29 arna0440-dos-code\_text oska0551-code\_text 0.000000e+00

37 arna0440-dos-code\_text supn0350-code\_text 0.000000e+00

46 arna0440-dos-code\_text vane0560-code\_text 0.000000e+00

56 arna0440-dos-code\_text wood0350-code\_text 0.000000e+00

67 arna0440-dos-code\_text wood043b-code\_text 0.000000e+00

3 arna0440-source\_text blac0350-code\_text 0.000000e+00

5 arna0440-source\_text dove0550-code\_text 0.000000e+00

8 arna0440-source\_text jame0551-code\_text 0.000000e+00

12 arna0440-source\_text long0500-code\_text 0.000000e+00

17 arna0440-source\_text muno0370-code\_text 9.239156e-02

23 arna0440-source\_text olss0551-code\_text 0.000000e+00

30 arna0440-source\_text oska0551-code\_text 0.000000e+00

38 arna0440-source\_text supn0350-code\_text 0.000000e+00

47 arna0440-source\_text vane0560-code\_text 1.434444e-02

57 arna0440-source\_text wood0350-code\_text 1.669722e-01

68 arna0440-source\_text wood043b-code\_text 9.867040e-03

6 blac0350-code\_text dove0550-code\_text 0.000000e+00

9 blac0350-code\_text jame0551-code\_text 0.000000e+00

13 blac0350-code\_text long0500-code\_text 0.000000e+00

18 blac0350-code\_text muno0370-code\_text 0.000000e+00

24 blac0350-code\_text olss0551-code\_text 0.000000e+00

31 blac0350-code\_text oska0551-code\_text 0.000000e+00

39 blac0350-code\_text supn0350-code\_text 1.000000e+00

48 blac0350-code\_text vane0560-code\_text 0.000000e+00

58 blac0350-code\_text wood0350-code\_text 0.000000e+00

69 blac0350-code\_text wood043b-code\_text 0.000000e+00

10 dove0550-code\_text jame0551-code\_text 0.000000e+00

14 dove0550-code\_text long0500-code\_text 0.000000e+00

19 dove0550-code\_text muno0370-code\_text 0.000000e+00

25 dove0550-code\_text olss0551-code\_text 0.000000e+00

32 dove0550-code\_text oska0551-code\_text 0.000000e+00

40 dove0550-code\_text supn0350-code\_text 0.000000e+00

49 dove0550-code\_text vane0560-code\_text 0.000000e+00

59 dove0550-code\_text wood0350-code\_text 0.000000e+00

70 dove0550-code\_text wood043b-code\_text 0.000000e+00

15 jame0551-code\_text long0500-code\_text 0.000000e+00

20 jame0551-code\_text muno0370-code\_text 0.000000e+00

26 jame0551-code\_text olss0551-code\_text 1.653302e-04

33 jame0551-code\_text oska0551-code\_text 0.000000e+00

41 jame0551-code\_text supn0350-code\_text 0.000000e+00

50 jame0551-code\_text vane0560-code\_text 3.028468e-05

60 jame0551-code\_text wood0350-code\_text 0.000000e+00

71 jame0551-code\_text wood043b-code\_text 0.000000e+00

21 long0500-code\_text muno0370-code\_text 0.000000e+00

27 long0500-code\_text olss0551-code\_text 0.000000e+00

34 long0500-code\_text oska0551-code\_text 6.969697e-01

42 long0500-code\_text supn0350-code\_text 0.000000e+00

51 long0500-code\_text vane0560-code\_text 0.000000e+00

61 long0500-code\_text wood0350-code\_text 1.831453e-02

72 long0500-code\_text wood043b-code\_text 0.000000e+00

28 muno0370-code\_text olss0551-code\_text 1.121076e-04

35 muno0370-code\_text oska0551-code\_text 0.000000e+00

43 muno0370-code\_text supn0350-code\_text 0.000000e+00

52 muno0370-code\_text vane0560-code\_text 4.135749e-02

62 muno0370-code\_text wood0350-code\_text 4.262097e-01

73 muno0370-code\_text wood043b-code\_text 5.947801e-02

36 olss0551-code\_text oska0551-code\_text 0.000000e+00

44 olss0551-code\_text supn0350-code\_text 0.000000e+00

53 olss0551-code\_text vane0560-code\_text 1.124290e-04

63 olss0551-code\_text wood0350-code\_text 0.000000e+00

74 olss0551-code\_text wood043b-code\_text 0.000000e+00

45 oska0551-code\_text supn0350-code\_text 0.000000e+00

54 oska0551-code\_text vane0560-code\_text 0.000000e+00

64 oska0551-code\_text wood0350-code\_text 1.761330e-02

75 oska0551-code\_text wood043b-code\_text 0.000000e+00

55 supn0350-code\_text vane0560-code\_text 0.000000e+00

65 supn0350-code\_text wood0350-code\_text 0.000000e+00

76 supn0350-code\_text wood043b-code\_text 0.000000e+00

66 vane0560-code\_text wood0350-code\_text 4.067647e-02

77 vane0560-code\_text wood043b-code\_text 2.433006e-02

78 wood0350-code\_text wood043b-code\_text 5.696615e-02

> df <- pairwise\_candidates(comparisons)

> View(df)

> write.csv(comparisons, file = "textreuse-comparisons.csv")