1.Shingles Answers:

# Calculate unique shingles for two documents.

shingles1 <- shingles('ABRACADABRA', 2)

shingles2 <- shingles('BRICABRAC', 2)

**1.a)**

# How many 2-shingles does ABRACADABRA have?

length(shingles1)

# 7

# ABRACADABRA

# AB, BR, RA, AC, CA, AD, DA, AB, BR, RA

# AB, BR, RA, AC, CA, AD, DA

1.b)

# How many 2-shingles does BRICABRAC have?

length(shingles2)

# 7

# BRICABRAC

# BR, RI, IC, CA, AB, BR, RA, AC

# BR, RI, IC, CA, AB, RA, AC

1.c)

# How many 2-shingles do they have in common?

intersect(shingles1, shingles2)

# 5

# Union:

# AB, BR, RA, AC, CA, AD, DA, RI, IC

1.d)

# What is the Jaccard similarity between the two documents"?

print(paste(length(intersect(shingles1, shingles2)), length(union(shingles1, shingles2)), sep = '/'))

length(intersect(shingles1, shingles2)) / length(union(shingles1, shingles2))

# 5/9