# Przykład merge()

# utwórz ramkę danych authors

# funkcja I() użyje nazw kolumn, aby uzyskać kolejność sortowania

authors <- data.frame(

surname = I(c("Tukey", "Venables", "Tierney", "Ripley", "McNeil")),

nationality = c("US", "Australia", "US", "UK", "Australia"),

deceased = c("yes", rep("no", 4)))

authorN <- within(authors, { name <- surname; rm(surname) })

# zmienił kolejność kolumn

# zmienił nazwę etykiety "surname" na "name" i usunął "surname"

books <- data.frame(

name = I(c("Tukey", "Venables", "Tierney",

"Ripley", "Ripley", "McNeil", "R Core")),

title = c("Exploratory Data Analysis",

"Modern Applied Statistics ...",

"LISP-STAT",

"Spatial Statistics", "Stochastic Simulation",

"Interactive Data Analysis",

"An Introduction to R"),

other.author = c(NA, "Ripley", NA, NA, NA, NA,

"Venables & Smith"))

# Przykłady łączenia damek danych

# 1. Funkcja merge() bez parametrów łączenia, jeśli kolumny mają identyczną nazwę:

m0 <- merge(authorN, books)

# pierwsza kolumna wskazuje po jakiej kolumnie złączono ramki

# potem brał kolejno kolumny z ramki authorN, następnie z books

# złączył po kolumnie names alfabetycznie, ale obserwacji jest 6.

# Spójrz na nazwiska:

# - jeśli w którejś kolumnie nie było danego nazwiska (R core) to pominął rekord,

# - jeśli w którejś kolumnie nazwisko było zdublowane, to utrzymał to podwojenie (Ripley)

# 2. Funkcja merge() z określonymi parametrami łączenia, gdy nazwy kolumn różnią się:

m1 <- merge(authors, books, by.x = "surname", by.y = "name")

# 3. Funkcja merge() z określonymi parametrami łączenia, gdy nazwy kolumn różnią się:

m2 <- merge(books, authors, by.x = "name", by.y = "surname")

stopifnot(exprs = {

identical(m0, m2[, names(m0)])

as.character(m1[, 1]) == as.character(m2[, 1])

all.equal(m1[, -1], m2[, -1][ names(m1)[-1] ])

identical(dim(merge(m1, m2, by = NULL)),

c(nrow(m1)\*nrow(m2), ncol(m1)+ncol(m2)))

})

# Nazwisko "R core" pojawia się dopiero, gdy użyjesz parametr all=TRUE:

merge(authors, books, by.x = "surname", by.y = "name", all = TRUE)

# Przykład użycia merge() z danymi nieporównywalnymi między sobą:

x <- data.frame(k1 = c(NA,NA,3,4,5), k2 = c(1,NA,NA,4,5), data = 1:5)

y <- data.frame(k1 = c(NA,2,NA,4,5), k2 = c(NA,NA,3,4,5), data = 1:5)

# NA's match. Połączy tylko te wiersze, które mają identyczną zawartość w kolumnach k1 i k2. Tutaj są to wiersze 4 i 5, a potem wiersz, który ma NA dwa razy w kolumnie k1 i k2, ale dla ramki x ma inną cyfrę i dla ramki y inną:

merge(x, y, by = c("k1","k2"))

# NA's match, więc 6 obserwacji, bardzo rozrzuconych:

merge(x, y, by = "k1")

# a jak się wstawi parametr incomparables = NA, to zostawi tylko identyczne wiersze:

merge(x, y, by = "k1", incomparables = NA)

# to samo co powyżej, tylko połączy po k2 (tu akurat są te same wartości co w k1):

merge(x, y, by = "k2", incomparables = NA) # 2 obserwacje