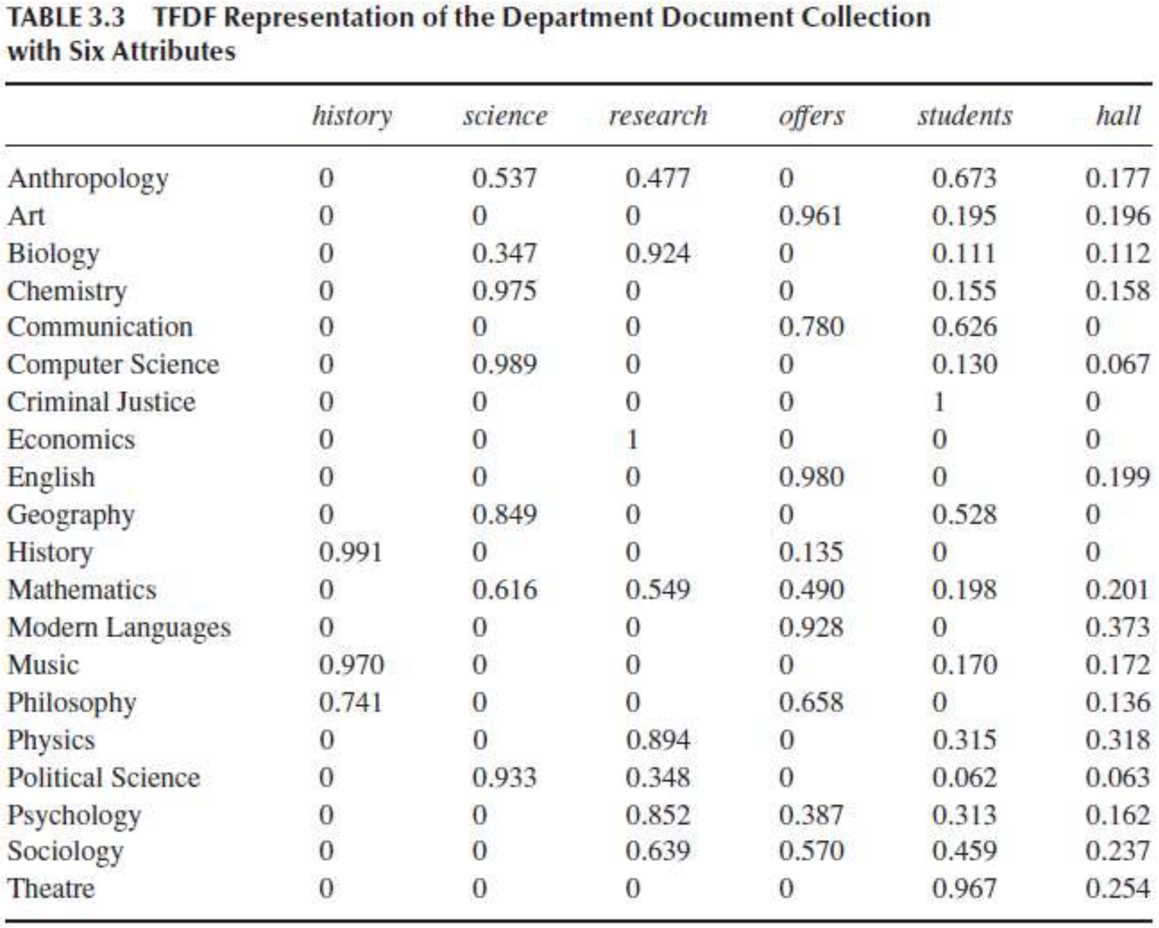

TABLE 3.3 TFDF Representation of the Department Document Collection with Six Attributes

	history	*science*	*research*	*offers*	*students*	*hall*
Anthropology	0	0.537	0.477	0	0.673	0.177
Art	0	0	0	0.961	0.195	0.196
Biology	0	0.347	0.924	0	0.111	0.112
Chemistry	0	0.975	0	0	0.155	0.158
Communication	0	0	0	0.780	0.626	0
Computer Science	0	0.989	0	0	0.130	0.067
Criminal Justice	0	0	0	0	1	0
Economics	0	0	1	0	0	0
English	0	0	0	0.980	0	0.199
Geography	0	0.849	0	0	0.528	0
History	0.991	0	0	0.135	0	0
Mathematics	0	0.616	0.549	0.490	0.198	0.201
Modern Languages	0	0	0	0.928	0	0.373
Music	0.970	0	0	0	0.170	0.172
Philosophy	0.741	0	0	0.658	0	0.136
Physics	0	0	0.894	0	0.315	0.318
Political Science	0	0.933	0.348	0	0.062	0.063
Psychology	0	0	0.852	0.387	0.313	0.162
Sociology	0	0	0.639	0.570	0.459	0.237
Theatre	0	0	0	0	0.967	0.254

Apply k means clustering for the above table (Sample code of K means clustering algorithm applied to iris data)

```
library(datasets)

head(iris)

#############

library(ggplot2)

#######################

str(iris) #view structure of dataset

################################

summary(iris)

head(iris)

################################

iris.new<- iris[,c(1,2,3,4)]
iris.class<- iris[,"Species"]

#################################

head(iris.new)

###############################

normalize <- function(x){

  return ((x-min(x))/(max(x)-min(x)))

}


iris.new$Sepal.Length<- normalize(iris.new$Sepal.Length)

iris.new$Sepal.Width<- normalize(iris.new$Sepal.Width)

iris.new$Petal.Length<- normalize(iris.new$Petal.Length)

iris.new$Petal.Width<- normalize(iris.new$Petal.Width)

head(iris.new)

##############################################################################

result<- kmeans(iris.new,3) #aplly k-means algorithm with no. of centroids(k)=3

result$size # gives no. of records in each cluster


result$centers # gives value of cluster center datapoint value(3 centers for k=3)

# gives value of cluster center datapoint value(3 centers for k=3)


result$cluster #gives cluster vector showing the custer where each record falls


par(mfrow=c(2,2), mar=c(5,4,2,2))

plot(iris.new[c(1,2)], col=result$cluster)# Plot to see how Sepal.Length and Sepal.Width data points have been distributed in clusters

plot(iris.new[c(1,2)], col=iris.class)# Plot to see how Sepal.Length and Sepal.Width data points have been distributed originally as per "class" attribute in dataset

plot(iris.new[c(3,4)], col=result$cluster)# Plot to see how Petal.Length and Petal.Width data points have been distributed in clusters

plot(iris.new[c(3,4)], col=iris.class)

table(result$cluster,iris.class)

#########################################################
```

```
ggplot(iris, aes(Petal.Length, Petal.Width, color = Species)) + geom_point()


#################################

set.seed(20)

irisCluster <- kmeans(iris[, 3:4], 3, nstart = 20)

table(irisCluster$cluster, iris$Species)
```