

# The most popular topics in IEEE publications in 2014

Satoko Kora  
Illinois State University  
skora@ilstu.edu

April 15, 2015

## Introduction

## Obtain data

In order to retrieve dataset from IEEE API, I used "IEEEER" package available on GitHub, developed by kbroman. To run the methods, it is required to install the package with devtools package for the first time.

```
##install.packages("devtools")
##library(devtools)
##install_github("ropensci/IEEEER")

#load a package to query IEEE Xplore Search Gateway
library(IEEEER, quietly=TRUE)

# make a data request query
ie3data <- suppressMessages(IEEE_search(query = list(pu="IEEE", ctype = "Journals", pys =
  2014, pye = 2014, oa=1, rs=1), limit=2000))

# extract Controlled Index Terms in academic journal metadata
terms<-ie3data[,c("controlledterms")]
```

## Scrub data

```
# split vectors of controlled index terms into an array
index <- unlist(strsplit(terms, "[|]"))

# Count frequencies of keywords and get top 5 keywords
sortedTable <- sort(table(index), TRUE)[1:5]

# Make a bar chart based on the data
df <- data.frame(sortedTable)
# get No.1 keyword
top<-df[1,1]
```

## Explore data

## Results

```
library(xtable)

## Printing a table using xtable ##
caption <- "Top 5 keywords in IEEE publications in 2014"
c1<-c("1.", "2.", "3.", "4.", "5.")
row.names(df) <- paste(c1, row.names(df), sep=" ")
colnames(df) <- c("Frequencies")
print(xtable(df, caption=caption),
      caption.placement="top", include.row.names=TRUE)
```

The following figure represents the top 5 keywords based on academic journals published in 2014. The data is retrieved by IEEE API, and the search query limits academic journals that are open-access, and published by IEEE in 2014. The top five keywords are: 59

Table 1: Top 5 keywords in IEEE publications in 2014

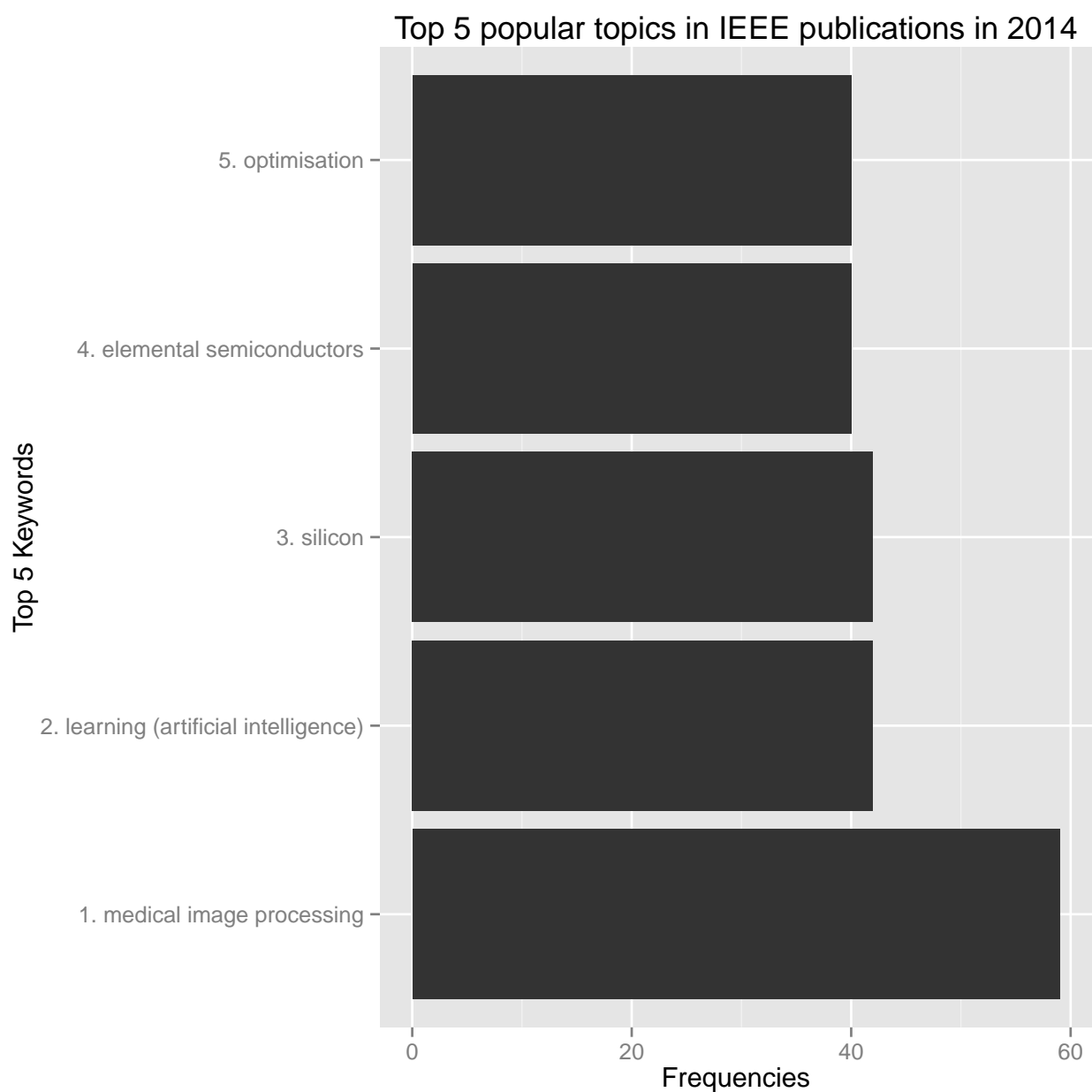
|                                       | Frequencies |
|---------------------------------------|-------------|
| 1. medical image processing           | 59          |
| 2. learning (artificial intelligence) | 42          |
| 3. silicon                            | 42          |
| 4. elemental semiconductors           | 40          |
| 5. optimisation                       | 40          |

```
# load a package to make a graph
library("ggplot2", quietly=TRUE)

### Labels for the graph

##c1←c("1.", "2.", "3.", "4.", "5.")
##cols←paste(c1, row.names(df), sep=" ")

g ← ggplot(df, aes(x=row.names(df), y=Frequencies))
g + xlab("Top 5 Keywords") + geom_bar(stat="identity") + coord_flip() + labs(title = "Top 5
popular topics in IEEE publications in 2014")
```



```
artilcles ← IEEE_search(query = list(pu="IEEE", ctype = "Journals", pys = 2014, pye = 2014, oa
=1,rs=1, cntrlterms="medical image processing"), limit=300)
```

```
retrieved batch 1
retrieved batch 2
retrieved batch 3
```

```
artilcles.gr ← artilcles[grepl("medical image processing", artilcles$controlledterms),]
affs←artilcles.gr$affiliations
arrAffs ← strsplit(affs, "[,]")
countries ← c(1:length(arrAffs))
j ←1
for(i in 1:length(arrAffs)){
  countries[i] ← arrAffs[[i]][length(arrAffs[[i]])]
}

## Count frequencies of keywords and get top 5 keywords
sortedTable ← sort(table(countries), TRUE)[1:5]

## Make a bar chart based on the data
df ← data.frame(sortedTable)

### Labels for the graph
colnames(df) ← c("Num")
c1←c("1.", "2.", "3.", "4.", "5.")
cols←paste(c1, row.names(df), sep=" ")

g ← ggplot(df, aes(x=cols, y=Num))
g + xlab("Countries") + ylab("Number of journals") + geom_bar(stat="identity")+ labs(title = "
Top 5 countries that produced medical image processing related journals in 2014")
```

Top 5 countries that produced medical image processing related journals in 201

