The most popular topics in IEEE publications in 2014

Satoko Kora
Illinois State University
skora@ilstu.edu

April 15, 2015

Introduction

IEEE stands for Institute of Electrical and Electronics Engineers. It is the worlds largest professional association dedicated to advancing technological innovation and excellence for the benefit of humanity. IEEE has a large variety of contributions to the advancement of technology, known as a vast group of IEEE members around 430,000 people over 160 countries, and as a leader of international standards of emerging technologies, and sponsored more than 1600 conferences.

In terms of publications, IEEE owns a third of the worlds technical literature in electrical engineering, computer science, and electronics. Approximately 180 transactions, journals, and magazines are published annucally. IEEE Xplore Digital Library is an online library containing more than 3.5 million documents from IEEE and IEEE journals, transactions, magazines, letters, conference proceedings, and active IEEE standards.

This paper will explore the worlds most highly cited IEEE journals and reveal the most popular topics in academic journals published by IEEE in 2014. For this research, IEEE Xplore Search Gateway service is used to retrieve statistical data. IEEE Xplore Search Gateway is a service providing information about all publications stored in IEEE Xplore Digital Library. Among abundant information about academic journals, I will retrieve controlled index term information which represents a set of keywords and main topics dealt with in each academic journal, and gather data showing how frequent each term is chosen as a topic in academic journals published in 2014. In this way, I would like to reveal the top 5 popular topics in IEEE journals in 2014.

Obtain data

In order to retrieve dataset from IEEE Xplore Search Gateway, I used "IEEER" package available on GitHub, developed by kbroman. The package contains methods to make a query to access to the library based on our criteria and handles memory consumption for users when they retrieve a large amount of data from the enormous library. 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/IEEER")

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

The following is sending a request to retrieve data about open-access academic journals published by IEEE in 2014.

```
# make a data request query and get data from IEEE Xplore Search Gateway
## pu=IEEE: published by IEEE, ctype="Journals": academic journals only
## pys=2014:Start value of publication year,  pye=2014:Start value of publication year
## oa=1: Open Access journals only, rs=1:Sequence number of first record to fetch is 1
## limit: maximun number of reterieved journals must ebe 2000
ie3data ← suppressMessages(IEEE_search(query = list(pu="IEEE", ctype = "Journals", pys =
    2014, pye = 2014,oa=1,rs=1), limit=2000))
```

The total number of corresponding journals is 1195.

Scrub data

After obtaining corresponding information, controlled index terms, which represents keyword and topics dealt within journals, are extracted from the data. Journal information is retrieved in a data frame form, and firstly the column "controlledterms" is extracted.

```
# extract only Controlled Index Terms data in academic journal information
terms←ie3data[,c("controlledterms")]
```

Usually an academic journal includes more than one index term, so multiple words are stored as one string sepated with "—" character. The following is splitting those strings and making one vector of all retrieved index terms.

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

In the following, a ranking of top 5 popular topics is created. Table function counts the number of times of each word's appearnce in the vector. The table is sorted by the frequency in descending order and top 5 terms are set into table. No.1 popular topic term is set to top varaiable.

```
# 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←row.names(df)[[1]]
top
```

```
[1] "medical image processing"
```

According to the processed data table, No.1 popular topic in 2014 is medical image processing, and the number of its appearance in journals in 2014 is c(59, 42, 42, 40, 40) times.

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.rownames=TRUE)
```

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

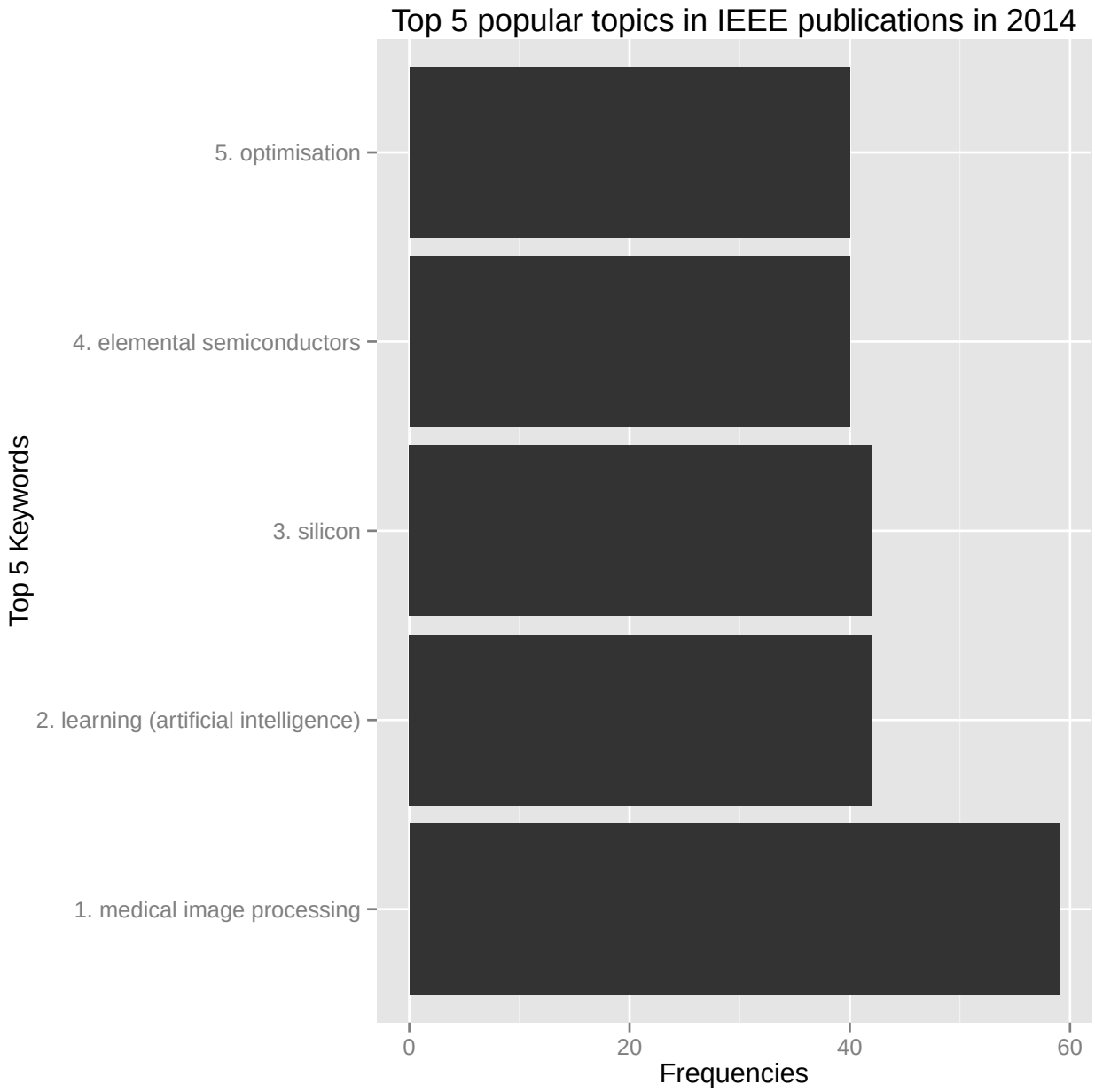
The following figure represents the top 5 keywords based on academic journals published in 2014. The data is retrieved by IEEE API, and the search query limits academic journals that are open-access, and published by IEEE in 2014. The top five keywords are: medical image processing

```
# 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[grep("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")
```