

SOTL-second pass

Ebner

May 27, 2016

The input file is “Combined clean 5-31-16.csv”

Print Initial dimensions

Make all words lower case

Remove preliminary stop words

Remove extra white space

Remove rows with empty abstracts (after cleaning)

Show dimensions after empty’s removal

Show summary statistics of word count in remaining abstracts

```
# read in the whole csv file for processing
orig <- read.csv("Combined clean 5-31-16.csv", stringsAsFactors = FALSE)
c <- orig
```

```
dim(c) # Rows and Columns in the file
```

```
## [1] 5310 27
```

```
# Clean the abstract data before processing
# make all lower case
c$abstract <- tolower(c$abstract)
# remove punctuation
c$abstract <- gsub("[^[:alnum:][:space:]]", " ", c$abstract)
# remove stop words
c$abstract <- removeWords(c$abstract, stopwords("english"))
# remove extra white space
c$abstract <- stripWhitespace(c$abstract)
# remove leading and trailing spaces
c$abstract <- gsub("^\\s+|\\s+$", "", c$abstract)

# remove rows where abstract is null
c <- c[which(c$abstract != ""),]
```

```
dim(c) # Rows and Columns after cleaning
```

```
## [1] 4433 27
```

```
statwords <- strsplit(c$abstract, "\\s+")
chunks.v <- unlist(lapply(statwords, length))
# Summary statistics of word count in the abstracts
summary(chunks.v)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.00   60.00   76.00   80.48   97.00  565.00
```

```

removeRows <- c()

# Take a look at short abstracts
for (i in 1:20) {
  w <- which(chunks.v == i)
  print(i)
  print(c[w,]$abstract)
  if (i < 8) {
    removeRows <- c(removeRows, w)
  }
}

```

```

## [1] 1
## [1] "absrtact" "abstract" "abstract" "abstract" "abstract" "abstract"
## [7] "abstract" "abstract" "abstract"
## [1] 2
## [1] "abstract available" "editors introduction" "1473 8376"
## [1] 3
## [1] "10 1177 1052562911411156"
## [1] 4
## character(0)
## [1] 5
## [1] "url http www econchallenge org"
## [2] "url http atlss umwblogs org"
## [3] "business ethics ouyang runping reviewed"
## [4] "business ethics zhou zucheng reviewed"
## [5] "business ethics xu dajian reviewed"
## [1] 6
## [1] "introduction business ethics zhang yinghang reviewed"
## [2] "contemporary business ethics su yong reviewed"
## [1] 7
## [1] "business ethics ethical business robert audi reviewed"
## [2] "ethical executive robert hoyk paul hersey reviewed"
## [3] "best practices designing effective ethics programmes reviewed"
## [4] "student response system srs often referred clicker"
## [1] 8
## [1] "getting wrong sign empirical work common phenomenon remarkably"
## [2] "introduction business ethics chen bingfu zhou zucheng reviewed"
## [3] "introduction business ethics 3rd edition joseph desjardins reviewed"
## [4] "business ethics real world approach andrew ghillyer reviewed"
## [1] 9
## [1] "url http sites google com site jgilberteconomics home excel"
## [2] "monty hall three door let s make deal game"
## [3] "term dilemma loosely used describe many challenges entrepreneurship pedagogy"
## [4] "moral issues business william h shaw vincent barry reviewed"
## [5] "understanding business ethics peter stanwick sarah d stanwick reviewed"
## [6] "moral leader challenges insights tools sandra j sucher reviewed"
## [1] 10
## [1] "entrepreneurial orientation associated high performance firms circumstances covin slevin 1989"
## [2] "story success five steps mastering ethics business leigh hafrey reviewed"
## [3] "american business values global perspective 6th edition gerald cavanagh reviewed"
## [4] "copy instructor notes can obtained request david oneill senatehall com"
## [1] 11

```

```

## [1] "survey current state education conducted kpmg education committee american accounting associati
## [2] "trend u s undergraduate economics degrees continued upward trajectory 2008 9"
## [3] "trend u s undergraduate economics degrees continued upward trajectory 2009 10"
## [4] "teaching moral leader literature based leadership course sandra j sucher reviewed"
## [5] "case studies business ethics edited al gini alexei m marcoux reviewed"
## [1] 12
## [1] "essentials business ethics creating organization high integrity superior performance dennis col
## [2] "business society stakeholders ethics public policy anne t lawrence james weber reviewed"
## [3] "business ethics case studies selected readings 7th edition marianne m jennings reviewed"
## [4] "hot seating form creative drama participants play imagine someone else s position"
## [5] "brain rules 12 principles surviving thriving work home school john medina reviewed"
## [1] 13
## [1] "case based real life scenario names dates company information disguised protect identity compan
## [2] "abstract author uses online securities market engage students exploration asset pricing microeco
## [3] "abstract review professional writings economists used passages plots characters themes ideas li
## [4] "faculty executive mba program emba investigating change emba curriculum motivation change provi
## [5] "analysing ethical codes uk professional bodies andy friedman sasha daly ruth andrzejewska review
## [6] "board ethical ceos create honest corporations patrizi porrini lorene hiris gina poncini reviewe
## [7] "article presents phaseouts table tax educational tool table compiles summarizes phaseouts limit
## [1] 14
## [1] "authors present simplified stripped version poker instructional classroom game although stripped
## [2] "present study sought contribute little existing literature fallan 2006 martin bartol 1986 e l"
## [3] "abstract learn nine feedback techniques can enhance teaching skills simultaneously raising stud
## [4] "managing business ethics straight talk right 5th edition l k trevino k nelson reviewed"
## [1] 15
## [1] "trend u s undergraduate economics degrees resumed upward trajectory fastest growth among privat
## [2] "trend u s undergraduate economics degrees continued upward trajectory 2010 2011 growing average
## [3] "article argues stand alone ethics course accounting details shortfalls questionable approach te
## [4] "leadership change responsibility joop remme stephanie jones beatrice van der heijden silvio de l
## [5] "paper describes particular design mandatory course business ethics mba students norwegian schoo
## [6] "paper identifies seven ways firms strategically use antitrust laws interests seven abuses illustr
## [1] 16
## [1] "charles crawford looks theories behind behaviour always ethical think might growing policy trend
## [1] 17
## [1] "associate editor online section journal economic education briefly describes successful submissi
## [1] 18
## [1] "author presents simple proof property method least squares variously known fwl frisch waugh love
## [2] "author describes senior level course designed promote student skills acting like economists alti
## [3] "develop creative innovative internet projects use electronic commerce vehicle teach entrepreneur
## [4] "short teaching note introduces visual exhibit help accounting students understand learn sfas 15
## [1] 19
## [1] "hanna davis team hired sports unlimited help revise markdown policies order increase store rever
## [2] "article examines use historical travel advertisements teaching tools students asked evaluate ad
## [1] 20
## [1] "skills os ability creative think critically recognize opportunities work effectively teams high

```

```

# Remove goofy abstracts.
c <- c[-removeRows,]
dim(c) # Rows and Columns after short abstract removal

```

```

## [1] 4409    27

```

```
# Now look at stats again.
statwords <- strsplit(c$abstract, "\\s+")
chunks.v <- unlist(lapply(statwords, length))
summary(chunks.v)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       8.0   60.0   76.0   80.9   97.0  565.0
```

Next, look at word frequency of the 11 disciplines