API setups and results for some public classifiers and taggers of unstructured text data

Alireza Ahadi

2023-02-20

1. Text analytics use cases

1.1 Sentiment analysis using uClassifer

uClassifier https://www.uclassify.com/ is a free machine learning web service where you can easily create and use text classifiers. Here is the code to perform sentiment analysis on a given text data, in this case simply a string:

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.1     v purrr   1.0.1
## v tibble  3.1.8     v dplyr   1.1.0
## v tidyr   1.3.0     v stringr 1.5.0
## v readr   2.1.4     v forcats 1.0.0
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
uClassifyURL = function(key, classifierName, text) {
    text = gsub(" ", "+", text)
    url = sprintf("https://api.uclassify.com/v1/uClassify/%s/classify/?readKey=%s&text=%s", classifierNa
    return(url)
  }

text = "I'm positive that my job search will lead me to the job that I truly deserve."
url = uClassifyURL('iGwAytFWYvvZ', 'sentiment', text)
raw = httr::GET(url)
content = rawToChar(raw$content)
content
```

```
## [1] "{\"negative\":0.28856,\"positive\":0.71144}"
```

1.2 Content classification using IAB-Taxonomy

Here we use the IAB_Taxonomy, which is available in uClassify API services:

```
library(tidyverse)

uClassifyURL = function(key, classifierName, text) {
    text = gsub(" ", "+", text)
    url = sprintf("https://api.uclassify.com/v1/uClassify/%s/classify/?readKey=%s&text=%s", classifierNa
    return(url)
  }


text = "I'm positive that my job search will lead me to the job that I truly deserve."
url = uClassifyURL('iGwAytFWYvvZ', 'IAB-Taxonomy-V2', text)
raw = httr::GET(url)
content = rawToChar(raw$content)


data = rjson::fromJSON(content)
data[199]
```

```
## $`hobbies and interests_board games and puzzles_11_8_1`
## [1] 3.12798e-07
```

2. O*NET Occupational Information

The Occupational Information Network (**O*NET**) is a free online database that contains hundreds of job definitions to help students, job seekers, businesses and workforce development professionals to understand today's world of work in the United States. Here I demo how you can get information about an occupation from ONET APIs. We use Bioscientist as an occupation, and get updated information about this occupation from O*NET:

```
library(ONETr)
```

```
## Loading required package: XML
```

```
## Loading required package: RCurl
```

```
##
## Attaching package: 'RCurl'
```

```
## The following object is masked from 'package:tidyr':
##
##     complete
```

```
## Loading required package: plyr
```

```
## ------------------------------------------------------------------------------
```

```
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
```

```
## ------------------------------------------------------------------------

##
## Attaching package: 'plyr'

## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize

## The following object is masked from 'package:purrr':
##
##     compact
```

```
setCreds("uts_edu_au","2393xyk")
```

```
## API credentials saved. You may now use package functions.
```

```
#Bioscientist: 15-2041.01
Bioscientist <- socSearch("15-2041.01")

skills <- skills(Bioscientist)
skills$name
```

```
##  [1] "Mathematics"                  "Active Learning"
##  [3] "Complex Problem Solving"      "Critical Thinking"
##  [5] "Judgment and Decision Making" "Reading Comprehension"
##  [7] "Science"                      "Speaking"
##  [9] "Active Listening"             "Writing"
```

```
skills$description
```

```
##  [1] "Using mathematics to solve problems."
##  [2] "Understanding the implications of new information for both current and future problem-solving a
##  [3] "Identifying complex problems and reviewing related information to develop and evaluate options
##  [4] "Using logic and reasoning to identify the strengths and weaknesses of alternative solutions, co
##  [5] "Considering the relative costs and benefits of potential actions to choose the most appropriate
##  [6] "Understanding written sentences and paragraphs in work-related documents."
##  [7] "Using scientific rules and methods to solve problems."
##  [8] "Talking to others to convey information effectively."
##  [9] "Giving full attention to what other people are saying, taking time to understand the points be
## [10] "Communicating effectively in writing as appropriate for the needs of the audience."
```

```
abilities <- abilities(Bioscientist)
abilities$name
```

```
##  [1] "Inductive Reasoning"    "Mathematical Reasoning" "Deductive Reasoning"
##  [4] "Oral Expression"        "Written Comprehension"  "Oral Comprehension"
##  [7] "Information Ordering"   "Problem Sensitivity"    "Speech Clarity"
## [10] "Written Expression"
```

```
abilities$description
```

```
##  [1] "The ability to combine pieces of information to form general rules or conclusions (includes fi
##  [2] "The ability to choose the right mathematical methods or formulas to solve a problem."
##  [3] "The ability to apply general rules to specific problems to produce answers that make sense."
##  [4] "The ability to communicate information and ideas in speaking so others will understand."
##  [5] "The ability to read and understand information and ideas presented in writing."
##  [6] "The ability to listen to and understand information and ideas presented through spoken words a
##  [7] "The ability to arrange things or actions in a certain order or pattern according to a specific
##  [8] "The ability to tell when something is wrong or is likely to go wrong. It does not involve solv
##  [9] "The ability to speak clearly so others can understand you."
## [10] "The ability to communicate information and ideas in writing so others will understand."
```

```
knowledge <- knowledge(Bioscientist)
knowledge$name
```

```
##  [1] "Mathematics"                   "English Language"
##  [3] "Computers and Electronics"     "Medicine and Dentistry"
##  [5] "Biology"                       "Education and Training"
##  [7] "Customer and Personal Service" "Administration and Management"
##  [9] "Administrative"                "Personnel and Human Resources"
```

```
knowledge$description
```

```
##  [1] "Knowledge of arithmetic, algebra, geometry, calculus, statistics, and their applications."
##  [2] "Knowledge of the structure and content of the English language including the meaning and spell
##  [3] "Knowledge of circuit boards, processors, chips, electronic equipment, and computer hardware an
##  [4] "Knowledge of the information and techniques needed to diagnose and treat human injuries, disea
##  [5] "Knowledge of plant and animal organisms, their tissues, cells, functions, interdependencies, a
##  [6] "Knowledge of principles and methods for curriculum and training design, teaching and instructi
##  [7] "Knowledge of principles and processes for providing customer and personal services. This inclu
##  [8] "Knowledge of business and management principles involved in strategic planning, resource alloc
##  [9] "Knowledge of administrative and office procedures and systems such as word processing, managing
## [10] "Knowledge of principles and procedures for personnel recruitment, selection, training, compens
```

```
tasks <- tasks(Bioscientist)
tasks$statement
```

```
##  [1] "Draw conclusions or make predictions, based on data summaries or statistical analyses."
##  [2] "Analyze clinical or survey data, using statistical approaches such as longitudinal analysis, m
##  [3] "Write detailed analysis plans and descriptions of analyses and findings for research protocols
##  [4] "Calculate sample size requirements for clinical studies."
##  [5] "Read current literature, attend meetings or conferences, and talk with colleagues to keep abre
##  [6] "Design research studies in collaboration with physicians, life scientists, or other professiona
##  [7] "Prepare tables and graphs to present clinical data or results."
##  [8] "Write program code to analyze data with statistical analysis software."
##  [9] "Provide biostatistical consultation to clients or colleagues."
## [10] "Review clinical or other medical research protocols and recommend appropriate statistical analy
```

```
education(Bioscientist)
```

```
##        .id              name score.text        score..attrs.scale
## 1 category   Master's degree         58 Percentage of Respondents
## 2 category   Doctoral degree         29 Percentage of Respondents
## 3 category Bachelor's degree         13 Percentage of Respondents
```

```
interests(Bioscientist)
```

```
##       .id          name
## 1 element Investigative
## 2 element  Conventional
## 3 element     Realistic
## 4 element      Artistic
## 5 element        Social
## 6 element  Enterprising
##
## 1
## 2
## 3 Realistic occupations frequently involve work activities that include practical, hands-on problems
## 4
## 5
## 6
##   score.text    score..attrs.scale score..attrs.important .attrs.id
## 1         95 Occupational Interest                   true   1.B.1.b
## 2         72 Occupational Interest                   true   1.B.1.f
## 3         33 Occupational Interest                  false   1.B.1.a
## 4         17 Occupational Interest                  false   1.B.1.c
## 5         11 Occupational Interest                  false   1.B.1.d
## 6         0  Occupational Interest                  false   1.B.1.e
##                                                                                .attrs.related
## 1 https://services.onetcenter.org/ws/online/occupations/15-2041.01/related/interests/1.B.1.b
## 2 https://services.onetcenter.org/ws/online/occupations/15-2041.01/related/interests/1.B.1.f
## 3 https://services.onetcenter.org/ws/online/occupations/15-2041.01/related/interests/1.B.1.a
## 4 https://services.onetcenter.org/ws/online/occupations/15-2041.01/related/interests/1.B.1.c
## 5 https://services.onetcenter.org/ws/online/occupations/15-2041.01/related/interests/1.B.1.d
## 6 https://services.onetcenter.org/ws/online/occupations/15-2041.01/related/interests/1.B.1.e
```

```
jobTitles(Bioscientist)
```

```
##     .id                          1
## 1 title              Biometrician
## 2 title Biostatistical Consultant
## 3 title           Biostatistician
## 4 title        Research Scientist
## 5 title     Statistical Scientist
```

```
technology <- technology(Bioscientist)
technology$title.text
```

```
##  [1] NA
```

```
##  [2] "Analytical or scientific software"
##  [3] "Presentation software"
##  [4] "Data base user interface and query software"
##  [5] "Operating system software"
##  [6] "Object or component oriented development software"
##  [7] "Spreadsheet software"
##  [8] "Web platform development software"
##  [9] "Medical software"
## [10] "Graphics or photo imaging software"
## [11] "Development environment software"
```

```
technology$example.text
```

```
##  [1] NA                              "The MathWorks MATLAB"
##  [3] "Microsoft PowerPoint"          "Structured query language SQL"
##  [5] "UNIX"                          "R"
##  [7] "Microsoft Excel"               "PHP"
##  [9] NA                              NA
## [11] "Microsoft Visual Studio"
```

```
tools <- tools(Bioscientist)
tools$title.text
```

```
## [1] "Scanners"            "Mainframe computers" "Notebook computers"
## [4] "Laser printers"      "Plotter printers"    "Desktop computers"
## [7] "Personal computers"
```

```
tools$example
```

```
## [1] "Computer data input scanners" NA
## [3] "Laptop computers"             "Computer laser printers"
## [5] "Plotters"                     NA
## [7] NA
```

```
workActivities <- workActivities(Bioscientist)
workActivities$name
```

```
##  [1] "Analyzing Data or Information"
##  [2] "Working with Computers"
##  [3] "Communicating with Supervisors, Peers, or Subordinates"
##  [4] "Making Decisions and Solving Problems"
##  [5] "Interpreting the Meaning of Information for Others"
##  [6] "Processing Information"
##  [7] "Updating and Using Relevant Knowledge"
##  [8] "Getting Information"
##  [9] "Organizing, Planning, and Prioritizing Work"
## [10] "Providing Consultation and Advice to Others"
```

```
workActivities$description
```

```
##  [1] "Identifying the underlying principles, reasons, or facts of information by breaking down inform
##  [2] "Using computers and computer systems (including hardware and software) to program, write softwa
##  [3] "Providing information to supervisors, co-workers, and subordinates by telephone, in written fo
##  [4] "Analyzing information and evaluating results to choose the best solution and solve problems."
##  [5] "Translating or explaining what information means and how it can be used."
##  [6] "Compiling, coding, categorizing, calculating, tabulating, auditing, or verifying information o
##  [7] "Keeping up-to-date technically and applying new knowledge to your job."
##  [8] "Observing, receiving, and otherwise obtaining information from all relevant sources."
##  [9] "Developing specific goals and plans to prioritize, organize, and accomplish your work."
## [10] "Providing guidance and expert advice to management or other groups on technical, systems-, or p
```