

Introduction to the **RKEA** Package

Ingo Feinerer

July 9, 2008

Abstract

A short introduction to the **RKEA** package.

Introduction

The **RKEA** package provides a R interface to KEA¹, a tool for keyword extraction in texts.

Loading the Package

Before actually working we need to load the package:

```
> library("RKEA")
```

Creating a Keyword Extraction Model

KEA needs a keyword extraction model for keyword extraction. You can build your own models by manually indexing the keywords in a small set of texts, and then call `createModel()`.

```
> library("tm")
> data(crude)
> keywords <- list(c("Diamond", "crude oil", "price"),
+                 c("OPEC", "oil", "price"),
+                 c("Texaco", "oil", "price", "decrease"),
+                 c("Marathon Petroleum", "crude", "decrease"),
+                 c("Houston Oil", "revenues", "decrease"),
+                 c("Kuwait", "OPEC", "quota"))
> tmpdir <- tempfile()
> dir.create(tmpdir)
> model <- file.path(tmpdir, "crudeModel")
> createModel(crude[1:6], keywords, model)
```

Please note that we just wrap the functionality of the original KEA program which always uses files for in- and output (and that is the reason you also need to use a directory in R as shown in the above example). We deliberately decided not

¹<http://www.nzdl.org/Kea/>

to modify the KEA java archive shipped with this R package for compatibility reasons. However this may induce some warnings in R (e.g., because some internal KEA paths might not be available) but nevertheless you should get the full functionality out of it.

Keyword Extraction

Once you have a KEA model you can extract keywords from texts.

```
> extractKeywords(crude, model)

[[1]]
 [1] "Diamond"           "cut its contract" "cut its"
 [4] "cut"               "its contract"    "reduction"
 [7] "contract"         "crude oil"      "dlrs a barrel"
[10] "prices"

[[2]]
 [1] "OPEC"              "problems"        "production"
 [4] "market's earlier" "is"              "be"
 [7] "analysts"         "not"            "but"
[10] "He"

[[3]]
 [1] "Texaco Canada"           "Texaco"
 [3] "Canada"                 "it lowered the contract price"
 [5] "crude oil"              "crude"
 [7] "price"                  "it"
 [9] "the"                    "for"

[[4]]
 [1] "it reduced the contract price" "reduced"
 [3] "grades of"                    "grades of crude"
 [5] "grades"                       "of crude"
 [7] "crude"                        "price"
 [9] "of"                            "it"

[[5]]
 [1] "trust's future"           "estimates the"
 [3] "trust's future net"      "reserves"
 [5] "estimates"              "Trust"
 [7] "Trust said"             "study"
 [9] "trust's future net revenues" "net"

[[6]]
 [1] "Kuwait"      "OPEC"      "He"      "bpd"      "has"
 [6] "Minister"    "Sheikh"    "Sheikh Ali" "Ali"      "members"

[[7]]
 [1] "says"           "Indonesia"    "economy"    "government"
```

[5] "report says" "report" "appears" "appears to be"
[9] "appears to" "measures"

[[8]]

[1] "yesterday's higher" "yesterday's higher levels"
[3] "riyal" "deposits were"
[5] "deposits" "yesterday's higher levels in"
[7] "in halting the current decline" "were"
[9] "at" "OPEC"

[[9]]

[1] "billion" "budget" "billion riyals"
[4] "riyals" "government" "Sheikh Abdul-Aziz"
[7] "Abdul-Aziz" "expenditure" "year's decline"
[10] "the government"

[[10]]

[1] "Saudi" "accord"
[3] "Nazer" "kingdom's commitment"
[5] "December's OPEC" "kingdom's commitment to"
[7] "December's OPEC accord" "SPA"
[9] "free" "free market"

[[11]]

[1] "Saudi" "exports" "Ju'aymah"
[4] "the sources" "from" "the sources said"
[7] "in January" "January" "Ju'aymah terminals"
[10] "bpd"

[[12]]

[1] "the official" "official" "oil ministers" "ministers"
[5] "Gulf" "Arab" "states" "Emirates"
[9] "crude oil" "of"

[[13]]

[1] "Saudi" "kingdom's commitment"
[3] "December's OPEC" "kingdom's commitment to"
[5] "December's OPEC accord" "Nazer"
[7] "kingdom's commitment to last" "December's OPEC accord to"
[9] "accord" "OPEC"

[[14]]

[1] "OPEC" "meeting" "oil prices" "that" "in"
[6] "prices" "oil" "a" "for" "the"

[[15]]

[1] "power" "The port" "port" "closed" "when" "after"
[7] "lines" "was" "of" "The"

[[16]]

```
[1] "group"                "strategic"
[3] "reserve to"            "one mln barrels as"
[5] "mln barrels"          "oil prices on the domestic"
[7] "reserve"              "present"
[9] "U.S"                  "as"
```

```
[[17]]
```

```
[1] "group"                "strategic"
[3] "reserve to"            "one mln barrels as"
[5] "mln barrels"          "oil prices on the domestic"
[7] "to raise the strategic reserve" "reserve"
[9] "U.S"                  "study"
```

```
[[18]]
```

```
[1] "Corp's Union"          "Corp's Union Oil Co"
[3] "Corp's Union Oil"     "Union"
[5] "the eastern region of" "lowered"
[7] "its posted"           "its posted prices"
[9] "its posted prices for" "posted prices for"
```

```
[[19]]
```

```
[1] "NYMEX"                "Exchange"
[3] "futures"              "a futures"
[5] "transaction"          "according to"
[7] "according"            "change"
[9] "hold a futures position" "hold a"
```

```
[[20]]
```

```
[1] "January"
[2] "in January"
[3] "mln barrels"
[4] "pct"
[5] "mln"
[6] "in"
[7] "Yacimientos Petroliferos Fiscales"
[8] "Yacimientos Petroliferos"
[9] "Yacimientos"
[10] "Petroliferos"
```

```
> unlink(tmpdir, recursive = TRUE)
```