Marking open-ended questions wuth Topic modelling and LDA

Text analysis/topic modelling

- Open-ended questions questions are more eficient/complete assessments.
- But are also harder to mark.
  - we want to use here text/topic analyis/minning...
- Text in `R` means a object `string`
  - so we are doing string manipulation
- *Moodle*, R/exams and open-ended questions...

Text processing and analysis

- String manipulation
- general pattern analysis
- key terms
- grouping ansewrs

String manipulation

```
# answers <- read.csv2("respostas_IC_conclusao-interpretacao.xml").
# head(answers)
```

word cloud

*for an overview on the answers

must have words

- to get marks for key words/sentences... and lose marks as well...

Latent Dirichlet Allocation (dicriminat analysis...) to mark `string` text answers

- to group students by similarities... to check for copying, cheating, ...

https://rpubs.com/Argaadya/topic_lda

https://rstudio-pubs-static.s3.amazonaws.com/100936_ae4251bba3da438bbfb870d4b9431f04.html

papers

https://scholar.harvard.edu/files/dtingley/files/topicmodelsopenendedexperiments.pdf

https://www.tandfonline.com/doi/full/10.1080/2573234X.2019.1590131

file:///C:/Users/Markus/Downloads/information-11-00550.pdf

References

- An Introduction to Text Processing and Analysis with R
  - https://m-clark.github.io/text-analysis-with-R/string-theory.html#basic-text-functionality
- Text Processing in R
  - http://www.mjdenny.com/Text_Processing_In_R.html