AUTOMATIC LABELLING OF TOPICS IN UNIVERSITY SUBJECTS TO DETECT WHICH TOPICS ARE MORE DIFFICULT TO LEARN

J. Martinez-Romo, J. Gomez-Martin, L. Plaza, L. Araujo, F. Lopez-Ostenero

Grupo de INnovación Docente en Estructuras de Datos y Algoritmos

Departamento de Lenguajes y Sistemas Informáticos

Universidad Nacional de Educación a Distancia (UNED)

Index

Context

The **UNED**, with more than **205,000 students**, has the largest student population in Spain and is one of the largest universities in Europe

Since the teaching is at **distance**, one of the objectives of the University is to **analyze the results** obtained by the students and to **reinforce** those aspects of the subjects that are more difficult

This work is included in a project in which the results obtained by the students in the tests of the degrees in **Computer Science** are analyzed to detect those **topics that are more difficult**

Objetives

METHODOLOGY BASED ON ARTIFICIAL INTELLIGENCE TECHNIQUES TO AUTOMATICALLY IDENTIFY THE ELEMENTS THAT ARE MOST COMPLICATED FOR STUDENTS TO LEARN

EXTRACT AND STATISTICALLY ANALYZE THE DIFFERENT TOPICS THAT ARE PART OF A COURSE

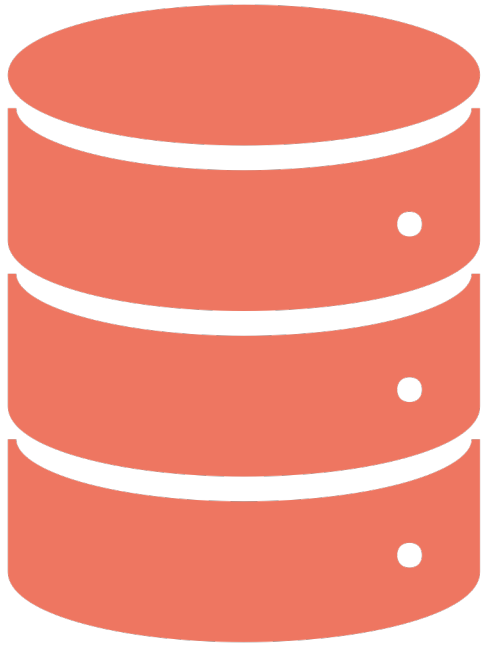
OPTIMIZE THE IMPROVEMENT OBTAINED WITH THE PRODUCTION OF NEW MATERIALS, FOCUSING ON THOSE ELEMENTS THAT ARE HARDER FOR THE STUDENTS TO LEARN

Methodology

- A fundamental part of the work is the preparation of a corpus of manually annotated exams to later evaluate the performance of the automatic topic annotation system
- This part requires, first, establishing the set of indicators or labels that will be assigned to the exam questions to characterize them
- Among the labels considered are the topic(s) of the subject to which the question relates, but also other aspects such as necessary prior knowledge not part of the subject

System developed

- System based on a generative model, that automatically extracts a set of topics from the exams of a subject.
- LDA (Latent Dirichlet Allocation) model
- The system would not need the manual establishment of labels on each of the examination questions

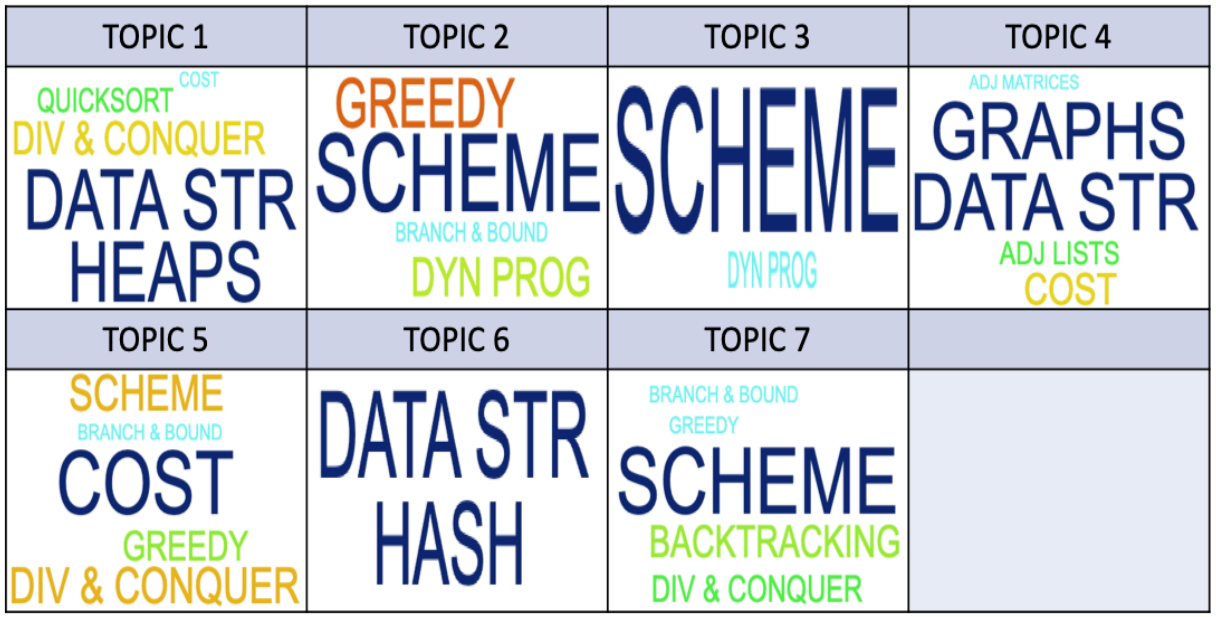


Results

Evaluation of the optimal number of topics according to the coherence score.

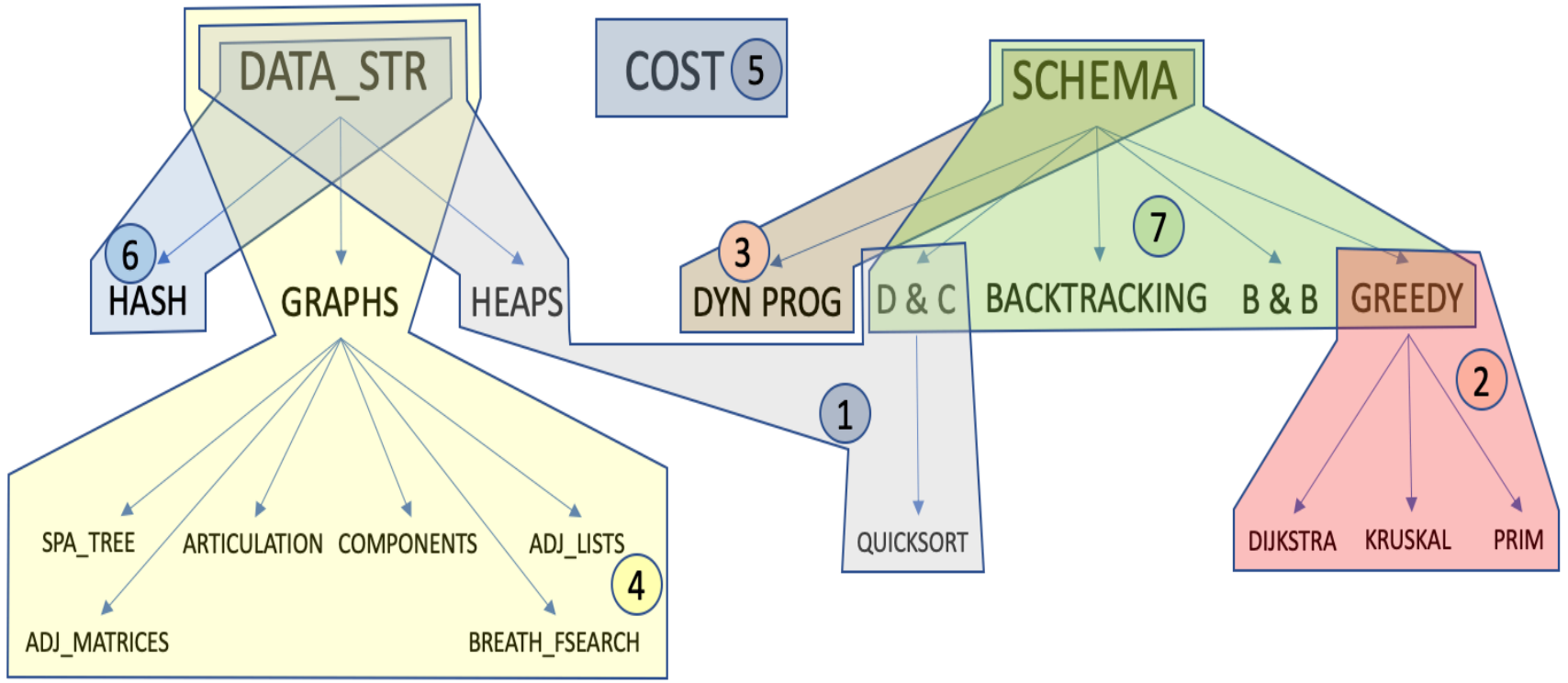
Results

- Topics extracted and represented each by a cloud of words formed by the labels that make up each topic



Results

- Label hierarchy and how the different topics (each represented by a color) affect each of the parts of this hierarchy



Conclusions

This work has shown a new methodology for the automatic identification of topics on exam questions from a university subject

This methodology automatically identifies the main topics covered during the course and allows for a statistical analysis of the results of evaluation tests

The system presented in this work detects mostly the topic of the syllabus to which each question corresponds, obtaining a great correlation with the topics assigned by the teachers

These results make it possible to generalize the automatic detection of topics in university subjects

Thank you for your interest