Self-Assessment tool with topic-driven navigation for algorithms learning

INEDA Group: Teaching innovation in data structure and algorithms

Fernando López-Ostenero, Laura Plaza, Lourdes Araujo y Juan Martínez-Romo

Motivation

1 **Popularity of online learning**

Which is gaining increasing attention, specially in higher education

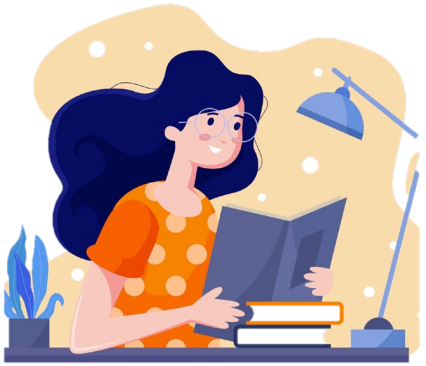
2 **The UNED University**

Which has distinct characteristics that make the learning more challenging

3 **The subject: Algorithms and data structures**

Which is crucial in the Computer Science programs and have an important impact on success

The boom of online learning



- ✓ Resources are available anywhere, any time
- ✓ Learners may study at their own pace and speed
- ✓ Helps to achieve equality and inclusion
- ✓ Facilitates lifelong learning
- ✓ Allows for higher level of personalization

- ✓ Overload of information
- ✓ Sentiment of loneliness
- ✓ Lack of motivation

The University

70 % of students are older than 30 y.o.

Most have a full time job and family duties

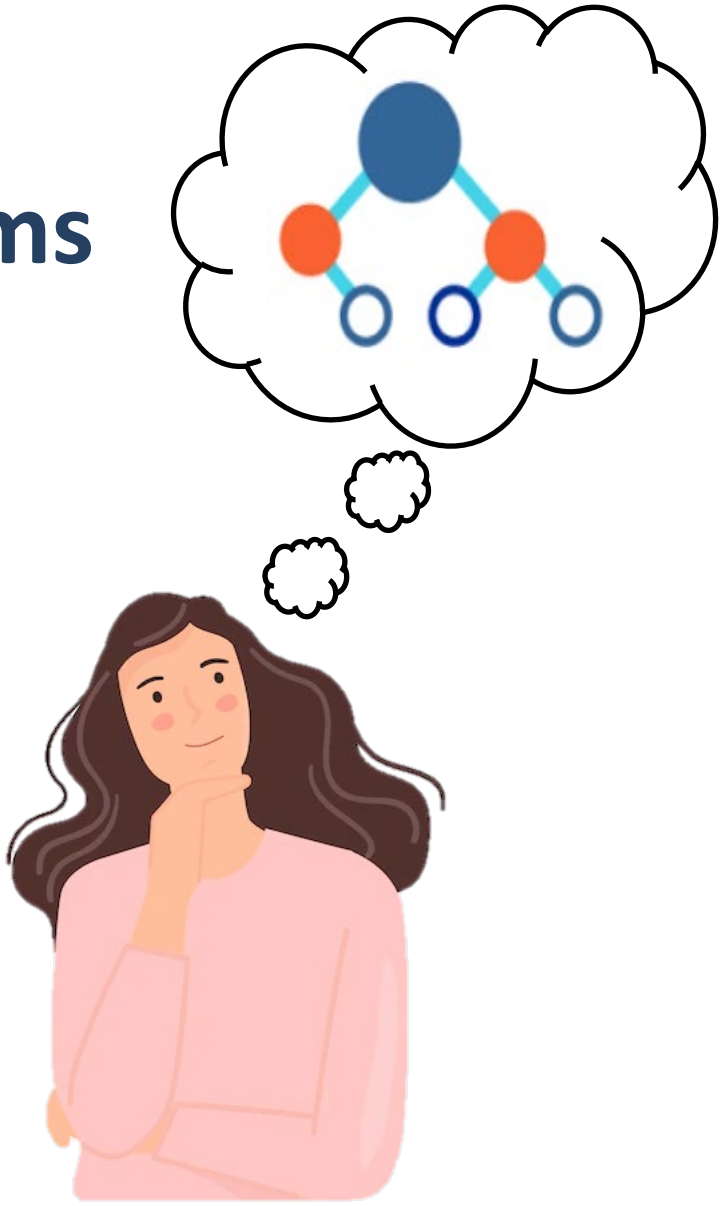
Significant number of disabled students

Very high drop-out rate

The subject

Data Structures and Algorithms

- ✓ Second year of CS undergraduate programs
- ✓ Given a problem, the best algorithmic scheme and data structures have to be selected to design the solution
- ✓ Crucial part of the CS program, but it includes topics of high difficulty



Motivation

- ✓ Online learning scenario
- ✓ Our students, with full-time jobs and family duties, have limited time to study
- ✓ The subject is difficult, requires a lot of training because of the need of reasoning, “intuition” and abstraction, where complex concepts are built upon simple ones

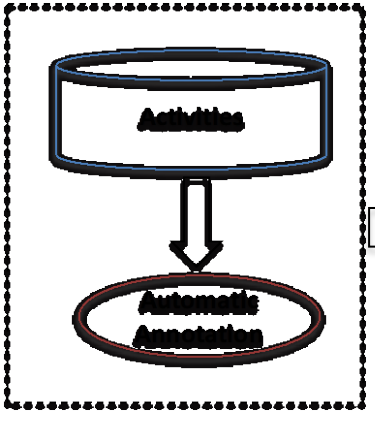
Development of motivational strategies and tools to reduce drop out

Self-assessment web tool

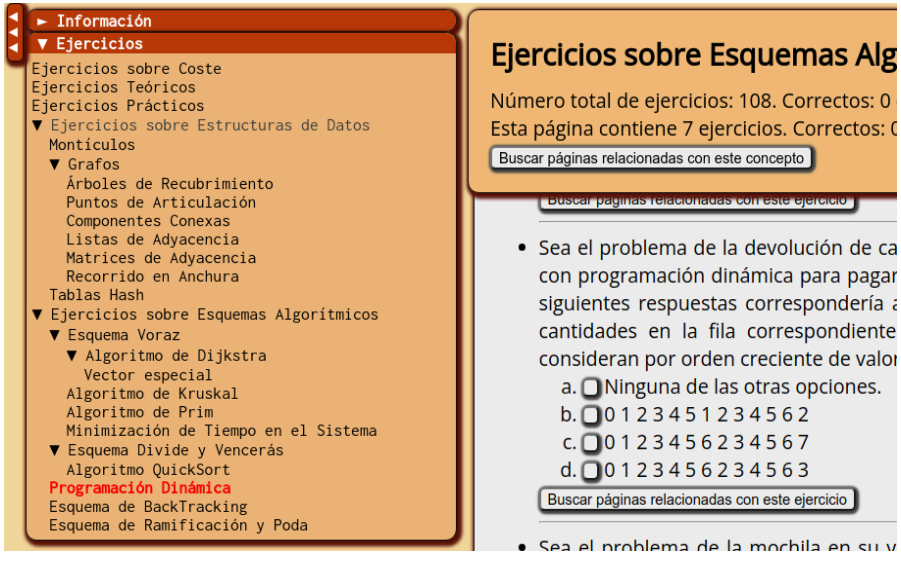
1. Visualizing the main concepts/topics as a hierarchy to facilitate understanding of the relations between them
2. Providing self-assessment activities for each topic to ensure the learning of the different concepts
3. Recommending reinforcement material from the web
4. Allowing practicing from anywhere and any device

Use of NLP techniques to automatically annotate and categorize activities into topics, so that it is easy to expand the tool and to transfer to other subjects

Development process




```
<QDB>
  ...
  <TEST date="2017-F-1S">
    ...
    <QUESTION n="2">
      <TEXT>Sea el problema de la devolución de
           cambio con monedas de valores 1,6 y 10
           solucionado con programación dinámica
           para pagar una cantidad de 12 unidades.
           Identifica cuál de las siguientes
           respuestas correspondería al contenido de
            la tabla de resultados parciales de
           cantidades en la fila correspondiente a
           la moneda de valor 6, si dichas monedas
           se consideran por orden creciente de
           valores:</TEXT>
      <LABELS>ESQUEMA,PD,PRACTICO</LABELS>
      <KEYPHRASES>monedas|valores|moneda de valor|
           respuestas|cantidad|contenido|fila
           correspondiente|tabla|programacion
           dinamica|orden|valor|problema</KEYPHRASES
           >
      <OPTION l="a" v="F">0 1 2 3 4 5 6 2 3 4 5 6
           3</OPTION>
      <OPTION l="b" v="F">0 1 2 3 4 5 6 2 3 4 5 6
           7</OPTION>
      <OPTION l="c" v="T">0 1 2 3 4 5 1 2 3 4 5 6
           2</OPTION>
      <OPTION l="d" v="F">Ninguna de las otras
           opciones.</OPTION>
    </QUESTION>
    ...
  </TEST>
  ...
</QDB>
```



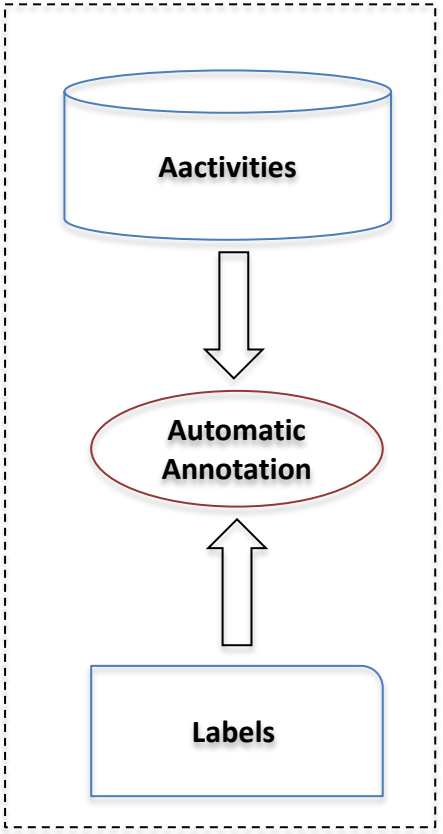
RECOMMENDATION

Step I:
Topic extraction and organization

- DATA STRUCTURES
  - HEAPS
    - HEAPSORT
  - GRAPHS
    - ADJACENCY MATRICES
    - ADJACENCY LISTS
    - SPANNING TREES
    - COMPONENTS
    - ARTICULATION
  - HASH
- ALGORITHM SCHEMES
  - SCHEME SELECTION
  - GREEDY
    - PRIM
    - KRUSKAL
    - DIJKSTRA
  - DIVIDE&CONQUER
    - MERGESORT
    - QUICKSORT
  - DYNAMIC PROGRAMMING
  - BACKTRACKING
  - BRANCH&BOUND
- COST

- Extraction of the main topics of the subject
- Organization as a hierarchy of concepts

Step II: Automatic labelling of activities



- Manual automatic annotation is a very time-consuming task
- We apply NLP to automatize this task

* We use as activities the questions from the exams from the past academic years

Extracting key phrases

- Words are converted to lower case, diacritics are eliminated, empty words (determiners, prepositions, etc.) are removed
- Words are tagged with their POS using the Penn Treebank POS tags
- A set of patterns is used to identify the key phrases

```
A B* C*
A = VB CC JJ | NN CC VB | VI
      NN* JJ*
B = IN NN NN* JJ* | CC NN*
C = JJ* RB* | CC VB | CC NN
```

Programación dinámica

Selecting of the most representative key phrases

- We use TF-IDF to select the most representative key phrases
  - Term Frequency counts how many times the key phrase appears in the actitivity
  - Inverse Document Frequency measures the number of questions in the database that contain the key phrase

$$TF\text{-}IDF(k,q) = tf(k,q) * idf(k)$$

$$idf(k) = \log(n / df(k)) + 1$$

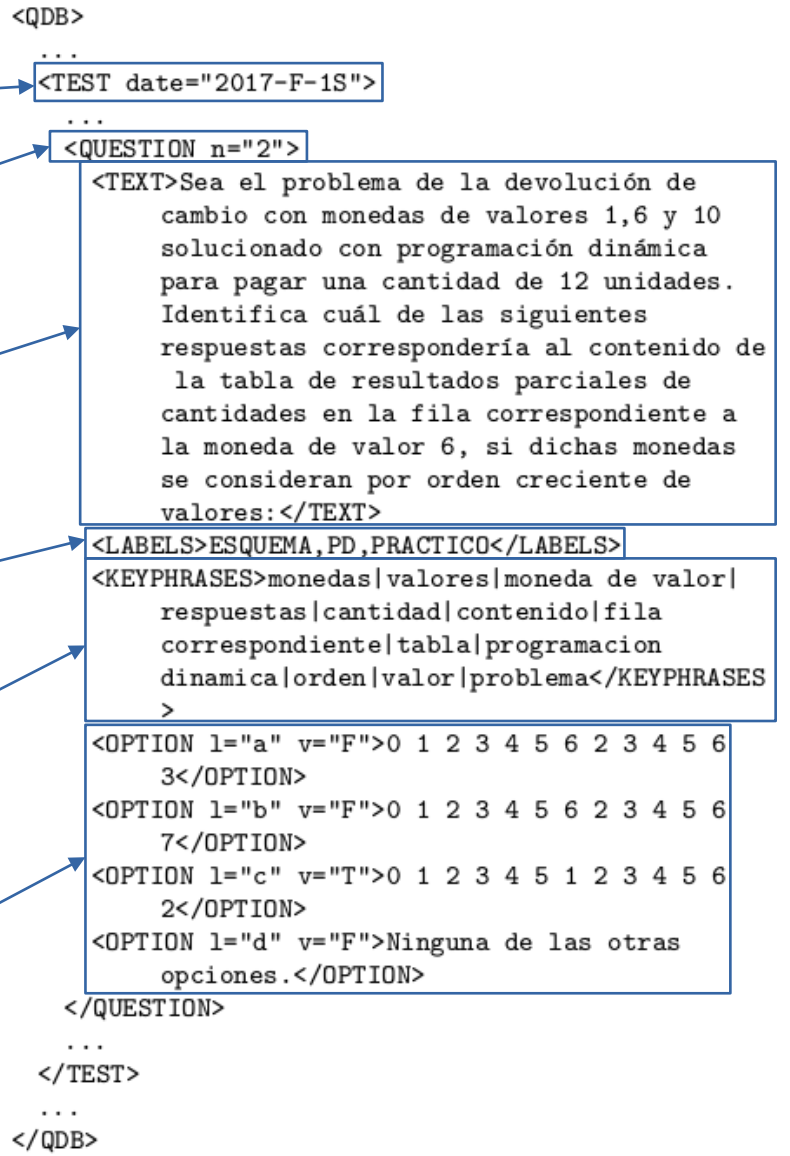
Labelling activities with topics

- We manually asign the most representative key phrases to the different topics in the hierarchy

- We automatically label each question in the database with the corresponding topics:
  - For each topic and question, we check if they have any key phrase in common
  - If so, both the topic and its ancestors in the hierarchy are assigned to the question

Step III: XML annotation scheme

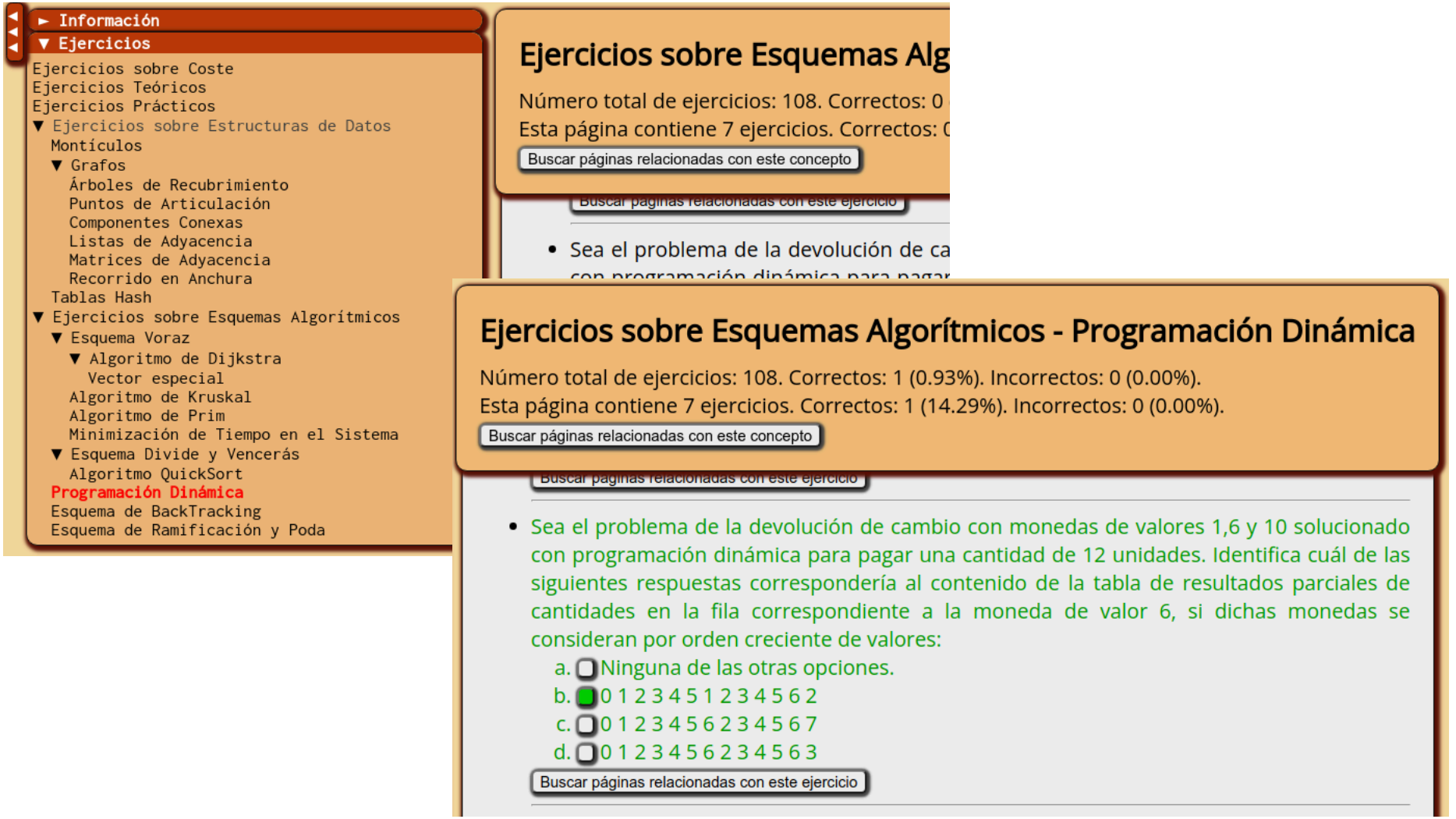
- TEST where the activity was used
- ID of the activity
- TEXT of the activity
- Labels of the TOPICS related to the activity
- Key phrases extracted from the activity
- Text, labels and truth value of the ANSWERS



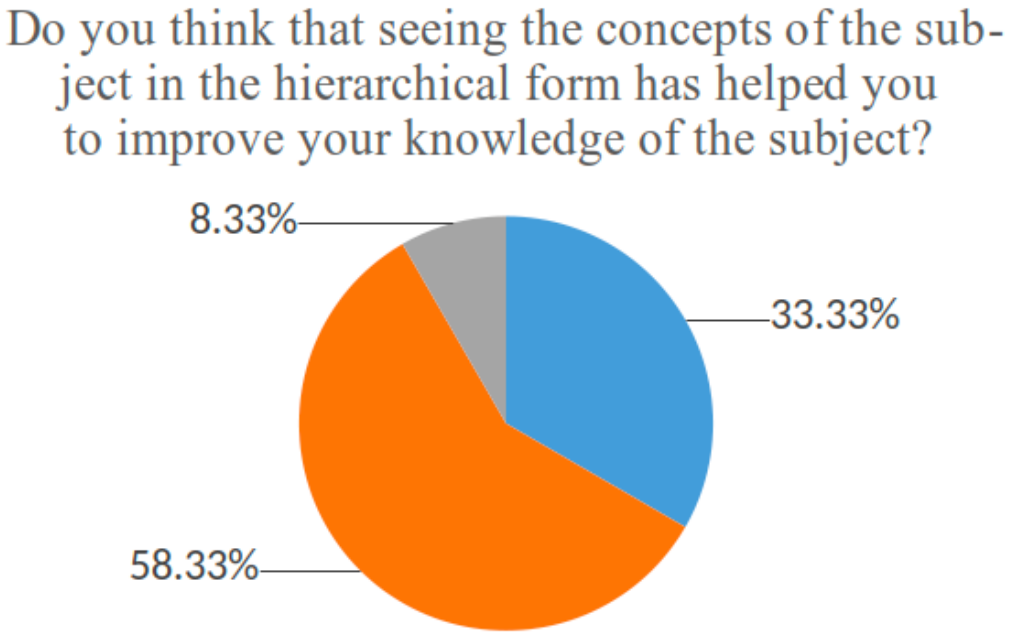
Step IV: Web recommender

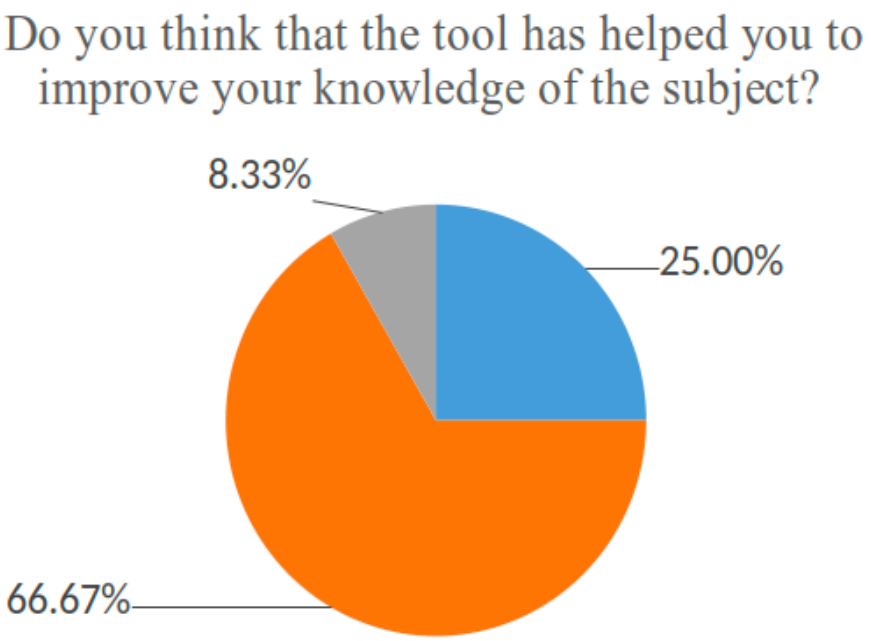
- Recommend reinforcement material for both specific activities and general topics
- Given an activity,
  - a query is built with the key phrases associated to it, joined using an OR operator
- Given a topic,
  - the top-5 most frequent key phrases extracted from all activities under the topic are used as query

Step V: The web tool



Evaluation

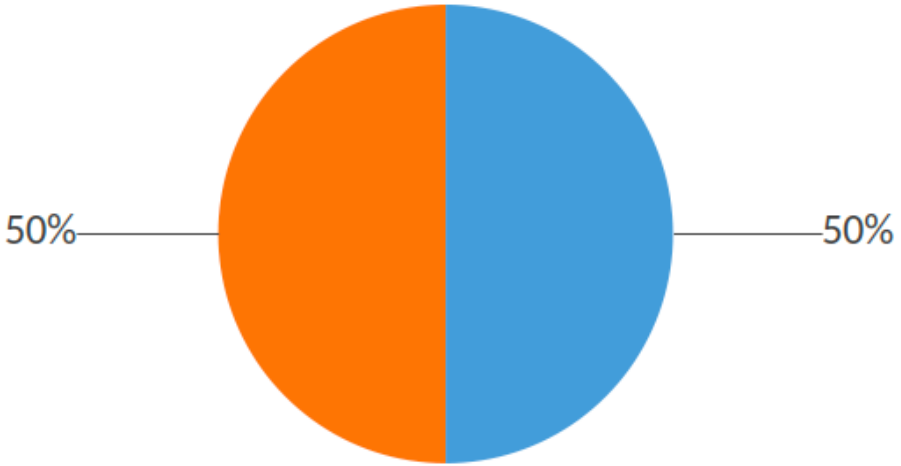




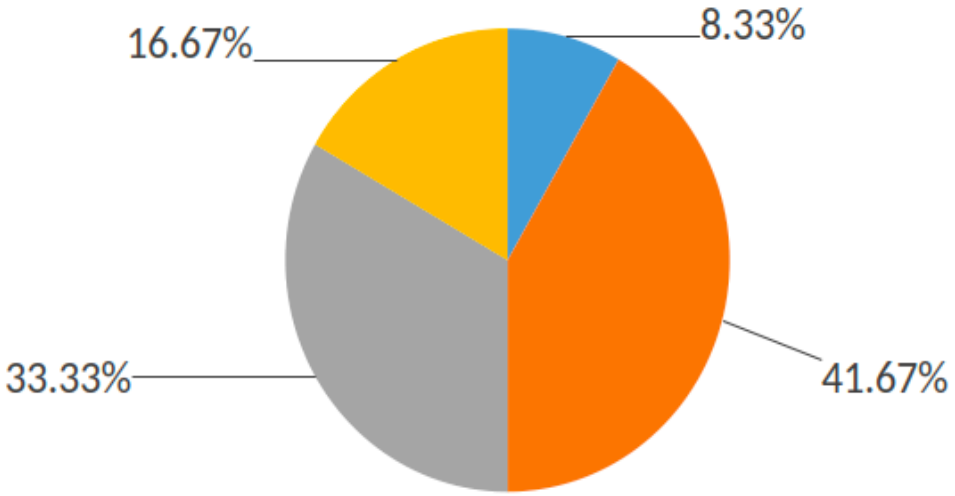
Very much Quite a lot A little Not at all

Evaluation

Did you find it easy to use?

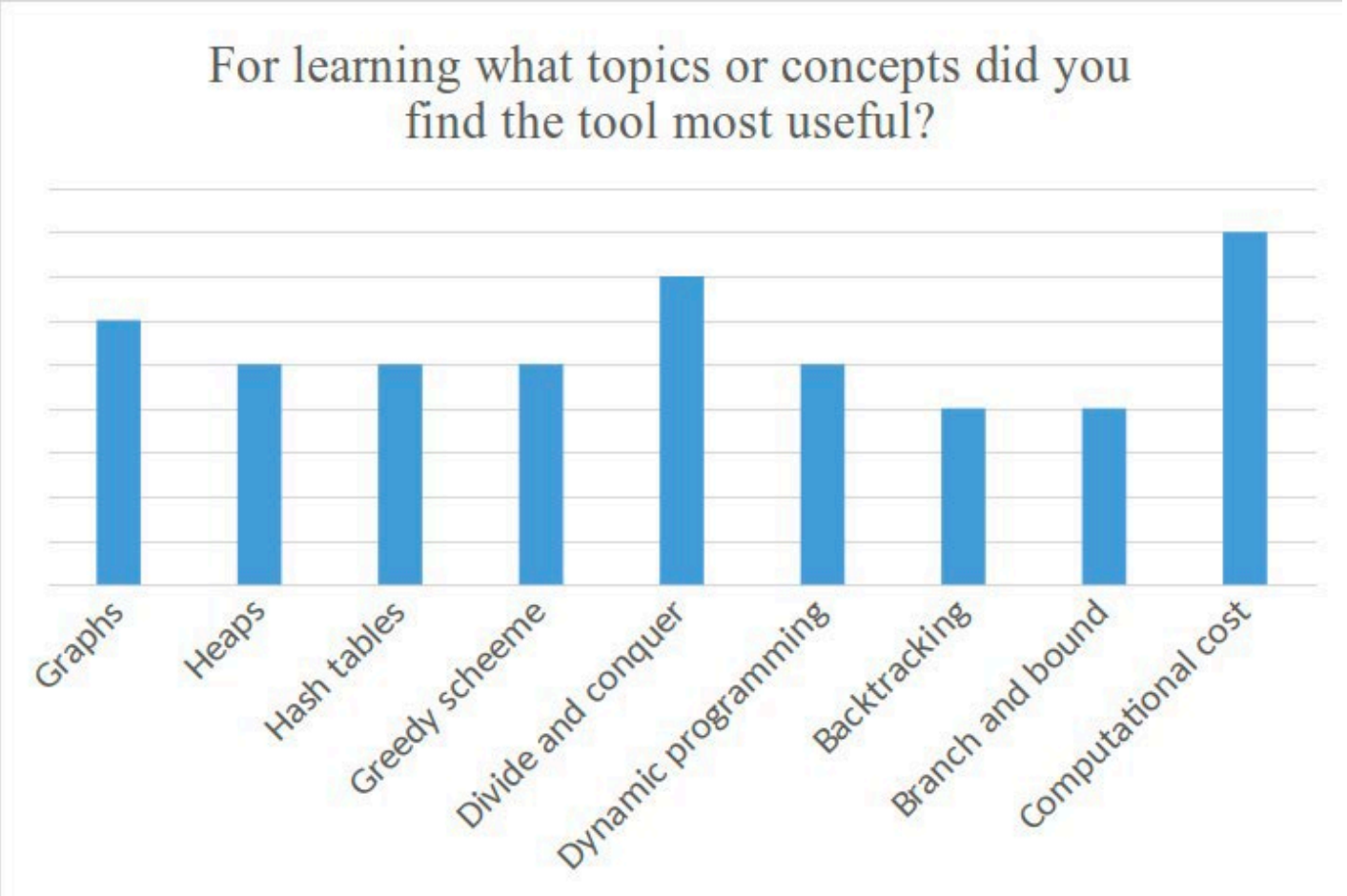


If you have used the utility to obtain more information from Internet, do you think that the information obtained has been useful?



Very much Quite a lot A little Not at all

Evaluation


For learning what topics or concepts did you find the tool most useful?
Graphs
Heaps
Hash tables
Greedy scheeme
Divide and conquer
Dynamic programming
Backtracking
Branch and bound
Computational cost

Conclusions

- We have presented a tool that combines navigation through the hierarchy of concepts of a subject with self-assessment activities
- The tool has had an excellent acceptance among students, who consider the tool to be highly useful for studying the subject and easy to use
- The tool is specially useful to study the most difficult topics

Future Work

- To organize questions in each topic by levels of difficulty
- To record the history of each student so that recommendation of activities may be personalized
- To implement the tool for other subjects

Thank you very much for your attention