

Проектна Задача 3

Тема:

Креирање и анализа на податочно множество со прашања и одговори од доменот на ПМ

Цели:

- ✓ оформување колекција на прашања и одговори користени на тестови за PMP сертификација
- ✓ примена врз повеќе формати како веб страни, табели, документи во PDF или EPUB
- ✓ примена на алгоритми за обработка и прочистување на текстот
- ✓ квантитативна анализа со приложување на статистики (табеларно и со графикони)

Испораки:

- ❖ документ со опис на методологија и што е сработено
- ❖ папка со оригинални документи и листа со линкови од веб страни
- ❖ рамномерно дистрибуирано податочно множество за прашања одговори
- ❖ код и тетратки со коментари

Корисни линкови:

- CommonsenseQA <https://aclanthology.org/N19-1421/>
- OpenBookQA <https://aclanthology.org/D18-1260/>
- MedQA <https://www.mdpi.com/2076-3417/11/14/6421>

Дополнителни инфо:

1. Прашањата се од тестови за полагање **PMP – Project Management Professional** Certification од PMI – **Project Management Institute** кои се комбинација од едноставни и комплексни т.н. ситуациони прашања со приложен опис и избор од 4 понудени можни одговори
2. Пример за прашање од тест со избор на 4 можни одговори:

You are a project manager of a complex project which involves many procurement contracts with subcontractors. During the execution phase, a subcontractor comes to you and says that he has completed his part of the work. You reviewed the facts and provided the subcontractor with a formal written notice that the contract has ended. What process is this?

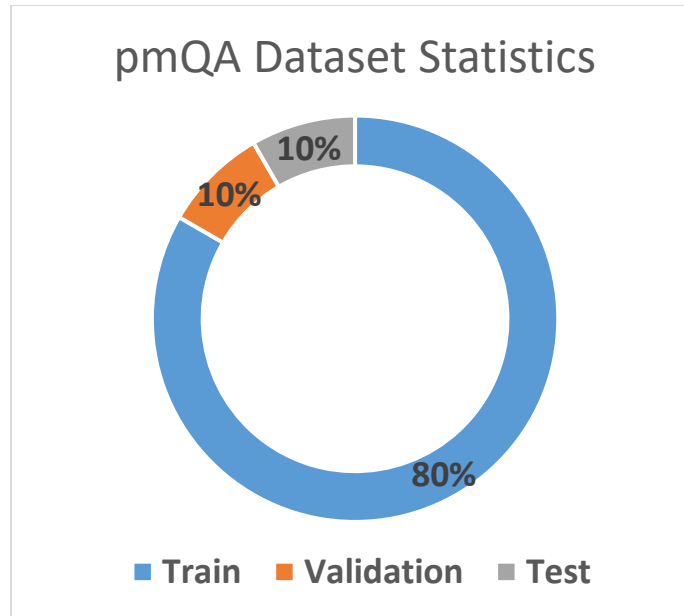
- (a) Validate scope
- (b) Verify scope
- (c) Close procurements
- (d) Control procurements

Answer-125: d

Explanation: In the control procurements process, the buyer issues a formal written notice that the contract has been completed.

Reference: The PMBOK Guide, 6th Edition, Page: 499

3. Материјали
 - a. Папка со PDF документи за тестови и припреми за полагање со прашања и одговори
 - b. Табела со листа на веб линкови што водат до веб страни со прашања и одговори кои треба да се извлечат (web scraping)
 - c. Примерок од податочно множество со прашања и одговори во .jsonl формат
4. Примери со чекори за прочистување на текстот (да се предложат и други чекори):
 - a. Текстот да биде само на англиски јазик, други јазици да се отфрлат
 - b. Отстранување на header, footer, и број на страна од документите (доколку има) за да не се појават во текстот
 - c. Отстранување на веб линкови и референци
 - d. Отстранување на специјални знаци, математички формули, празни места, итн.
 - e. Пополнување на празните места _____ FITB (Fill In The Blanks)
 - f. Отстранување на дупликат прашања
5. Квантитативна статистичка анализа на ртQA со визуелизација:
 - a. присуство на сврзувачки термини (на пр. и, но)
 - b. присуство на термини за негација (на пр. не, никогаш)
 - c. присуство на услови за заштита (на пример, понекогаш, можеби)
6. Рамномерно дистрибуирано податочно множество за прашања одговори



7. Структура и формат на податочно множество pmQA

Податочните под-множества да бидат испорачани во .jsonl формат {train,dev,test}.jsonl (<https://jsonlines.org/>) каде што секој примерок (една линија) е запишан како JSON. Во продолжение има пример за JSON структурата на едно прашање со 4 можни одговори:

```
{"source":  
  "CertChamp",  
  "PMBOK": 6,  
  "year": 2019,  
  "type": "Final Mock Test",  
  "qid": "3",  
  "question": {"stem": "Which of the following statements are not true regarding Network Diagrams?:"},  
  "choices": [  
    {"label": "A", "text": "Network Diagrams determine risks and causes of issues as they occur."},  
    {"label": "B", "text": "Network diagrams are used to show interdependecnies of all activities."},  
    {"label": "C", "text": "Network diagrams depict the workflow so everyone knows the sequence in which activities need to be performed"},  
    {"label": "D", "text": "They help plan , organize and control the project and can help track schedule progress"}  
  ],  
  "answerKey": "A",
```

"explanation": "The correct answer is A - All the other statements are valid regarding Network Diagrams .",

"statements": [

{ "label": true, "statement": "Which of the following statements are not true regarding Network Diagrams?:
Network Diagrams determine risks and causes of issues as they occur." },

{ "label": false, "statement": "Which of the following statements are not true regarding Network Diagrams?:
Network diagrams are used to show interdependecnies of all activities." },

{ "label": false, "statement": "Which of the following statements are not true regarding Network Diagrams?:
Network diagrams are used to show interdependecnies of all activities." },

{ "label": false, "statement": "Which of the following statements are not true regarding Network Diagrams?: They
help plan , organize and control the project and can help track schedule progress" }

]

}