

NLP - Project 1

Due Date: **15th October**

Project Limitations

- Can only use *regex* library from Python 3.10

General Outline



1. **Analyze three publicly available crime novels/stories by famous crime novelists that will be assigned to our team shortly. You can find novels at the Project Gutenberg site: <http://www.gutenberg.org/>.**
2. **Ask ChatRegex the questions listed under **Prompts**, and produce precise results and in natural, well-structured English.**



To effectively conduct this analysis, you should find resources and read the plot summaries of each novel so you can make your search more effective. If plot summaries are unavailable, use regex to search for clues and report how well/fast that approach worked.



Do not hardcode exact strings for prompts. Allow for prompts asking "Describe when the ..." and direct questions "When the ..." or more personal "Tell me when the ...".

Crime novels

- The Sign of the Four by Arthur Conan Doyle
- The Murder on the Links by Agatha Christie
- The Man in the Brown Suit by Agatha Christie

Prompts

1. When does the investigator (or a pair) occur for the first time - chapter #, the sentence(s) # in a chapter
2. When is the crime first mentioned - the type of the crime and the details - chapter #, the sentence(s) # in a chapter

3. When is the perpetrator first mentioned - chapter #, the sentence(s) # in a chapter
4. What are the three words that occur around the perpetrator on each mention (i.e., the three words preceding and the three words following the mention of a perpetrator)
5. When and how the detective/detectives and the perpetrators co-occur - chapter #, the sentence(s) # in a chapter
6. When are other suspects first introduced - chapter #, the sentence(s) # in a chapter

Project checklist

- ☐ Source code
- ☐ 2-page long report (maximum), excluding references
- ☐ 10 min results presentation + 2 min Q&A