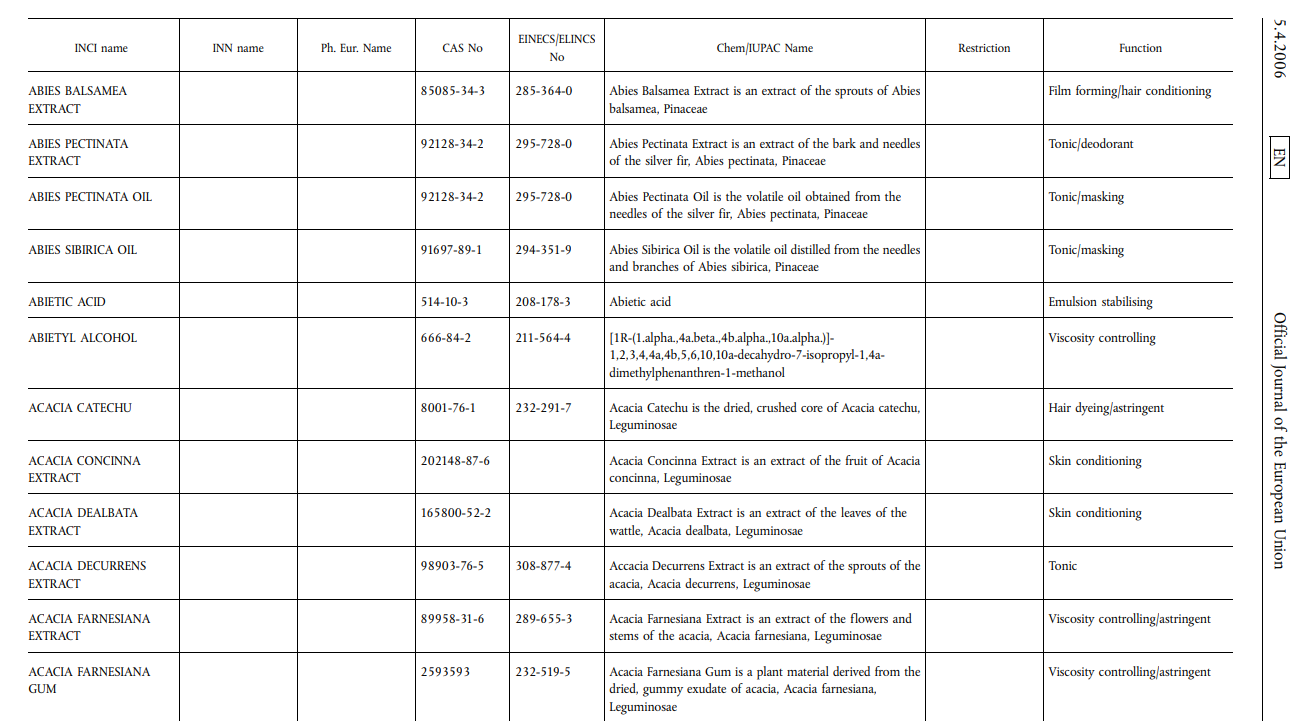
Project Weekly Progress Report Agile – Scrum

|  |  |
| --- | --- |
| **Semester** | **W2023, SEM-2** |
| **Course Code** | **AML-2404** |
| **Section** | **Section 2** |
| **Group Name** | **D** |
| **Student names/Student IDs** | **Jash Vaghasiya - C0884733**  **Nivedini Kathagonda - C0872720**  **Keval Parmar - C0882386**  **Monil Rupawala - C0882370**  **Sai Divya Madhuri Guntupalli - C0882360** |
| **Reporting Week** | **7** |
| **Team Lead for the reporting week** | **Monil Rupawala** |

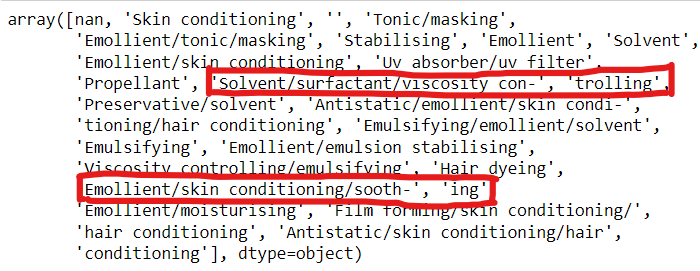
**1. Progress Made in Reporting** **Week:**

The preprocessing stage of the data received from PDF documents of European skincare items saw substantial advancements during week 8 of the project. Due to the mismatch between our recommendation system and intended recommendations, we chose to use this resource. We used the Camelot library to extract text from the PDFs as part of the data preprocessing. Row by row, the extraction procedure produced separate rows for constituents with descriptions spanning more than two lines. This development has created the framework for more data analysis and improving our recommendation system.



**2. Difficulties Encountered in Reporting Week:**

However, dealing with the problem of separating ingredient information into different rows is now proving difficult. Camelot's text extraction process resulted in other rows for ingredients on numerous lines, which caused the data to be represented in pieces. This makes it challenging to match the correct elements with the relevant goods. We investigate various options, including text processing methods and NLP tools, to detect and reconstruct the fragmented ingredient information reliably. It is imperative to address this issue to guarantee the integrity and dependability of the data used for the recommendation system.



The ingredient information must also be included in the recommendation system's later phases and the data preparation step. It affects the feature extraction process since splitting the ingredients across numerous rows makes extracting significant characteristics difficult. It also makes it harder to provide recommendations that accurately match user input with the right products. To manage the fragmented ingredient data effectively and guarantee the development of reliable suggestions for consumers, the solution to this challenge necessitates thorough analysis and experimentation with various methodologies and algorithms.