**Práctica Unidad 1 - Extracción y Procesamiento de Texto**

**Etapa 1 (Consolidación de repo compartido)**

1. Según los grupos conformados en relación al juego cada grupo deberá

generar un repositorio en **drive** donde guardar los recursos obtenidos a continuación enumerados. Este repositorio deberá ser compartido por el grupo con los profesores de la cátedra en el rol de editor.

[jpmanson@gmail.com](mailto:jpmanson@gmail.com)

[alan.geary.b@gmail.com](mailto:alan.geary.b@gmail.com)

[constantinoferrucci@gmail.com](mailto:constantinoferrucci@gmail.com)

[dolores.sollberger@gmail.com](mailto:dolores.sollberger@gmail.com)

**(nota: en las siguientes tareas considerar la extracción de recursos en varios idiomas)**

* Extraer texto de documentos tipo pdf, txt, word o formatos de imágenes con las librerías adecuadas que se consideren relevantes.
* Extraer texto de videos de reviews y tutoriales.
* Evaluar diferentes tutoriales teniendo en cuenta que existen en varios idiomas.
* Extraer texto de foros con Web Scraping ([ejemplo](https://colab.research.google.com/drive/1cxJ365KUzlTEpwLMMXSLN7Fa1oGHRZVB?usp=sharing)).
* Extraer estadísticas. La idea es que generen un **formato tabular** de información que crean relevante sobre datos del juego.
* Extraer relaciones entre juegos, creadores y otros sujetos implicados ([ejemplo](https://boardgamegeek.com/boardgame/414317/harmonies/credits) en la sección créditos).

1. Generar un documento “**readme.txt**” en el repositorio, que justifique la decisión de los datos recolectados e indique, a un potencial usuario futuro, en las características relevantes de los recursos extraídos para cada ítem. Esta tarea se debe continuar en las prácticas realizadas en próximas unidades.