* Your goal
* The data source and web technology it is based upon
* The approach/technology that you will use for downloading
* How you plan to clean the data (scraping)
* How you would analyze the extrcated information

L’objectiu és obtenir informació de la pàgina web

https://www.gamespot.com/articles/game-release-dates-2016/1100-6442373/

del jocs editats durant l’any 2016. Volem crear una base de dades agregada de diferents aspectes per a cada joc (nom del joc, editora del joc, genere del joc, nota de la revista, nota dels usuaris, opinió dels usuaris, consoles per a les que ha sortit cada joc,...).

Per a escrapejar la web, farem servir la llibreria rvest i probablement necessitarem expressions regulars per a netejar les dades, a fi d’obtenir la base de dades final. La idea es recórrer tots els links a les pagines pròpies dels jocs des de la llista per obtenir més informació, com notes, opinions o preus.

De cara a l’anàlisi, farem un anàlisi descriptiu de les dades, i mirarem diverses associacions com per exemple la correlació existent entre la nota de la revista i la nota dels usuaris.

Roger Borràs

Jose Segura