

Task Description

1. Run [ReVerb](#) on 5 documents from each web site, camera reviews, and 2 forum sites in the weapons domain.
2. For each document, list the extractions and score them as useful or useless. (A useful extraction is one we could use to build a knowledge graph, use your judgment.)
3. Make a note of information that you think should have been extracted, but wasn't.

Preparation

Five documents from websites for “camera reviews”

1. [Nikon's New D5 and D500 Push the Boundaries of DSLR](#) at [dpreview](#)
2. [Canon 5DS R Review](#) at [imaging-resource](#)
3. [Nikon D750](#) at [pcmag](#)
4. [Sony Alpha A7R II Digital Camera Review](#) at [reviewed](#)
5. [Sony Cyber-shot RX10 review](#) at [engadget](#)

Two forum sites in the weapons domain

1. [Korth has finally made a PPC gun!](#) at forum of [calguns.net](#)
2. [Best pump shotgun ever?](#) at forum of [armslist](#)

Outputs

Raw Folder

Raw html files downloaded from websites, ReVerb can process these html files without considering HTML tags.

Output Folder

Outputs from ReVerb extraction, preprocessed for further using.

Processing Folder

Process outputs from ReVerb in Python code, which directly transfer preprocessed output from ReVerb into csv data.

Report Folder

Reports for each document, including extraction, score and notes.