**What is the seed URL(s)?**

Seed Url is …

For artists, my seed url is <https://americanart.si.edu/art/artists>.

For artworks, my seed url is <https://americanart.si.edu/art/browse?classification=All&mediums=All&location=All&on_view=0>

**How did you manage to only collect artwork/artist pages? How did you discard irrelevant pages?**

My seed url only contains information of artworks and artist, I go through every page continuously from seed urls and go into every link of artworks, when the number of collected webpage reach the requirement I stop the program