What is the seed URL(s) you used for each task?

Task1:

'https://www.rogerebert.com/collections/drama-movies',
'https://www.rogerebert.com/collections/comedy-movies',

Task2:

'https://www.metacritic.com/browse/movies/people/popular'

How did you manage to only collect movie or cast pages?

First I set-up allow domains to avoid crawling websites outside Metacritic or Rogerebert. Second I only crawl pages from index page about movie or cast. So the crawling pages are always the about movie or cast pages.

Did you need to discard irrelevant pages? If so, how?

I don't need to discard irrelevant pages because I use index page to parse movie or cast links, and crawl content from those links, so they should not include irrelevant pages.

Did you collect the required number of pages? If you were not able to do so, please describe and explain your issues.

Yes. Although Rogerebert contents limit movies relate to drama and comedy, I crawled both of them. So I had enough number of pages. Metacritic is just has enough cast pages for me to crawl.