**Code Review**

File Name: web.py  
Version Name: aa77c2f  
Location: team01-Project/src/web.py

Documentation:

Lines 13-14 can include more information on the limitations of the depth

Line 16 should include more context

Line 27 should read “that have already been checked”

Line 30 should read, “and add the obtained articles”

Line 27 should read “that have been added to the database”

Lines 42-48 have a good and an adequate description

Line 60 should include information on scenarios where the article could not be fetched

Lines 63-64 need more clarity

Line 76 should be “published date”

Lines 85-87 are good

Line 108 should be “then this page is ..  
Lines 147-148 more detail needed

Coding Style:

The code is clear and well structured; variables are used correctly.

Correctness:  
Line 45 attempts to crawl articles with the same depth; the program will continue to crawl for indefinitely long period of time – the depth should be reduced by one every time this function is called.

Line 10 provides a list of page types that are invalid, but there are potentially other types.

Testing  
There are only a few tests provided in test\_web\_crawler – testing different depths. Furthermore, only cnn.com seems to be tested and no testing related to adding references between articles and sources were included (despite crawl seemingly having this functionality).

Suggestions:

1. More tests can be provided to test different websites – the format of websites could vary by a large amount throughout different sites. Both local/national and international websites could be tested.
2. Specific types of invalid urls could be provided. Perhaps a page with just a video or sites that contain no links.
3. Tests involving references need to be added.

Summary:

Fix the points mentioned in documentation, correctness and testing; otherwise well-structured code.