1. Crawler.getLinks()

* Input: Document của jsoup
* Output: Tách links trong Document, chuẩn hóa, lưu vào Queue

1. Crawler.getLinks()
2. Input: Document của jsoup
3. Output: Tách links trong Document, chuẩn hóa, lưu vào Queue
4. Crawler.request()
   1. Input: url
   2. Output: Document jsoup
5. Frontier.shouldVisit()
   1. Input: url
   2. Output: True/False -> Kiểm tra xem có nên crawl url đó hay ko
6. Frontier.checkCrawledUrl()
   1. Input: url
   2. Output: True/false -> Kiểm tra url đã tồn tại trong hashset crawledUrl hay ko
7. Frontier.checkTraps()
   1. Input: url
   2. Output: True/false -> Kiểm tra url có bẫy hay ko
8. Frontier.filterUrl()
   1. Input: url
   2. Output: True/False -> Kiểm tra khi request url thì trả về loại file hợp lệ ko. Danh sách loại file cho phép tải (\*.mp3, \*.html…) được chứa trong Frontier.fileTypes
9. Storage.saveFile()
   1. Input: url, content (nội dung cần ghi), depth: Độ sâu của url
   2. Output: Lưu content ra file có tên là url (thay tất cả ‘/’ thành ‘.’), chứ trong thư mục tên ‘depth[depth]’.
   3. Quy định thư mục
      1. Thư mục ‘depth0’ chứa trang có depth = 0 và thư mục depth1
      2. Thư mục ‘depth1’ chứa trang có depth = 1 và thư mục depth2
      3. …