You will need to extract the link structure from the downloaded web pages in order to compute pagerank. A large number of tools are available for this task. Here we provide an example on how to use such tool to extract outgoing links from a downloaded html page.

We will use the open source jsoup library for this task. You can download or integrate it into your maven project following the instructions at https://jsoup.org/download.

Following is a sample code which processes a downloaded abc-main.html file to extract outgoing urls.

```
import org.jsoup.Jsoup;
import org.jsoup.nodes.Document;
import org.jsoup.nodes.Element;
import org.jsoup.select.Elements;

import java.io.File;

/**
 * Created by charith on 9/29/16.
 */
public class ExtractLinks {

    public static void main(String[] args) throws Exception{


        File file = new File("/home/charith/Desktop/CSCI572/simple-page/abc-main.html");
        Document doc = Jsoup.parse(file,"UTF-8", "http://abcnews.go.com/");

        Elements links = doc.select("a[href]");
        Elements media = doc.select("[src]");
        Elements imports = doc.select("link[href]");

        print("\nMedia: (%d)", media.size());
        for (Element src : media) {
            if (src.tagName().equals("img"))
                print(" * %s: <%s> %sx%s (%s)",
                        src.tagName(), src.attr("abs:src"), src.attr("width"), src.attr("heig
                        trim(src.attr("alt"), 20));
            else
                print(" * %s: <%s>", src.tagName(), src.attr("abs:src"));
        }

        print("\nImports: (%d)", imports.size());
        for (Element link : imports) {
            print(" * %s <%s> (%s)", link.tagName(),link.attr("abs:href"), link.attr("rel"));
        }

        print("\nLinks: (%d)", links.size());
        for (Element link : links) {
            print(" * a: <%s>  (%s)", link.attr("abs:href"), trim(link.text(), 35));
        }
    }


    private static void print(String msg, Object... args) {
        System.out.println(String.format(msg, args));
    }

    private static String trim(String s, int width) {
        if (s.length() > width)
            return s.substring(0, width-1) + ".";
        else
            return s;
    }

}
```

You will get an output like below:

Media: (72)
 * script: <http://abcnews.go.com/abc-main_files/bk-coretag.js>
 * script: <http://abcnews.go.com/abc-main_files/get>
 * script: <http://abcnews.go.com/abc-main_files/cb=gapi.loaded_0>
 * script: <http://abcnews.go.com/abc-main_files/vrs.js>
 * script: <http://abcnews.go.com/abc-main_files/analytics.js>
……

Media: (72)
 * script: <http://abcnews.go.com/abc-main_files/bk-coretag.js>
 * script: <http://abcnews.go.com/abc-main_files/get>
 * script: <http://abcnews.go.com/abc-main_files/cb=gapi.loaded_0>
 * script: <http://abcnews.go.com/abc-main_files/vrs.js>
 * script: <http://abcnews.go.com/abc-main_files/analytics.js>
…….

Links: (339)
* a: <http://abcnews.go.com/US/photos/nj-transit-train-crashes-hoboken-terminal-42446739>
(PHOTOS: NJ Transit Train Crashes I.)
 * a: <http://abcnews.go.com/US/photos/nj-transit-train-crashes-hoboken-terminal-42446739> ()
 * a:
<http://abcnews.go.com/US/passenger-describes-harrowing-scene-nj-transit-crash/story?id=424
46427> (Passengers Describe Harrowing Scen.)
 * a:
<http://abcnews.go.com/US/passenger-describes-harrowing-scene-nj-transit-crash/story?id=424
46427> ()
 * a: <http://abcnews.go.com/#> (Sections)
……..

Please refer to Jsoup API documentation for further reference: https://jsoup.org/apidocs/