Web spider to extract the content of hidden urls of website

In order to find out the solution process of given problem I spent alot more time on internet finally I reached the solution.

1)The problem was to handle dynamically loading the content of the html document without changing the url of the page. This took my maximum amount of time in comparison to others. Initially I found a solution to this problem, using with PhantomJs and BeautifulSoup but unfortunatly it didn't work with me.
http://kochi-coders.com/2014/05/06/scraping-a-javascript-enabled-web-page-using-beautiful-soup-and-phantomjs/
Then I found another blog which contained the that solution of this problem which worked with me finally. It was using the concept of 'Python Network Programming'. The section which contains the information 'Downloading Pages Through Form Submission ' with 'Mechanize' was helpful for me.
http://rhodesmill.org/brandon/chapters/screen-scraping/

2)The very first thing for me that what is the 'web crawler/spider' and how to build it? so I searched for it on the web and got some solutions from youtube, some pdfs and ofcourse stackoverflow etc.
https://www.youtube.com/watch?v=SFas42HBtMg
http://www.springer.com/?SGWID=4-102-45-113505-p28710520
many answers of questions asked stackoverflow.

3) Then I developed a simple web crawler which take a link from a frontier queue and after scrapping, extract all hrefs from html document and put these hrefs in to frontier queue after checking for duplication. And the links that I have already visited, stored them in a visited List. The inputs to this crawler are the original link as seed and number of depth till which you want to crawl. Then I thought something about global crawling and local based on netloc part which can be found by urllib2.urlsplit function. Since it is not required form me so I did it for only learning crawling with multiprocessing in pyhton. By this crawler, I was not able to crawl peginated content of hidden links.

4) After finding the solution of javascript forms submission through python scripts using mechanize then I started to build the crawler according to our requirements Initially I built it in such a way so that it can work with only the given website\seed. Then I tried to build for every page which is having the dynamic content change property. And tried on some other pages one of them is - http://data.fingal.ie/ViewDataSets/ but this crawler can't identify the class name of that div which contain the dynamic content so it parse the whole page and extract all the hrefs which are refering to some other pages. But If we need only the content of

dynamic division of the webpage then it is able to extract all the data you want but some other not required data such as the hrefs of that html document which are outside the paginated division.

5)Since my job is to make the crawler work for at least this given website http://trai.gov.in/Content/PressReleases.aspx so in the final version I restrict some functionality of the script and inserted the class name of paginated division of given website. Now at this stage the crawler works perfactly and able to download all the pdf file of pressRelease section of orignal website. And in the output, at runtime crawler shows the current status of ongoing downloads of each pdf file and after the successfuly running of crwaler we get a directory named 'PDFs' in the same directory where the script is. But when I change the classname of peginated division according to html document then It shows the result according to that website.

6)Further I want to change it in such a way such that it will be download all perticuler type of files in single running and put them in seprate directories according to their extension in the same directory where the script is. Development of this crwaler has gone through such processes so that it can be easily extendiable in future.