

Capire il codice

- Dalla riga 2 alla riga 7 vengono importati tutti i pacchetti per il corretto funzionamento del programma.
- Dalla riga 13 alla 15 viene ottenuto l'html della pagina.
 - Nella riga 13 viene assegnato alla variabile **url** il link del sito così da poterlo richiamare più facilmente in seguito.
 - Nella riga 14 si dà il comando per ottenere i dati grezzi dal sito.
 - Nella riga 15 i dati del sito importati vengono formattati in sintassi html, su questa sintassi si andrà a lavorare.
- Nella riga 22 si estraggono i 20 link dall'HTML della pagina.
- Dall riga 29 alla 38 si accede singolarmente ad ogni pagina contenente il txt tramite il corrispettivo link, se ne estraggono i dati e si salvano a loro volta in un nuovo formato txt sul proprio computer.
Più precisamente:
 - Nelle righe 30,34,38 sono stati inseriti dei print per controllare il funzionamento del programma.
I print sono stati inseriti al fine di avere un feedback sul funzionamento del programma dato che in questo esempio i dati estratti sono molti ed il programma su computer più lenti potrebbe sembrare bloccato mentre in realtà sta lavorando.
In questo modo si controlla che tutto proceda evitando riavvi inutili.
 - Nella riga 31 si compone il link della pagina su cui sono i dati, come comporlo è frutto di una osservazione empirica.
 - Nelle righe 31,32 si prende la pagina web e si scarica il suo contenuto, in questo caso sarà già in formato txt.