① CRAWL the web …

② Efficiently index the documents (D). How?

→ Store all the docs in a folder

Any other ideas?