IndexingWebPages

How to index a web site

Lucene doesn't directly support this, you need to use a spider like regain, SearchBlox or Nutch to accomplish this.

HTTrack is a useful, free spider with many features. Also see the Lucene FAQ