

weight focused irrelevant pages
optimization text relevance uses
file crawling fetched relevant
seed determine html used search table whether
crawlers link
relevancy queue updateable lsfc equation
determines updated
scoring url topic
classification crawl site naive
crawled links words system store
classifier threshold nb anchor harvest