



the Internet

URL filtering

regular expressions

deduplication

topic classification



database

top-down



database

experiment execution

experiment design

literature mining



problem statement

bottom-up