



the Internet

regular expressions

URL filtering

deduplication

topic classification



database

synthetic  
data

top-down

synthetic  
data



database

experiment execution

experiment design

literature mining

e.g. find a polymer with  
this mass distribution



problem statement

bottom-up