Deliverable 1- Basic web search using Nutch

Gaurang Patel
03/10/09
CS297
Deliverable 1

• Previously Error in path.
• /WEB-INF/classes/nutch-site.xml
  
  
  <property>
    <name>searcher.dir</name>
    <value>C:\nutch-0.9\nutch-0.9\crawl.sports.yahoo-5-1000</value>
  </property>

• Search using crawled results implemented
Clustering approach..

• Experiments with Carrot2
• carrot2-java-api-3.0
  – Examples
• Carrot2 workbench
Clustering approach..

• Carrot2 clustering plugin
• /WEB-INF/classes/nutch-default.xml
  – Set value of Property plugin.includes to clustering-carrot2