CSE508 – Assignment 1

Samarth Bharadwaj

Indexing process

1) Using build argument of galago script file
    a. Galago build –destpath -corpuspath
2) Indexing process creates two posting listing, one without stemming and other with stemming
3) Indexes are stored at a desired location

Index Size:

Wiki-small: 32.1 Mb
Wiki-large: 590 Mb

Time taken

Wiki-small: few minutes
Wiki-large: 4-5 hours

Query	Intent	Top 5 results	Remark
breakdown alfred	Show Alfred Hitchcock page	Alfred Hitchcock filmography - Wikipedia, the free encyclopedia Common Rider/Against All Authority Split - Wikipedia, the free encyclopedia Live at the Roxy WC2 - Wikipedia, the free encyclopedia Lake Alfred, Florida - Wikipedia, the free encyclopedia BV - Wikipedia, the free encyclopedia	Disregards other pages with alfred
breakdown director	Show Alfred Hitchcock page	Alfred Hitchcock filmography - Wikipedia, the free encyclopedia Common Rider/Against All Authority Split - Wikipedia, the free encyclopedia Apocalypse Now - Wikipedia, the free encyclopedia The Pit and the Pendulum (1961 film) - Wikipedia, the free encyclopedia Live at the Roxy WC2 - Wikipedia, the free encyclopedia	Finds the director's page correctly
vision computer **or** **computer vision**	Does order matter?	Point Distribution Model - Wikipedia, the free encyclopedia NUS School of Computing - Wikipedia, the free encyclopedia Manarat International University - Wikipedia, the free encyclopedia Neuros Technology - Wikipedia, the free encyclopedia Visions of Atlantis - Wikipedia, the free encyclopedia	Order is immaterial
Game sport	Disambiguation between computer games and sports	Sporting Praia Cruz - Wikipedia, the free encyclopedia Truman Sports Complex - Wikipedia, the free encyclopedia Jason Koumas - Wikipedia, the free encyclopedia Edwin Moses - Wikipedia, the free encyclopedia	Noun with 'sport' is selected
bowl game	Is NFL in top 5?	Fiesta Bowl - Wikipedia, the free encyclopedia East Carolina Pirates football - Wikipedia, the free encyclopedia National Football Conference West Division - Wikipedia, the free encyclopedia 2006 Purdue Boilermakers football team - Wikipedia, the free encyclopedia	Yes