16th Jan, 2012

## CSE508 – Assignment 1

Samarth Bharadwaj

### Indexing process

1. Using build argument of galago script file
   1. 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](http://localhost:50764/document?identifier=Common_Rider_Against_All_Authority_Split_0539)  [Live at the Roxy WC2 - Wikipedia, the free encyclopedia](http://localhost:50764/document?identifier=Live_at_the_Roxy_WC2_64a3)  [Lake Alfred, Florida - Wikipedia, the free encyclopedia](http://localhost:50764/document?identifier=Lake_Alfred,_Florida_c826)  BV - Wikipedia, the free encyclopedia | Disregards other pages with alfred |
| breakdown director | Show Alfred Hitchcock page | [Alfred Hitchcock filmography - Wikipedia, the free encyclopedia](http://localhost:50764/document?identifier=Alfred_Hitchcock_filmography_b0bf)  [Common Rider/Against All Authority Split - Wikipedia, the free encyclopedia](http://localhost:50764/document?identifier=Common_Rider_Against_All_Authority_Split_0539)  [Apocalypse Now - Wikipedia, the free encyclopedia](http://localhost:50764/document?identifier=Apocalypse_Now_a6a5)  [The Pit and the Pendulum (1961 film) - Wikipedia, the free encyclopedia](http://localhost:50764/document?identifier=The_Pit_and_the_Pendulum_%281961_film%29_e0db)  [Live at the Roxy WC2 - Wikipedia, the free encyclopedia](http://localhost:50764/document?identifier=Live_at_the_Roxy_WC2_64a3) | Finds the director’s page correctly |
| vision computer  or  computer vision | Does order matter? | [Point Distribution Model - Wikipedia, the free encyclopedia](http://localhost:50764/document?identifier=Point_Distribution_Model_df77)  [NUS School of Computing - Wikipedia, the free encyclopedia](http://localhost:50764/document?identifier=NUS_School_of_Computing_e6f0)  [Manarat International University - Wikipedia, the free encyclopedia](http://localhost:50764/document?identifier=Manarat_International_University_1618)  [Neuros Technology - Wikipedia, the free encyclopedia](http://localhost:50764/document?identifier=Neuros_Technology_a3a7)  [Visions of Atlantis - Wikipedia, the free encyclopedia](http://localhost:50764/document?identifier=Visions_of_Atlantis_62af) | Order is immaterial |
| Game sport | Disambiguation between computer games and sports | [Sporting Praia Cruz - Wikipedia, the free encyclopedia](http://localhost:50764/document?identifier=Sporting_Praia_Cruz_00e5)  [Truman Sports Complex - Wikipedia, the free encyclopedia](http://localhost:50764/document?identifier=Truman_Sports_Complex_d0a0)  [Jason Koumas - Wikipedia, the free encyclopedia](http://localhost:50764/document?identifier=Jason_Koumas_6351)  [Edwin Moses - Wikipedia, the free encyclopedia](http://localhost:50764/document?identifier=Edwin_Moses_0128) | Noun with ‘sport’ is selected |
| bowl game | Is NFL in top 5? | [Fiesta Bowl - Wikipedia, the free encyclopedia](http://localhost:50764/document?identifier=Fiesta_Bowl_75b3)  [East Carolina Pirates football - Wikipedia, the free encyclopedia](http://localhost:50764/document?identifier=East_Carolina_Pirates_football_a547)  [National Football Conference West Division - Wikipedia, the free encyclopedia](http://localhost:50764/document?identifier=National_Football_Conference_West_Division_e28a)  [2006 Purdue Boilermakers football team - Wikipedia, the free encyclopedia](http://localhost:50764/document?identifier=2006_Purdue_Boilermakers_football_team_41f1) | Yes |