

Jayden Hyman, Sabrina Le, Nhan Nguyen, Eric Tran

CS 121: Information Retrieval

Professor Lopes

5 November 2025

## Assignment 2: Web Crawler Report

### **Number of Unique Pages**

count: 18100

### **Longest Page**

url: <https://ics.uci.edu/~gbowker/classification>

Word count: 34819

### **50 Most Common Words**

1. events, 380771
2. day, 119463
3. support, 60266
4. media, 48822
5. wiki, 47060
6. files, 38735
7. tools, 37885
8. search, 36315
9. page, 25312
10. manager, 24518

11. png, 22524
12. file, 22021
13. history, 21203
14. pasted, 20926
15. upload, 19975
16. security, 19885
17. view, 19336
18. software, 19081
19. wics, 18185
20. services, 18181
21. revisions, 17575
22. date, 17377
23. show, 17364
24. virtual, 17283
25. code, 17210
26. environments, 17122
27. instructional, 16917
28. accounts, 16297
29. hardware, 16293
30. group, 16063
31. labs, 15821
32. windows, 15567
33. linux, 15407

- 34. projects, 15079
- 35. log, 15011
- 36. calendar, 14927
- 37. us, 14134
- 38. contact, 13657
- 39. sitemap, 13584
- 40. courses, 13433
- 41. site, 13233
- 42. user, 12963
- 43. apply, 12954
- 44. found, 12861
- 45. month, 12828
- 46. old, 12754
- 47. bytes, 12731
- 48. innovate, 12709
- 49. ghc, 12702
- 50. recent, 12689

### **Subdomains in uci.edu Domain**

accessibility.ics.uci.edu, 2

code.ics.uci.edu, 1

courselisting.ics.uci.edu, 1

cs.ics.uci.edu, 1

cs.uci.edu, 1

cwicsocal18.ics.uci.edu, 12

ics.uci.edu, 341

informatics.uci.edu, 1

insite.ics.uci.edu, 6

intranet.ics.uci.edu, 697

malek.ics.uci.edu, 1

mcs.ics.uci.edu, 2

mhcid.ics.uci.edu, 9

mswe.ics.uci.edu, 8

observium.ics.uci.edu, 1

stat.ics.uci.edu, 2

student-council.ics.uci.edu, 1

studentcouncil.ics.uci.edu, 1

summeracademy.ics.uci.edu, 6

wics.ics.uci.edu, 4906

wiki.ics.uci.edu, 11825

www.cs.uci.edu, 1

www.ics.uci.edu, 44

www.informatics.uci.edu, 213

www.stat.uci.edu, 17

