

COMPSCI 121 A2 Web Crawler Report

1) Unique Pages

Total unique pages: 15600

2) Longest Page

Longest page in terms of words: <http://ics.uci.edu/~cs224>, 78184 words

3) 50 Most Common Words

research, 23224
student, 13165
computer, 12978
uci, 12425
information, 11355
students, 11197
ics, 11021
e, 10880
software, 10800
graduate, 10375
events, 10351
data, 10333
n, 10133
undergraduate, 9688
physics, 9058
science, 8997
news, 8996
us, 8646
courses, 8414
ramesh, 8358
may, 8204
p, 8161
b, 8085
department, 8063
search, 7922
engineering, 7749
computing, 7729
learning, 7717
university, 7641
informatics, 7558
people, 7367

l, 6921
irvine, 6889
one, 6878
new, 6784
faculty, 6701
systems, 6669
projects, 6631
time, 6589
contact, 6527
home, 6427
current, 6415
past, 6282
community, 6223
c, 6030
design, 5949
r, 5847
read, 5781
books, 5633
june, 5467

4) Subdomain Counts

accessibility.ics.uci.edu, 6
acoi.ics.uci.edu, 75
aiclub.ics.uci.edu, 1
archive-beta.ics.uci.edu, 9
archive.ics.uci.edu, 189
asterix.ics.uci.edu, 7
betapro.proteomics.ics.uci.edu, 2
capstone.cs.uci.edu, 8
cdb.ics.uci.edu, 43
cecs.uci.edu, 12
cert.ics.uci.edu, 30
chemdb.ics.uci.edu, 42
chenli.ics.uci.edu, 10
circadiomics.ics.uci.edu, 6
cloudberry.ics.uci.edu, 41
cml.ics.uci.edu, 163
code.ics.uci.edu, 13
computableplant.ics.uci.edu, 96
courselisting.ics.uci.edu, 5
create.ics.uci.edu, 5

cs.ics.uci.edu, 12
cs.uci.edu, 7
cwicsocal18.ics.uci.edu, 12
cyberclub.ics.uci.edu, 11
cybert.ics.uci.edu, 24
demskey.eecs.uci.edu, 1
dgillen.ics.uci.edu, 10
ds4all.ics.uci.edu, 5
duttgroup.ics.uci.edu, 59
dynamo.ics.uci.edu, 31
economics.uci.edu, 1
edgelab.ics.uci.edu, 7
eecs.uci.edu, 501
emj.ics.uci.edu, 39
evoke.ics.uci.edu, 2
flamingo.ics.uci.edu, 21
fr.ics.uci.edu, 3
futurehealth.ics.uci.edu, 155
graphics.ics.uci.edu, 8
graphmod.ics.uci.edu, 1
hack.ics.uci.edu, 2
hai.ics.uci.edu, 6
hobbes.ics.uci.edu, 10
hpi.ics.uci.edu, 5
i-sensorium.ics.uci.edu, 5
icde2023.ics.uci.edu, 45
ics.uci.edu, 7985
industryshowcase.ics.uci.edu, 21
informatics.ics.uci.edu, 2
informatics.uci.edu, 1070
insite.ics.uci.edu, 9
intranet.ics.uci.edu, 1
ipubmed.ics.uci.edu, 1
isg.ics.uci.edu, 126
jdsylab.physics.uci.edu, 18
jgarcia.ics.uci.edu, 8
luci.ics.uci.edu, 4
mailman.ics.uci.edu, 40
malek.ics.uci.edu, 1
mcs.ics.uci.edu, 11

mdogucu.ics.uci.edu, 1
mds.ics.uci.edu, 8
mhcid.ics.uci.edu, 19
mlphysics.ics.uci.edu, 86
motifmap.ics.uci.edu, 1
mswe.ics.uci.edu, 10
mupro.proteomics.ics.uci.edu, 3
nalini.ics.uci.edu, 7
newport.eecs.uci.edu, 1
ngs.ics.uci.edu, 1842
oai.ics.uci.edu, 5
pepito.proteomics.ics.uci.edu, 5
physics.uci.edu, 970
plrg.eecs.uci.edu, 10
plrg.ics.uci.edu, 16
psearch.ics.uci.edu, 1
radicle.ics.uci.edu, 6
redmiles.ics.uci.edu, 1
riscit.ics.uci.edu, 3
scale.ics.uci.edu, 7
scratch.proteomics.ics.uci.edu, 4
seal.ics.uci.edu, 45
selectpro.proteomics.ics.uci.edu, 5
sherlock.ics.uci.edu, 7
stairs.ics.uci.edu, 4
stat.ics.uci.edu, 17
stat.uci.edu, 14
statconsulting.ics.uci.edu, 1
student-council.ics.uci.edu, 1
summeracademy.ics.uci.edu, 5
tad.ics.uci.edu, 2
tastier.ics.uci.edu, 1
tutoring.ics.uci.edu, 5
ugradforms.ics.uci.edu, 1
unite.ics.uci.edu, 10
vision.ics.uci.edu, 210
wics.ics.uci.edu, 1189
wiki.ics.uci.edu, 1
www-db.ics.uci.edu, 23
xtune.ics.uci.edu, 6