Assignment 2 Report

How many unique pages did you find?
Number of unique pages: 5204

What is the longest page in terms of the number of words?
Longest page: http://www.ics.uci.edu/~kay/wordlist.txt that has 385018 words

What are the 50 most common words in the entire set of pages crawled under these domains ?
The top 50 words are:
research, 17844
html, 12872
update, 12303
software, 11496
automatic, 11252
ics, 10631
student, 10236
information, 9886
will, 9208
computer, 9124
students, 9091
data, 8516
graduate, 7629
may, 7613
engineering, 7547
undergraduate, 6972
informatics, 6729
can, 6578
science, 6424
news, 6360
projects, 6141
support, 5804
ramesh, 5696
department, 5604
people, 5511
irvine, 5391
us, 5376
courses, 5319
events, 5311
learning, 5269
design, 5236
uci, 5207
systems, 5174

current, 5137
computing, 5005
new, 4995
search, 4916
june, 4880
ph.d, 4779
past, 4719
one, 4599
september, 4552
books, 4507
july, 4455
university, 4331
march, 4286
april, 4255
november, 4231
contact, 4201
august, 4199

How many subdomains did you find in the uci.edu domain?
Number of subdomains: 52
The subdomains are:
accessibility.ics.uci.edu, 5
acoi.ics.uci.edu, 81
archive.ics.uci.edu, 8
asterix.ics.uci.edu, 5
cert.ics.uci.edu, 25
checkin.ics.uci.edu, 3
cml.ics.uci.edu, 36
code.ics.uci.edu, 13
computableplant.ics.uci.edu, 86
courselisting.ics.uci.edu, 3
cs.ics.uci.edu, 11
cs.uci.edu, 1
dgillen.ics.uci.edu, 8
duttgroup.ics.uci.edu, 46
emj.ics.uci.edu, 33
flamingo.ics.uci.edu, 12
fr.ics.uci.edu, 2
frost.ics.uci.edu, 1
futurehealth.ics.uci.edu, 171
gitlab.ics.uci.edu, 36
grape.ics.uci.edu, 10
graphics.ics.uci.edu, 1
hai.ics.uci.edu, 5

helpdesk.ics.uci.edu, 1
hpi.ics.uci.edu, 4
hub.ics.uci.edu, 3
iasl.ics.uci.edu, 1
ics.uci.edu, 228
informatics.ics.uci.edu, 1
informatics.uci.edu, 2
intranet.ics.uci.edu, 4
isg.ics.uci.edu, 137
mailman.ics.uci.edu, 17
mhcid.ics.uci.edu, 17
mswe.ics.uci.edu, 10
nalini.ics.uci.edu, 6
ngs.ics.uci.edu, 987
observium.ics.uci.edu, 1
password.ics.uci.edu, 2
seal.ics.uci.edu, 47
stat.uci.edu, 1
statconsulting.ics.uci.edu, 4
support.ics.uci.edu, 10
swiki.ics.uci.edu, 200
vision.ics.uci.edu, 210
wics.ics.uci.edu, 480
wiki.ics.uci.edu, 203
www-db.ics.uci.edu, 24
www.cs.uci.edu, 7
www.ics.uci.edu, 1210
www.informatics.uci.edu, 552
www.stat.uci.edu, 233