

## A2 Web Crawler Report

1. How many unique pages did you find?

Number of unique pages: 7359

2. What is the longest page in terms of the number of words?

The longest page by number of words was <http://www.ics.uci.edu/~kay/wordlist.txt>

3. What are the 50 most common words in the entire set of pages crawled under these domains?

research - 30871  
information - 21646  
computer - 19409  
software - 18951  
2020 - 18686  
students - 16322  
student - 15729  
2018 - 15644  
2021 - 14738  
data - 14594  
design - 14556  
events - 14408  
2019 - 13986  
ramesh - 13969  
engineering - 12842  
2016 - 12652  
may - 12151  
informatics - 11872  
2017 - 11724  
news - 11575  
says - 11566  
science - 11406  
2015 - 11107  
graduate - 10895  
reply - 10715  
current - 10394  
undergraduate - 10371  
march - 10169

systems - 10001  
search - 9804  
read - 9787  
courses - 9766  
page - 9736  
april - 9586  
past - 9423  
new - 9215  
uci - 9175  
projects - 9080  
department - 9017  
people - 9005  
books - 8880  
january - 8821  
computing - 8770  
february - 8516  
time - 8439  
november - 8430  
september - 8369  
support - 8309  
december - 8226  
october - 7975

4. How many subdomains did you find in the ics.uci.edu domain?

Found 129 subdomains

acoi.ics.uci.edu - 1413  
aiclub.ics.uci.edu - 6  
alumni.ics.uci.edu - 10  
archive.ics.uci.edu - 16  
asterix-gerrit.ics.uci.edu - 1  
asterix.ics.uci.edu - 34  
asterixdb.ics.uci.edu - 2  
auge.ics.uci.edu - 2  
awareness.ics.uci.edu - 2  
calendar.ics.uci.edu - 141  
cbcl.ics.uci.edu - 4  
cert.ics.uci.edu - 9  
cgvw.ics.uci.edu - 3

checkmate.ics.uci.edu - 12  
chenli.ics.uci.edu - 32  
cloudberry.ics.uci.edu - 1493  
cml.ics.uci.edu - 4751  
cocoa-krispies.ics.uci.edu - 1  
codeexchange.ics.uci.edu - 8  
computableplant.ics.uci.edu - 347  
coronavirustwittermap.ics.uci.edu - 90  
cradl.ics.uci.edu - 241  
create.ics.uci.edu - 73  
cwicsocal18.ics.uci.edu - 1352  
cyberclub.ics.uci.edu - 9  
dataguard.ics.uci.edu - 1  
dataprotector.ics.uci.edu - 1  
dblp.ics.uci.edu - 2  
dejavu.ics.uci.edu - 1  
duttgroup.ics.uci.edu - 2848  
dynamo.ics.uci.edu - 4  
elms.ics.uci.edu - 2  
emj-pc.ics.uci.edu - 1  
emj.ics.uci.edu - 1626  
esl.ics.uci.edu - 4  
evoke.ics.uci.edu - 22526  
flamingo.ics.uci.edu - 16  
fr.ics.uci.edu - 18  
frost.ics.uci.edu - 1  
futurehealth.ics.uci.edu - 28  
givargis.ics.uci.edu - 1  
grape.ics.uci.edu - 23  
graphics.ics.uci.edu - 21  
graphmod.ics.uci.edu - 1  
hack.ics.uci.edu - 1  
hai.ics.uci.edu - 4  
halps.ics.uci.edu - 1  
hana.ics.uci.edu - 61  
hanalab.ics.uci.edu - 1  
helpdesk.ics.uci.edu - 17  
hombao.ics.uci.edu - 8  
honors.ics.uci.edu - 1  
i-sensorium.ics.uci.edu - 2

iasl.ics.uci.edu - 9  
informatics.ics.uci.edu - 34  
intranet.ics.uci.edu - 341  
ipf.ics.uci.edu - 8  
ipubmed.ics.uci.edu - 8  
isg.ics.uci.edu - 3413  
jgarcia.ics.uci.edu - 493  
jujube.ics.uci.edu - 2  
kdd.ics.uci.edu - 1  
keys.ics.uci.edu - 2  
luci.ics.uci.edu - 34  
mailman.ics.uci.edu - 9  
malek.ics.uci.edu - 16  
map125.ics.uci.edu - 2  
mapgrid.ics.uci.edu - 2  
mcs.ics.uci.edu - 2399  
mdogucu.ics.uci.edu - 2  
mds.ics.uci.edu - 460  
metaviz.ics.uci.edu - 1  
mhcid.ics.uci.edu - 1311  
mondego.ics.uci.edu - 13  
mse.ics.uci.edu - 1038  
mswe.ics.uci.edu - 853  
nalini.ics.uci.edu - 90  
ngs.ics.uci.edu - 106822  
nile.ics.uci.edu - 2  
omni.ics.uci.edu - 1  
password.ics.uci.edu - 17  
pasteur.ics.uci.edu - 2  
perennialpolycultures.ics.uci.edu - 2  
plrg.ics.uci.edu - 260  
psearch.ics.uci.edu - 4  
redmiles.ics.uci.edu - 44  
riscit.ics.uci.edu - 1  
satware.ics.uci.edu - 1  
scale.ics.uci.edu - 1  
sconce.ics.uci.edu - 7  
sdel.ics.uci.edu - 6505  
se.ics.uci.edu - 1  
seal.ics.uci.edu - 12

seraja.ics.uci.edu - 1  
sherlock.ics.uci.edu - 5  
sli.ics.uci.edu - 27081  
soc.ics.uci.edu - 1  
sourcerer.ics.uci.edu - 2  
sprout.ics.uci.edu - 9  
stairs.ics.uci.edu - 1  
statconsulting.ics.uci.edu - 218  
student-council.ics.uci.edu - 4  
studentcouncil.ics.uci.edu - 2  
support.ics.uci.edu - 5  
swiki.ics.uci.edu - 2  
tastier.ics.uci.edu - 2  
tippers.ics.uci.edu - 2  
tippersweb.ics.uci.edu - 582  
transformativeplay.ics.uci.edu - 966  
tutors.ics.uci.edu - 1  
ugradforms.ics.uci.edu - 3  
unite.ics.uci.edu - 220  
vision.ics.uci.edu - 32  
wearablegames.ics.uci.edu - 166  
wics.ics.uci.edu - 35702  
www-db.ics.uci.edu - 168  
www-db.ics.uci.edu:8000 - 1  
www-db.ics.uci.edu:8003 - 1  
www-db.ics.uci.edu:8004 - 1  
www-db.ics.uci.edu:8006 - 1  
www-netra.ics.uci.edu - 4  
www.cert.ics.uci.edu - 7  
www.graphics.ics.uci.edu - 2  
www.ics.uci.edu - 14460  
www.informatics.ics.uci.edu - 7  
www.informatics.uci.edu - 202936  
www.isg.ics.uci.edu - 1  
xtune.ics.uci.edu - 33  
yarra.ics.uci.edu - 5