

Jason Davis 22336416
Joya Monishi 23598856
Irene E. Kim 21544176

Assignment 2: Web Crawler Report

QUESTION 1 : Number of Unique URLs - 5911

QUESTION 2 : Longest Webpage

<https://evoke.ics.uci.edu/qs-personal-data-landscapes-poster/#comment-70732> 168725 words

QUESTION 3: 50 Most Common Words ['research', '2020', '2021', 'student', '2018', '2019', '2016', 'students', '2017', 'says', '2015', 'reply', 'information', 'engineering', 'bs', 'news', 'march', 'graduate', 'april', 'undergraduate', 'informatics', 'software', 'courses', 'january', 'computer', 'science', 'february', 'past', 'books', 'may', 'phd', 'events', 'november', 'december', 'september', 'design', 'ramesh', 'projects', 'june', 'current', 'october', 'will', 'august', 'july', 'policies', 'pm', 'search', 'computing', 'jain', 'data']

QUESTION 4: Subdomains and Number of Pages Within

acoi.ics.uci.edu	45
aiclub.ics.uci.edu	1
asterix.ics.uci.edu	1
cbcl.ics.uci.edu	2
cert.ics.uci.edu	2
chenli.ics.uci.edu	1
cloudberry.ics.uci.edu	8
cml.ics.uci.edu	137
computableplant.ics.uci.edu	6
coronavirustwittermap.ics.uci.edu	1
cradl.ics.uci.edu	13
create.ics.uci.edu	3
cyberclub.ics.uci.edu	1
dejavu.ics.uci.edu	1
duttgroup.ics.uci.edu	78
dynamo.ics.uci.edu	1
emj.ics.uci.edu	27
esl.ics.uci.edu	1
evoke.ics.uci.edu	68
fr.ics.uci.edu	2
futurehealth.ics.uci.edu	2
grape.ics.uci.edu	11
graphics.ics.uci.edu	2
graphmod.ics.uci.edu	1
hai.ics.uci.edu	2

hana.ics.uci.edu	13
iasl.ics.uci.edu	5
ics.uci.edu	482
informatics.ics.uci.edu	1
ipf.ics.uci.edu	1
ipubmed.ics.uci.edu	1
isg.ics.uci.edu	93
jgarcia.ics.uci.edu	5
luci.ics.uci.edu	3
malek.ics.uci.edu	2
mcs.ics.uci.edu	38
mds.ics.uci.edu	7
mhcids.ics.uci.edu	9
mondego.ics.uci.edu	3
mse.ics.uci.edu	1
mswe.ics.uci.edu	17
ngs.ics.uci.edu	1777
perennialpolycultures.ics.uci.edu	1
plrg.ics.uci.edu	11
redmiles.ics.uci.edu	5
riscit.ics.uci.edu	1
scale.ics.uci.edu	1
sdcl.ics.uci.edu	69
seal.ics.uci.edu	5
sli.ics.uci.edu	244
sourcerer.ics.uci.edu	1
sprout.ics.uci.edu	1
statconsulting.ics.uci.edu	4
studentcouncil.ics.uci.edu	1
support.ics.uci.edu	1
transformativeplay.ics.uci.edu	39
vision.ics.uci.edu	1
wearablegames.ics.uci.edu	8
wics.ics.uci.edu	367
xtune.ics.uci.edu	2