Team: Yuan Wang (38713273), Yan Xiao (35655533)

Assignment 2 Report:

Q1: We have found 12333 unique pages

Q2: The longest page in terms of words is <http://www.ics.uci.edu/~kay/wordlist.txt>, and there are 384954 words in it.

Q3: Here’s the following 50 most common words in the entire set of pages ordered by frequency:

research: 33859

events: 24439

computer: 24047

information: 21425

ramesh: 19439

students: 17064

software: 16513

will: 15628

student: 15360

search: 15237

uci: 15012

data: 13952

current: 13941

news: 13342

science: 12373

past: 12373

engineering: 12062

may: 11831

us: 11726

courses: 11539

informatics: 11348

graduate: 11031

books: 10972

can: 10920

undergraduate: 10608

time: 10536

code: 10445

jain: 10396

ics: 10288

october: 10257

people: 10204

design: 10124

computing: 10070

projects: 9911

september: 9871

contact: 9597

august: 9509

read: 9450

department: 9139

ph: 9050

june: 8958

support: 8826

july: 8732

new: 8502

upcoming: 8291

march: 8177

course: 8129

april: 8092

november: 8077

one: 8042

Q4: Here’s the list of subdomains ordered alphabetically:

accessibility.ics.uci.edu: 1

acoi.ics.uci.edu: 56

aiclub.ics.uci.edu: 1

archive.ics.uci.edu: 5

asterix.ics.uci.edu: 6

cbcl.ics.uci.edu: 3

cert.ics.uci.edu: 2

checkmate.ics.uci.edu: 1

chenli.ics.uci.edu: 1

cloudberry.ics.uci.edu: 59

cml.ics.uci.edu: 175

code.ics.uci.edu: 12

computableplant.ics.uci.edu: 32

cradl.ics.uci.edu: 20

create.ics.uci.edu: 6

cwicsocal18.ics.uci.edu: 12

cyberclub.ics.uci.edu: 2

dejavu.ics.uci.edu: 1

dgillen.ics.uci.edu: 19

duttgroup.ics.uci.edu: 107

dynamo.ics.uci.edu: 1

emj.ics.uci.edu: 45

esl.ics.uci.edu: 1

flamingo.ics.uci.edu: 6

fr.ics.uci.edu: 3

frost.ics.uci.edu: 1

futurehealth.ics.uci.edu: 4

grape.ics.uci.edu: 12

graphics.ics.uci.edu: 3

graphmod.ics.uci.edu: 1

hai.ics.uci.edu: 3

helpdesk.ics.uci.edu: 1

i-sensorium.ics.uci.edu: 1

iasl.ics.uci.edu: 6

industryshowcase.ics.uci.edu: 20

informatics.ics.uci.edu: 1

intranet.ics.uci.edu: 1

ipf.ics.uci.edu: 2

ipubmed.ics.uci.edu: 1

isg.ics.uci.edu: 138

jgarcia.ics.uci.edu: 23

luci.ics.uci.edu: 4

malek.ics.uci.edu: 1

mcs.ics.uci.edu: 31

mdogucu.ics.uci.edu: 1

mds.ics.uci.edu: 11

mhcid.ics.uci.edu: 16

mondego.ics.uci.edu: 3

mse.ics.uci.edu: 2

mswe.ics.uci.edu: 16

nalini.ics.uci.edu: 7

ngs.ics.uci.edu: 3034

perennialpolycultures.ics.uci.edu: 1

plrg.ics.uci.edu: 14

psearch.ics.uci.edu: 1

redmiles.ics.uci.edu: 4

riscit.ics.uci.edu: 1

scale.ics.uci.edu: 1

sconce.ics.uci.edu: 2

sdcl.ics.uci.edu: 208

seal.ics.uci.edu: 6

sherlock.ics.uci.edu: 1

sli.ics.uci.edu: 386

sourcerer.ics.uci.edu: 1

sprout.ics.uci.edu: 2

stairs.ics.uci.edu: 1

statconsulting.ics.uci.edu: 5

student-council.ics.uci.edu: 1

studentcouncil.ics.uci.edu: 3

swiki.ics.uci.edu: 1

tastier.ics.uci.edu: 1

transformativeplay.ics.uci.edu: 59

tutors.ics.uci.edu: 1

ugradforms.ics.uci.edu: 1

unite.ics.uci.edu: 10

vision.ics.uci.edu: 7

wearablegames.ics.uci.edu: 11

wics.ics.uci.edu: 2614

www-db.ics.uci.edu: 11

www.cert.ics.uci.edu: 1

www.graphics.ics.uci.edu: 1

www.ics.uci.edu: 1168

www.informatics.ics.uci.edu: 1

www.informatics.uci.edu: 1009

xtune.ics.uci.edu: 6