**How to compile and run the code:**

The below library dependencies were used to compile this project:

oauth.signpost:signpost-commonshttp4:1.2.1.2

org.apache.httpcomponents:httpcore:4.3

org.apache.httpcomponents:httpclient:4.3-alpha1

oauth.signpost:signpost-core:1.2.1.2

commons-logging:commons-logging:1.1.3

commons-io:commons-io:2.4

commons-codec:commons-codec:1.6

org.json:json:20180813

org.jsoup:jsoup:1.11.3

Please run the main method in the SearchTweets.java file.

1.) javac SearchTweets.java

2.) java SearchTweets

The data.txt contains the tweets that were extracted using the API and the query word “#oscar”. Please include the txt file in the same folder to start crawling from the extracted links.

The tweet extractor code is in searchTweets() method.

The crawler method contains the crawler code and keeps track of the incoming and outgoing links from a particular URL.

**Unique Links:** 11343

Links of document type: 1830

Links of image type: 2408

Links of HTML/ text type: 7105

**Freq distribution of links by Domain name:**

www.addtoany.com: 9

www.estnetawards.co.uk: 1

bouldercityreview.com: 1

forums.craigslist.org: 1

wp.me: 2

it.pinterest.com: 1

www.paula.cl: 1

autoinsuranceband.org: 1

www.ivoox.com: 198

mx.ivoox.com: 5

www.business.com: 1

signin.ebay.com: 1

www.6pm.com: 1

b2serve.com: 1

Tameraott7702076.Wikidot.com: 1

www.diarioconcepcion.cl: 1

ilmetropolitano.it: 1

hi-in.facebook.com: 1

music.amazon.com: 1

astardiamonds.com: 229

ivoox.a2hosted.com: 1

www.pillpack.com: 1

kck.st: 1

imovision.com.br: 2

thethemefoundry.com: 1

losspreventionmedia.com: 2

portalrefil.com.br: 67

snaporazblast.wordpress.com: 1

www.vimeo.com: 2

www.amazoninspire.com: 1

www.billmelater.com: 1

zh-cn.facebook.com: 1

www.vendhq.com: 2

www.junglee.com: 1

www.kickstarter.com: 49

aws.amazon.com: 1

monumentscripts.com: 21

www.youfuture.it: 2

www.elle.com: 1

advertising.amazon.com: 2

www.beethovenfm.cl: 1

cgi6.ebay.com: 1

anagrammer.uservoice.com: 1

account.reviewjournal.com: 7

www.thedigitaltransformationpeople.com: 1

www.dailymotion.com: 5

www.fabrica101.com: 1

www.houseofviolin.com: 1

dwnlds.co: 63

open.spotify.com: 4

rapids.amazon.com: 1

www.grupocopesa.cl: 2

culto.latercera.com: 1

sport.periodicodaily.com: 76

www.ilmetropolitano.it: 161

www.fictioncorporation.com.br: 1

www.biut.cl: 1

www.kirmizikurdele.org: 1

www.d-ddaily.com: 1

blog.aboutamazon.com: 2

pages.ebay.com: 20

www.uktech-awards.co.uk: 1

musicnsouls.com: 142

www.epossystemsguide.co.uk: 1

ja-jp.facebook.com: 1

nrfbigshow.nrf.com: 1

www.buzzfeednews.com: 1

cgi1.ebay.com: 1

pvtimes.com: 1

instagram.com: 1

nevadapreps.com: 2

www.cineroad.it: 2

hornet.com: 1

www.alexa.com: 1

especiales.latercera.com: 1

www.ebay.com: 54

pt-br.facebook.com: 1

www.glamorama.cl: 1

awards.bankingtech.com: 1

laclaquetayyo.blogspot.com: 21

avisoslegales.latercera.com: 1

buff.ly: 2

www.shinystat.com: 1

pinterest.com: 4

www.abebooks.com: 1

affiliate-program.amazon.com: 1

suscripcioneslt.latercera.com: 1

wordpress.com: 2

ocsnext.ebay.com: 2

www.learningtechnologies.co.uk: 1

istara.fm: 1

www.goodreads.com: 1

pulsar.ebay.com: 36

rateyourmusic.com: 1

marketscale.com: 1

rarbg.unblocked2.me: 1

apparelmag.com: 1

www.ihlservices.com: 3

www.hotfrog.com: 1

www.bizbuysell.com: 1

www.hannush.com: 1

rover.ebay.com: 1

www.latercera.com: 4

www.eastdane.com: 1

ivoox.zendesk.com: 1

www.1cares.asia: 1

www.allmusic.com: 1

alanbice46022563.wikidot.com: 1

partnernetwork.ebay.com: 1

mouse.latercera.com: 1

jobs.reviewjournal.com: 2

www.cinemablend.com: 1

discettazionierranti.blogspot.com: 4863

www.campaigntechawards.com: 1

mesg.ebay.com: 1

www.facebook.com: 114

cpbldi.com: 21

www.topcatmedia.com: 1

tv.youtube.com: 4

plus.google.com: 15

www.aboutamazon.com: 1

lacuarta.com: 1

www.linkedin.com: 2

cart.payments.ebay.com: 1

luminouspr.com: 63

gzone.com.tr: 205

petersjohns63.carbonmade.com: 1

post.craigslist.org: 1

kdp.amazon.com: 1

amzn.to: 1

www.bookdepository.com: 1

www.woot.com: 1

www.anagrammer.com: 89

www.audible.com: 1

es-la.facebook.com: 1

play.google.com: 3

www.theriverside.co.uk: 62

britishlegalitawards.com: 1

www.theretaildigest.com: 1

www.zappos.com: 1

ow.ly: 5

mailchimp.com: 1

bestoflasvegas.com: 2

blog.feedspot.com: 1

www.iubenda.com: 2

my.ebay.com: 2

creditapply.paypal.com: 2

disqus.com: 1

community.ebay.com: 1

resolutioncenter.ebay.com: 1

www.retailroi.org: 1

besttayo.com: 13

ar-ar.facebook.com: 1

www.google.com: 8

cbldc.io: 63

www.youtube.com: 262

gaysofturkey.com: 1

www.fabric.com: 1

radiofobia.com.br: 1

remates-propiedades-tgr.latercera.com: 1

it-it.facebook.com: 1

track.webgains.com: 42

www.radiozero.cl: 1

www.boxofficemojo.com: 1

www.ebayinc.com: 2

worldin2019.economist.com: 2

www.shopbop.com: 1

www.lahoramujeres.cl: 15

bit.ly: 33

cgi5.ebay.com: 1

redirect.viglink.com: 50

review-journal-store.myshopify.com: 1

www.bbfc.co.uk: 1

www.pinterest.com: 1

qu.ebay.com: 1

www.blogger.com: 21

www.souq.com: 1

www.acx.com: 1

support.google.com: 4

luxurylv.com: 1

esquentamente.com.br: 1

canaldenuncias.grupocopesa.cl: 2

www.cityfeet.com: 2

www.papodegordo.com.br: 1

vip.wordpress.com: 1

announcements.ebay.com: 1

www.giovis.com: 4

offer.ebay.com: 2

fr-fr.facebook.com: 1

pro.imdb.com: 1

twitter.com: 38

businesspress.vegas: 1

accounts.google.com: 20

www.mymovies.it: 1

www.craigslist.org: 6

feedback.ebay.com: 1

accounts.craigslist.org: 1

automattic.com: 1

toronto.craigslist.org: 9

www.amazon.jobs: 1

themegrill.com: 1

www.amazon.com: 211

www.instagram.com: 55

www.imdb.com: 4

checkout.reviewjournal.com: 1

contact.ebay.com: 2

bitsc.io: 21

eltiempolv.com: 1

www.comixology.com: 1

youtu.be: 1

obits.reviewjournal.com: 4

classifieds.reviewjournal.com: 12

messenger.com: 1

ko-kr.facebook.com: 1

developers.facebook.com: 1

trustsealinfo.websecurity.norton.com: 1

www.wholefoodsmarket.com: 1

www.dpreview.com: 1

www.elmagacin.com: 1

www.lahora.cl: 1

www.bfi.org.uk: 1

www.windowsphone.com: 2

www.tonydonofrio.com: 151

www.duna.cl: 1

noticiasalsur.co: 164

alphavillepc.wordpress.com: 181

www.reviewjournal.com: 285

itunes.apple.com: 3

www.createspace.com: 1

www.tomsguide.com: 1

goo.gl: 3

services.amazon.com: 1

www.gotechawards.co.uk: 1

wordpress.org: 1

l.facebook.com: 1

www.dameskarlette.com: 1

paranerdia.com.br: 1

primenow.amazon.com: 2

developer.amazon.com: 1

videodirect.amazon.com: 1

www.pulso.cl: 1

www.icarito.cl: 1

www.disneylatino.com: 1

flipbooks.reviewjournal.com: 3

www.astardiamonds.com: 1

de-de.facebook.com: 1

feedback.anagrammer.com: 2

remates.latercera.com: 1

lasvegasreviewjournal.adperfect.com: 5

**Average link depth:** 0.9938071780436313

**Incoming top 25 links:**

http://noticiasalsur.co/author/periodista1/: 39

https://astardiamonds.com/en/#: 28

https://cbldc.io/49a280a#: 17

https://dwnlds.co/4beb359#: 17

https://dwnlds.co/c292207#: 17

https://bitsc.io/71d2806#: 17

https://cbldc.io/b6c82d6#: 17

https://cbldc.io/ae67826#: 17

https://cpbldi.com/cf514f2#: 17

https://dwnlds.co/d99fe22#: 17

https://portalrefil.com.br/2019/02/28/exibidores-da-europa-dizem-que-os-oscars-de-roma-desvalorizam-o-premio#: 15

http://gzone.com.tr/recep-ozdas-green-booku-yazdi-kocanizi-benimle-paylastiginiz-icin-tesekkurler/#: 13

https://www.kickstarter.com/projects/rocketmigration/rocket-migration?ref=project\_link#: 13

http://musicnsouls.com/9-musicians-who-are-celebrating-national-puppy-day/#: 13

https://www.kickstarter.com/trust?ref=project\_reward: 12

https://sport.periodicodaily.com/in-forma-con-lady-gaga-in-5-step/#: 12

http://musicnsouls.com/9-musicians-who-are-celebrating-national-puppy-day/: 10

https://bit.ly/2W8mbW6: 8

https://www.youtube.com/: 8

https://www.tonydonofrio.com/tags/retail.html: 7

https://bit.ly/2CoKDvf: 7

https://www.tonydonofrio.com/tags/omnichannel.html: 7

https://bit.ly/2OcbxLj: 7

http://noticiasalsur.co/julio-verne-genio-de-la-literatura-y-la-ciencia-ficcion/: 7

https://www.facebook.com/5550296508/posts/10159150431856509?sfns=mo#: 7

**Outgoing top 25 links:**

https://discettazionierranti.blogspot.com/2014/05/: 5533

https://discettazionierranti.blogspot.com/search/label/giovis: 5489

https://discettazionierranti.blogspot.com/search/label/Agerola: 5458

https://discettazionierranti.blogspot.com/2013/: 5418

https://discettazionierranti.blogspot.com/2014/11/: 5404

https://discettazionierranti.blogspot.com/2014/07/: 5394

https://discettazionierranti.blogspot.com/2014/: 5382

https://discettazionierranti.blogspot.com/2014/12/: 5382

https://discettazionierranti.blogspot.com/2014/06/: 5380

https://discettazionierranti.blogspot.com/search/label/MaraTrail: 5370

https://discettazionierranti.blogspot.com/search/label/Trek: 5363

https://discettazionierranti.blogspot.com/2015/05/: 5355

https://discettazionierranti.blogspot.com/2015/01/: 5355

https://discettazionierranti.blogspot.com/search/label/100km: 5344

https://discettazionierranti.blogspot.com/search/label/Pogerola: 5328

https://discettazionierranti.blogspot.com/search/label/tradizionale: 5324

https://discettazionierranti.blogspot.com/search/label/camminate: 5321

https://discettazionierranti.blogspot.com/search/label/cucina: 5317

https://discettazionierranti.blogspot.com/search/label/Punta%20Campanella: 5316

https://discettazionierranti.blogspot.com/search/label/cartine: 5316

https://discettazionierranti.blogspot.com/search/label/Giovanni%20Visetti: 5315

https://discettazionierranti.blogspot.com/search/label/Bomerano: 5312

https://discettazionierranti.blogspot.com/search/label/Sentiero%20Dei: 5307

https://discettazionierranti.blogspot.com/search/label/Menorca: 5304

https://discettazionierranti.blogspot.com/2015/03/: 5302