

DSE 203 Group 8: Alibaba

Faezeah Ghazi
Jimmy Quach
Kevin Liu
Ying Cui

Outline

- ❖ **Introduction**
- ❖ **Text Extraction and NLP**
- ❖ **Structured Text Extraction**
- ❖ **Neo4J Graph Creation**
- ❖ **Queries**
- ❖ **Lessons Learned**

Introduction

Introduction: Project

- Construct a knowledge graph as a property - graph view from heterogeneous data sources
 - graph is comprised of technology companies and their related metadata
- Additional nodes within the graph are based off of NAICS data
- Additional nodes within the graph should be created based off of NLP text extractions
- From the graph created, create interesting queries and insights that can be formed due to traversing the graph

Introduction: Alibaba



Text Extraction and NLP

Text Extraction and NLP: Pre-processing

Pre-processing Text

1. Sentence segmentation.
2. Normalization and tokenization.
3. Part-of-speech (POS) tagging.
4. Chunked tokens
5. Named entity recognition.
6. Feature Extraction
7. Word2Vec

Part of Speech Tagging

```
1 #for token in introduction_doc:  
2     #print (token, token.tag_, token.pos_, spacy.explain(token.tag_))
```

ebay JJ ADJ adjective
eebay JJ ADJ adjective
american JJ ADJ adjective
multinational JJ ADJ adjective
ecommerce NN NOUN noun, singular or mass
corporation NN NOUN noun, singular or mass
base NN NOUN noun, singular or mass
san FW X foreign word
jose NN NOUN noun, singular or mass
california NN NOUN noun, singular or mass
facilitate NN NOUN noun, singular or mass
consumertoconsumer NN NOUN noun, singular or mass
businessstoconsumer NN NOUN noun, singular or mass
sale NN NOUN noun, singular or mass
website NN NOUN noun, singular or mass
ebay NN NOUN noun, singular or mass
found VBD VERB verb, past tense
pierre JJ ADJ adjective



Chunking:

```
1 for chunk in introduction_doc.noun_chunks:  
2     print (chunk)  
  
eBay EEbay american multinational ecommerce corporation base  
San Jose California  
consumertoconsumer businessstoconsumer sale website  
eBay  
Pierre Omidyar autumn notable success story dotcom bubble  
eBay multibilliondollar business operation country company  
wide variety good service worldwide website free buyer seller charge fee list item 1  
er free listing item soldIn addition  
original auctionstyle sale website  
evolve expand  
instant Buy shopping shopping  
Universal Product Code ISBN kind sku number Halfcom  
online classify advertisement Kijiji eBay  
online event ticket trading StubHub service  
eBay  
online money transfer service PayPal  
wholly own subsidiary eBay History s AuctionWeb  
California September Frenchborn IranianAmerican computer programmer
```

Finding nouns and Adjectives

```
nouns = []
adjectives = []
for token in introduction_doc:
    if token.pos_ == 'NOUN':
        nouns.append(token)
    if token.pos_ == 'ADJ':
        adjectives.append(token)

print(nouns,adjectives)

ommerce, corporation, base, jose, california, facilitate, consumertoconsumer, businessstoco
mer, sale, website, ebay, autumn, success, story, dotcom, bubble, business, operation, cou
', company, website, auction, shopping, website, people, business, sell, variety, service,
site, buyer, seller, charge, fee, list, item, limit, number, listing, item, soldin, additi
auctionstyle, sale, website, evolve, expand, buy, shopping, shopping, product, code, sku,
per, halfcom, advertisement, kijiji, event, ticket, trading, stubhub, service, money, tran
:, service, paypal, history, auctionweb, september, computer, programmer, site, item, lase
pointer, contact, bidder, understand, laser, pointer, email, buyer, collector, break, lase
pointer, auction, site, persontoperson, transaction, volume, ebay, internet, service, prov
:, inform, upgrade, business, account, volume, traffic, website, result, price, increase,
ch, start, charge, meet, animosity, result, chris, employee, process, mail, check, jeffre
skoll, president, company, november, ebay, thirdparty, licensing, deal, company, travel, a
on, smartmarket, technology, sell, plane, ticket, travel, product, growth, site, host, au
n, company, service, september, site, echo, bay, technology, group, firm, omidyar, try, r
```

Text Extraction and NLP: Technology

Beautiful Soup

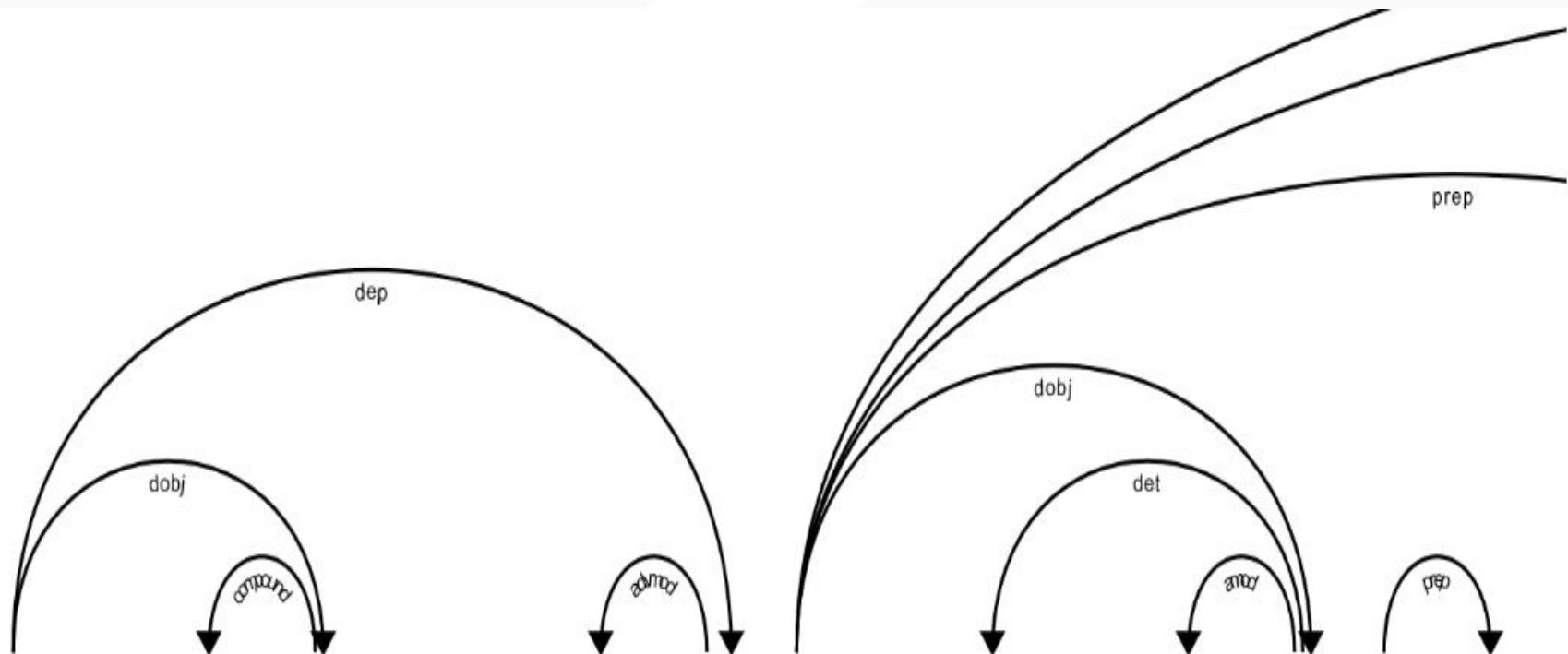
Spacy

Textacy

Neuralcoref

<p>Alibaba Group Holding Limited, (also known as Alibaba Group and as Alibaba), is a Chinese multinational conglomerate holding company specializing in e-commerce, retail, Internet, and technology. Founded on 4 April 1999 in Hangzhou, Zhejiang, the company provides consumer-to-consumer (C2C), business-to-consumer (B2C), and business-to-business (B2B) sales services via web portals, as well as electronic payment services, shopping search engines and cloud computing services. It owns and operates a diverse array of businesses around the world in numerous sectors, and is named as one of the world's most admired companies by <i>Fortune</i>. ^{#91;3#93;} ^{#91;4#93;} </p><p>At closing time on the date of its initial public offering (IPO) – US\$25 billion – the world's highest in history, 19 September 2014, Alibaba's market value was US\$231 billion. ^{#91;5#93;} As of 19 December#160;2018^{#91;update#93;}, Alibaba's market cap stood at US\$352.28 billion. ^{#91;6#93;} It is one of the top 10 most valuable and is the 59th biggest public companies in the world by Global 2000 list. ^{#91;7#93;} In January 2018, Alibaba became the second Asian company to break the US\$500 billion valuation mark, after its competitor Tencent. ^{#91;8#93;} As of 2018^{#91;update#93;}, Alibaba has the 9th highest global brand value. ^{#91;9#93;}

Word dependencies



Firing	Mr.	Strzok,	however,	removes	a	favorite	target	of
VERB	PROPN	PROPN	ADV	VERB	DET	ADJ	NOUN	ADP

Visualize the entities of the entire texts

opensource security design properly document recent independent security audit AIM BlackBerry Messenger Ebuddy PERSON XMS Hushmail Kik
Messenger PERSON Viber Yahoo Messenger PERSON score pointsas August DATE Skype ORG support endtoend encryption platform Service
Peoples Republic GPE China September DATE user China GPE try download Skype ORG software client redirect site TOM ORG online joint
venture chinese wireless operator Skype ORG modify chinese version download TOM ORG client participate Chinas ORG internet censorship
monitoring text message Skype ORG user China GPE message exchange user outside country Niklas Zennstrom PERSON chief executive
Skype PERSON tell reporter TOM ORG implement text filter market regulation state thing that s certain thing way jeopardise privacy security usersln
October DATE report TOM ORG save message content Skype ORG text conversation server apparently focus conversation contain political issue
Tibet Falun PERSON Gong Taiwan GPE independence Chinese Communist Party ORG save message contain personally identifiable information
message sender recipient include IP address username landline phone number entire content text message include time date message Information Skype

Text Extraction and NLP: Corpus Prep

Lemma sentences

Resolve coref clusters

Alibaba Group Holding Limited, (also known as Alibaba Group and as Alibaba), is a Chinese multinational conglomerate holding company specializing in e-commerce, retail, Internet, and technology. Founded on 4 April 1999 in Hangzhou, Zhejiang, the company provides consumer-to-consumer (C2C), business-to-consumer (B2C), and business-to-business (B2B) sales services via web portals, as well as electronic payment services, shopping search engines and cloud computing services. It owns and operates a diverse array of businesses around the world in numerous sectors, and is named as one of the world's most admired companies by Fortune.[3][4]

Alibaba Group Holding Limited , (also know as Alibaba Group and as Alibaba) , is a Chinese multinational conglomerate hold company specialize in e - commerce , retail , Internet , and technology . found on 4 April 1999 in Hangzhou , Zhejiang , Alibaba provide consumer - to - consumer (C2C) , business - to - consumer (B2C) , and business - to - business (B2B) sales services via web portals , as well as electronic payment services , shopping search engines an d cloud computing services . Alibaba own and operate a diverse array of businesses around the world in numerous secto rs , and is name as one of the world ' s most admired companies by Fortune.[3][4]

Text Extraction and NLP: Methods

Method 1: Parsing the POS Tree

```
{  
    "after": [  
        "as",  
        ")"  
    ],  
    "before": [  
        "Alibaba Group Holding Limited"  
    ],  
    "prep": [  
        "as Alibaba Group and as Alibaba",  
        "as Alibaba",  
        "in e-commerce , retail , Internet , and technology"  
    ],  
    "verb": "known"  
},  
{  
    "after": [  
        "a Chinese multinational conglomerate holding company"  
    ],  
    "before": [  
        "Alibaba Group Holding Limited"  
    ],  
    "prep": [  
        "as Alibaba Group and as Alibaba",  
        "as Alibaba",  
        "in e-commerce , retail , Internet , and technology"  
    ],  
    "verb": "is"  
},
```

Method 2: Creating noun verb patterns

```
{  
    "after": [  
        "Alibaba Group",  
        "Alibaba"  
    ],  
    "before": [  
        "Alibaba Group Holding Limited"  
    ],  
    "verb": "known"  
},  
{  
    "after": [  
        "Chinese multinational conglomerate holding company"  
    ],  
    "before": [  
        "Alibaba Group",  
        "Alibaba"  
    ],  
    "verb": "is"  
},  
{  
    "after": [  
        "ecommerce",  
        "retail",  
        "Internet",  
        "technology"  
    ],  
    "before": [  
        "Chinese multinational conglomerate holding company"  
    ],  
    "verb": "specializing"  
}
```

Text Extraction and NLP: Methods 2

Method 3: Attributes

```
[  
  ["Alibaba",  
   "is",  
   "a good name"],  
  [  
    ["Alibaba",  
     "is",  
     "a kind, smart business person, and Alibaba helped the village"],  
    [  
      ["Alibaba",  
       "is",  
       "a kind, smart business person, and Alibaba helped the village"],  
      [  
        ["Alibaba",  
         "was",  
         "the biggest US IPO in history, bigger than Google, Facebook, and Twitter combined"],  
        [  
          ["Lazada Group",  
           "is",  
           "a Singaporean e-commerce company founded by Rocket Internet in 2011"]  
        ]]  
    ]]  
]
```

Method 4: Money

```
[  
  [  
    [  
      ["Alibaba's market value"],  
      ["US$231 billion"],  
      [  
        ["closing time",  
         "IPO",  
         "history",  
         "Alibaba's market value"]  
      ]],  
    ]],  
  [  
    [  
      ["stood"],  
      ["352"]  
    ]],  
  [  
    [  
      ["buying"],  
      ["US$1 billion"]  
    ]],  
  ]]  
]
```

Structured Text Extraction

Scraping Info Box, Bottom Infobox, Table of Contents

Alibaba Group Holding Limited	
 Alibaba Group 阿里巴巴集团	
	
"Taobao City", the main corporate campus for	
Alibaba Group at Xixi, Hangzhou.	
Native name	阿里巴巴集团控股有限公司
Type	Public
Traded as	NYSE: BABA SEHK: 9988
ISIN	US01609W1027
Industry	Conglomerate
Founded	4 April 1999; 20 years ago Hangzhou, Zhejiang
Founder	Jack Ma, principal co-founder
Contents [hide]	
1	Naming
2	History
3	Companies and affiliated entities
3.1	E-commerce and retail service platforms
3.2	Cloud computing and artificial intelligence technology
3.3	FinTech and online payment platforms
3.4	Entertainment services
3.5	Internet services
3.6	Others
4	Corporate governance
5	Controversies
5.1	Gold Supplier membership
5.2	Uranium sales
5.3	Counterfeit items and scams
5.4	Class action on IPO
6	See also
7	References
8	External links

V · T · E	Alibaba Group
People	J. Michael Evans · Peng Lei · Jonathan Lu · Jack Ma · Joseph Tsai · Maggie Wei Wu · Daniel Zhang
Services	Alibaba Cloud · AliExpress · AliMusic · AliOS · Alipay · AliGenie · Taobao · Tmall (Tmall Genie) · Xuexi Qiangguo
Subsidiaries	Alibaba Pictures · Amblin Partners (Amblin Entertainment · Amblin Television · DreamWorks Pictures · DreamWorks Animation · Heyi Pictures · Lazada Group · Shenma · South China Morning Post · Tudou · UCWeb (UC Browser) · World Electricity Supply Company

Scraping Info Box, Bottom Infobox, Table of Contents

- The TA scripts were generally helpful for a starting point
- However data was of course still messy and not entirely useful for graph queries
 - Revenue: US\$232.887 billion (2018) converting to 232887000000
 - Employee Count: Employees 1,000–5,000 (July 2016) to 5000
- Fields called different things
 - Employees vs Number of Employee for infobox
 - Table of Content has good data on controversies but key were “Controversy”, “Legal Issues”, “Criticism”, etc

Scraping Info Box, Bottom Infobox, Table of Contents

-Different levels of nesting so had to traverse through the data with that in mind

V*T*E			Amazon
People	Current	Jeff Bezos · Doug Gurr · Tony Hsieh · Werner Vogels	
	Former	Rick Dalzell · Paul Davis · Brian McBride · Christopher North · Nadia Shouraboura · Ram Shriram	
Facilities	Doppler · Day 1 · HQ2 · Spheres		
Products and services	Subsidiaries	A9.com · AbeBooks · Alexa Internet · Amazon Game Studios · Amazon Lab126 · Amazon Robotics · ComiXology · Digital Photography Review · Double Helix Games · Fresh · Goodreads · Graph · Souq.com · Twitch · Woot.com · Zappos	
	Cloud computing	Web Services (AMI · Amazon Aurora · Beanstalk · CloudFront · DynamoDB · EBS · EC2 · EF · Product Advertising API · RDS · Redshift · Rekognition · Route 53 · S3 · SageMaker · SES · Simple Queue Service · Simple Storage Service · Simple Workflow Service)	
	Services	Amazon.com (China) · Alexa · Appstore · Digital Game Store · Drive · Fire OS · Payments · Kindle Store · Marketplace · Music · Silk · Storywriter · Wireless	
	Devices	Cloud Cam · Echo (Show) · Echo Buds · Echo Loop · Fire (Fire HD · Fire HDX) · Fire TV (Stick · Stick+) · Kindle Paperwhite · Kindle Voyage · Kindle Oasis · Kindle Fire · Kindle Fire HD · Kindle Fire HDX · Kindle Fire HDX 8.9 · Kindle Fire 7 · Kindle Voyage 2018 · Kindle Oasis 2017 · Kindle Paperwhite 2018 · Kindle Paperwhite 2013 · Kindle Voyage 2013 · Kindle Oasis 2013 · Kindle Fire HD 2013 · Kindle Fire HDX 2013 · Kindle Fire 2013 · Kindle Paperwhite 2013)	
	Technology	1-Click · Dynamo · Obidos	
	Media	Amazon Publishing · Amazon Studios · Breakthrough Novel Award · Best Books of the Year · Kindle Direct Publishing · Kindle Scout · Kindle Single · Kindle Unlocked · Kindle Unlimited · Kindle Direct Publishing · Kindle Scout · Kindle Single · Kindle Unlocked · Kindle Unlimited	
	Retail	Amazon Books · Amazon Go · Treasure Truck · Whole Foods Market	
	Logistics	Amazon Air · Amazon Prime Air	
	Former	43 Things · Askville · Amapedia · Amie Street (Songza) · CDNow · Dash buttons · Dash wand · LivingSocial · LoveFilm · MobiPocket · PlanetAll · Reflexive Entertainment · Sellaband · Shelfari · Shutterfly · Tivo · Vudu	
Other	Amazon Light · ASIN · Criticism (tax) · Fishbowl · History of Amazon · LibraryThing · List of Amazon brands · List of mergers and acquisitions by Amazon · Locker · MacKenzie Bezos · Perfect 10, Inc. v. Amazon.com, Inc.		

V*T*E			Alibaba Group
People	J. Michael Evans · Peng Lei · Jonathan Lu · Jack Ma · Joseph Tsai · Maggie Wei Wu · Daniel Zhang		
Services	Alibaba Cloud · AliExpress · AliMusic · AliOS · Alipay · AliGenie · Taobao · Tmall (Tmall Genie) · Xuexi Qi		
Subsidiaries	Alibaba Pictures · Amblin Partners (Amblin Entertainment · Amblin Television · DreamWorks Pictures · Dr. Seuss Entertainment · Heyi Pictures · Lazada Group · Shenma · South China Morning Post · Tudou · UCWeb (UC Browser) · WPS Office		

NAICS scraping

- To scrape NAICS, we utilized selenium.
- Started with the base script then had to expand. Based off whether there is code section. If not I had to grab the first company, then the NAICS section from that.

The image shows two screenshots of web interfaces. On the left is the SICCODE.com homepage, featuring a search bar with 'alibaba' and a 'SEARCH' button. Below it, a table lists companies: Alibaba Com Inc (San Mateo, CA), Alibaba Locks and Keys (Alexandria, VA), and Alibaba Restaurant (Memphis, TN). On the right is a search results page for 'amazon'. It has a search bar with 'amazon' and a 'SEARCH' button. Below it, a table lists companies: Amazon.com Inc (Seattle, WA) and Amazon Park (Rochester, NH). Both pages have tabs for 'Companies', 'SIC', and 'NAICS'.

SICCODE.com

DATA SERVICES ▾ WHY SICCODE !

alibaba

SEARCH

Search In: All SIC NAICS Companies More ▾

Companies	
Alibaba Com Inc	San Mateo, CA, 94402
Alibaba Locks and Keys	Alexandria, VA, 22314
Alibaba Restaurant	Memphis, TN, 38134

Matching NAICS Codes

No Results Found

Matching SIC Codes

No Results Found

amazon

SEARCH

Search In: All SIC NAICS Companies More ▾

Companies	
Amazon.com Inc	Seattle, WA, 98109
Amazon Park	Rochester, NH, 03867

Matching NAICS Codes

443142 – Electronics Stores

NAICS scraping

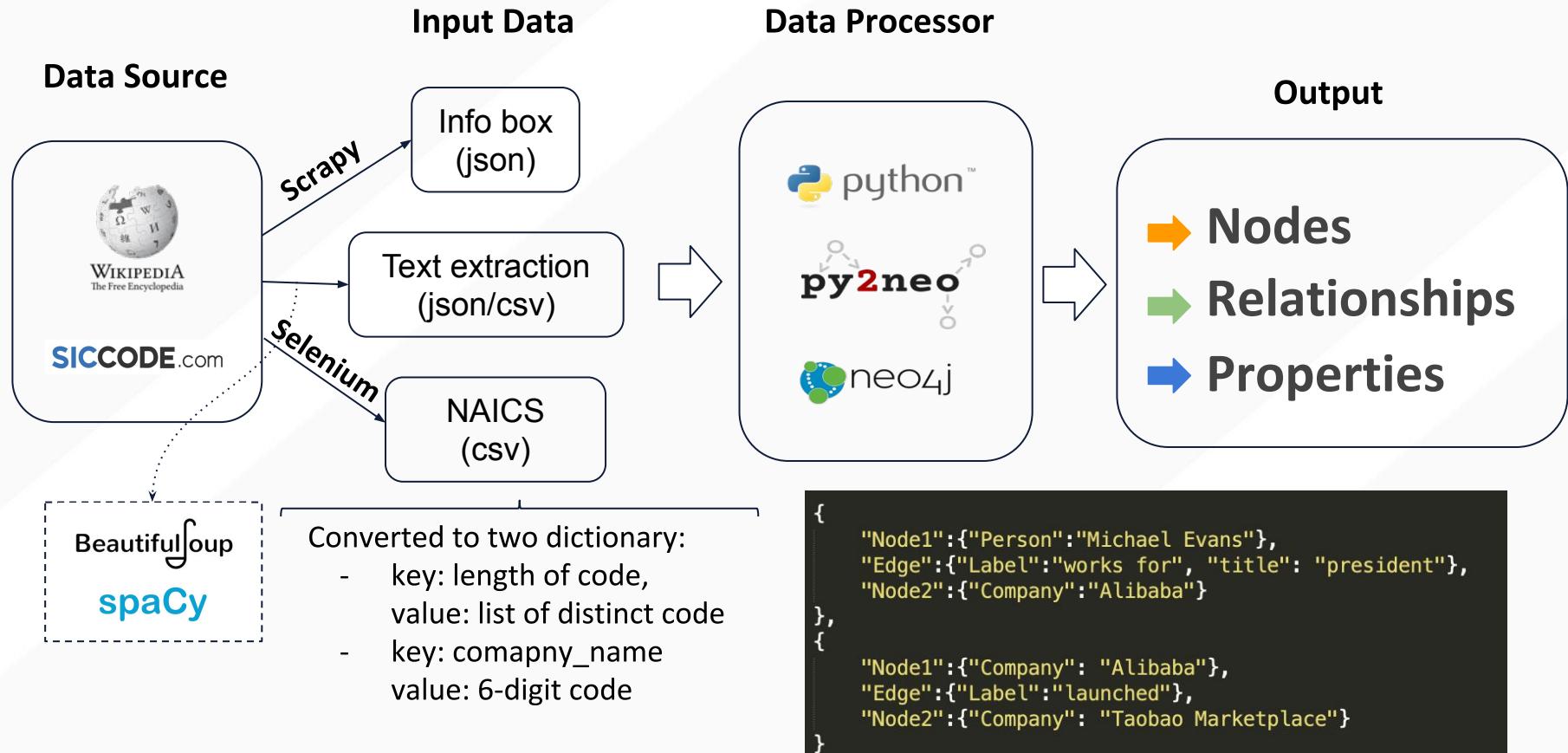
- After that, navigated 4 more times with each iteration removing a code, signifying a wider industry group
- An issue was that clicking on the buttons weren't as reliable as navigating to the urls in the A tag, because of Ads, awkward wait functions, etc

The screenshot shows a web browser window with the URL siccode.com/naics-code/443142/electronics-stores. The page header includes the SICCODE logo and a search bar. Below the header, there are four navigation tabs: 'Description' (underlined), 'Hierarchy', 'Industry Examples', and 'Cross-Ref'. The main content area is titled 'Hierarchy Navigation' and shows the category 'Parent Category' as '44314 - Electronics and Appliance Stores'.

The screenshot shows a web browser window with the URL siccode.com/naics-code/44314/electronics-appliance-stores. The page header includes the SICCODE logo and a search bar. The 'NAICS CODE 44314' and 'NAICS Code 44314 - Ele...' are displayed prominently. Below the header, there are four navigation tabs: 'Description' (underlined), 'Hierarchy', 'Industry Examples', and 'Cross-Ref'. The main content area is titled 'Hierarchy Navigation' and shows the categories 'Parent Category' as '4431 - Electronics and Appliance Stores' and 'Subcategories' as '443141 - Household Appliance Stores'.

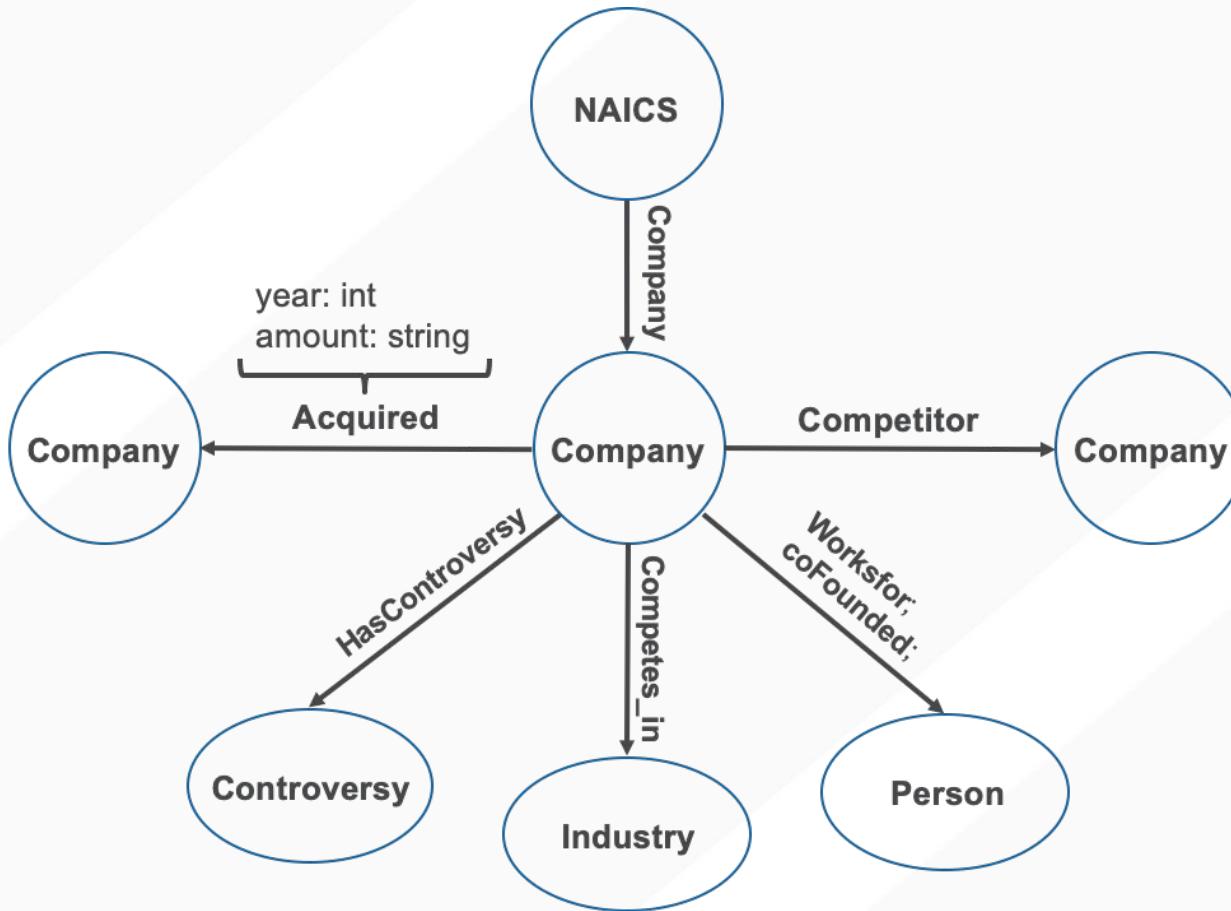
Neo4j Graph Creation

Data Processing

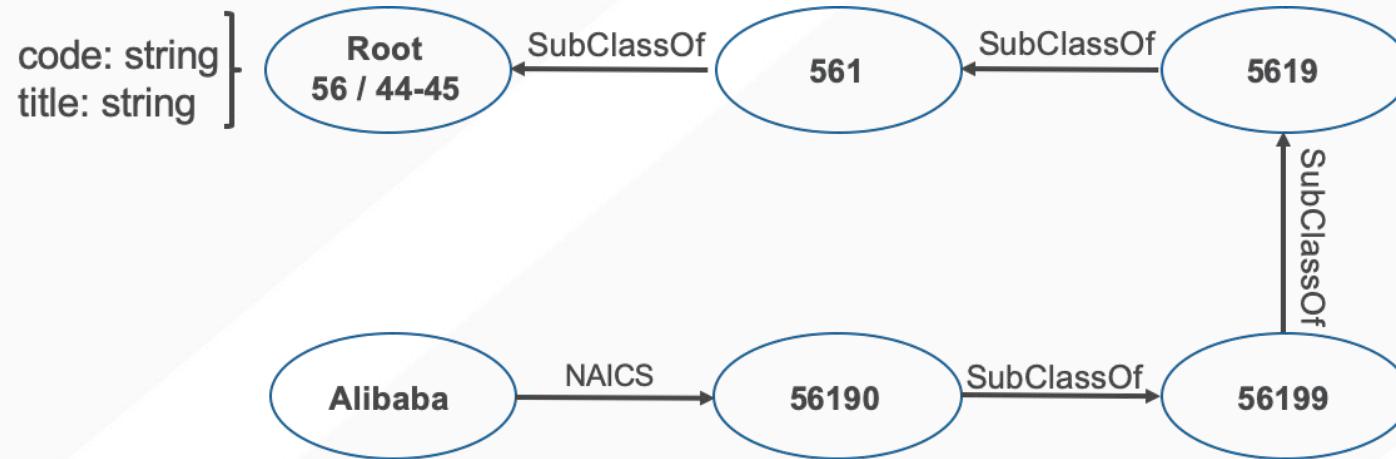


DATA SCHEMA

name: string
type: string
parent: string
founded: string
founder: string
products: string
industry: string
services: string
revenue: string
employees: int
website: string



NAICS CODE STRUCTURE



Company	NAICS_CODE	NAICS_TITLE
Alibaba	561990	All Other Support Services
Alibaba	56199	All Other Support Services
Alibaba	5619	Other Support Services
Alibaba	561	Administrative and Support Services
Alibaba	56	Administrative and Support and Waste Management and Remediation Services

Relationships - Companies

Main_company.json

Main_Company
Alibaba

Alibaba_Compетitor.json

Company	Competitor
Alibaba	eBay
Alibaba	Amazon
Alibaba	Walmart

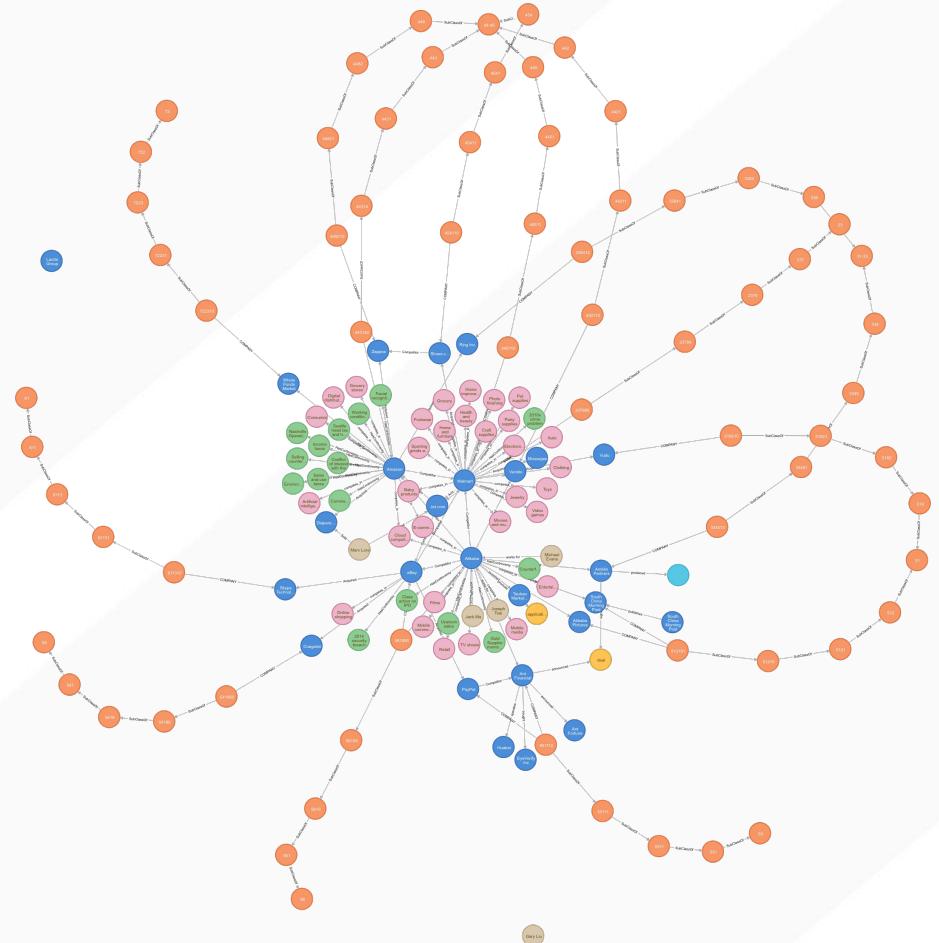
Alibaba_Acquisition.json

Company	Acquisition
Alibaba	Ant Financial
Alibaba	Alibaba Pictures
Alibaba	Amblin Partners
Alibaba	South China Morning Post

Company	Acquisition
eBay	PayPal
eBay	Skype Technologies
eBay	Craigslist
Amazon	Ring Inc.
Amazon	Whole Foods Market
Amazon	Zappos
Walmart	Shoes.com
Walmart	Vudu
Walmart	Jet.com



GRAPH DATABASE



- Company: 31
- NAICS: 69
- Industry: 31
- Controversy: 16
- Person: 5
- Object: 2
- Movie: 1

Queries

Query_1

What industries are in common between acquired company and competitor acquired company?

```
MATCH    (c:Company)-[:Competitor]-(c1:Company)-[:Acquired]->(c3:Company),  
        (c3:Company)-[:COMPANY]-(:NAICS)-[:SubClassOf*4]->(n3:NAICS),  
        (c:Company)-[:Acquired]->(c2:Company),  
        (c2:Company)-[:COMPANY]-(:NAICS)-[:SubClassOf*4]->(n2:NAICS)  
WHERE    n2=n3  
RETURN   c.name as company,  
        collect(distinct n2.code) as common_naics_code,  
        collect(distinct n2.title) as common_industry
```

Result:

"company"	"common_naics_code"	"common_industry"
"Alibaba"	["51", "31-33", "44-45"]	["Information", "Manufacturing", "Retail Trade"]
"Walmart"	["51", "44-45"]	["Information", "Retail Trade"]
"Amazon"	["31-33", "44-45"]	["Manufacturing", "Retail Trade"]

Query_2

Which industry has most controversy?

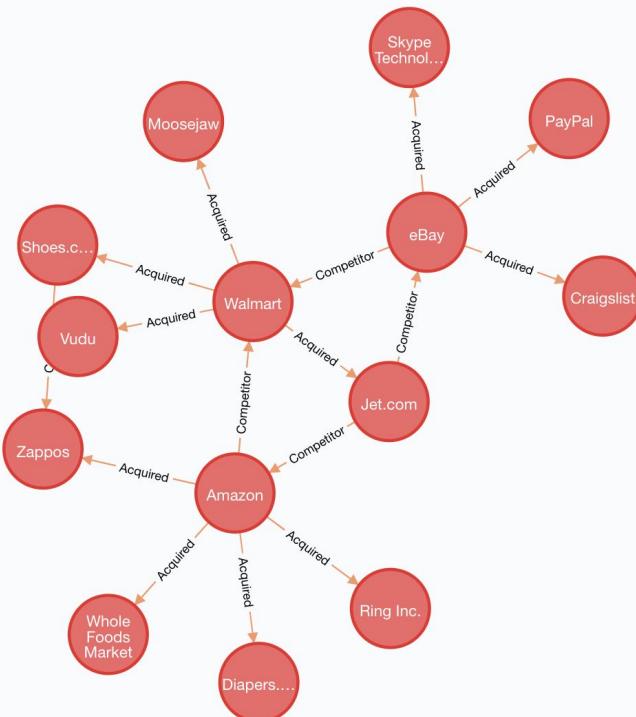
```
MATCH      (c:Company)-[:COMPANY]-(n1:NAICS)-[:SubClassOf*4]->(n2:NAICS),  
          (c:Company)-[:HasControversy]->(con:Controversy)  
RETURN    n2.code,n2.title, size(collect(con))as num_controversy  
ORDER BY  num_controversy desc
```

RESULT:

"n2.code"	"n2.title"	"num_controversy"
"44-45"	"Retail Trade"	11
"56"	"Administrative and Support and Waste Management and Remediation Services"	5
"61"	"Educational Services"	5
"51"	"Information"	4
"55"	"Management of Companies and Enterprises"	2
"72"	"Accommodation and Food Services"	1

Query_3

Who are the international competitors does our company (Alibaba) have?



```
match (ali:Company)  
{name:"Alibaba"}-[e:Competitor]->(c:Company)  
where c.country <> ali.country with c  
match (c) - [r:Acquired *0..5] - (c2:Company)  
return distinct c2.name  
order by c2.name
```

Query_4

What does the average revenue look like for companies grouped by NAICS-2 code?

```
MATCH      (c:Company)-[:COMPANY]-(n1:NAICS)-[:SubClassOf*4]->(n4:NAICS)
RETURN     n4.code as NAICS_2_CODE, n4.title as NAICS_TITLE,
           size(collect(c.name)) as NUM_COMPANY,
           AVG(c.revenue_num) as AVERAGE_REVENUE
ORDER BY   AVERAGE_REVENUE ASC
```

"NAICS_2_CODE"	"NAICS_TITLE"	"NUM_COMPANY"	"AVERAGE_REVENUE"
"54"	"Professional, Scientific, and Technical Services"	1	694000000.0
"61"	"Educational Services"	1	2000000000.0
"55"	"Management of Companies and Enterprises"	2	15451000000.0
"56"	"Administrative and Support and Waste Management and Remediation Services"	2	33449000000.0
"44-45"	"Retail Trade"	5	249764000000.0

Query_5

Do the number of industries a company is in influence how many employees it has?

```
MATCH      (c:Company)-[ci:competes_in]->(pi:Product_Industry_Name)  
WHERE      c.employees IS NOT NULL  
WITH       c, size(collect(pi.product_industry_name)) AS COMPANY_COMPETES_IN_COUNT  
RETURN     COMPANY_COMPETES_IN_COUNT, AVG(c.employee_num) as AVERAGE_EMPLOYEE  
ORDER BY   COMPANY_COMPETES_IN_COUNT ASC
```

"COMPANY_COMPETES_IN_COUNT"	"AVERAGE_EMPLOYEE"
1	733.333333333334
2	52500.0
6	328250.0
7	1300.0
8	101958.0
18	2200000.0

Query_6

Which companies compete on the most product areas?

```
MATCH (c1:Company)-[ci1:competes_in]->(pi:Product_Industry_Name)<-[ci2:competes_in]-(c2:Company)
WHERE id(c1) < id(c2)

WITH c1, c2, size(collect(pi.product_industry_name)) as num_product_compete_in

WITH max(num_product_compete_in) as max_num_product_compete_in

MATCH (c3:Company)-[ci3:competes_in]->(pi2:Product_Industry_Name)<-[ci4:competes_in]-(c4:Company)
WHERE id(c3) < id(c4)

WITH c3, c4, max_num_product_compete_in, size(collect(pi2.product_industry_name)) as NUM_PRODUCTS_COMPETE_IN

WHERE NUM_PRODUCTS_COMPETE_IN = max_num_product_compete_in

RETURN c3.name as COMPANY1, c4.name as COMPANY2, NUM_PRODUCTS_COMPETE_IN
```

"COMPANY1"	"COMPANY2"	"NUM_PRODUCTS_COMPETE_IN"
"Alibaba"	"Amazon"	2

Lessons Learned

Lessons Learned

Get data from different resources and Active learning and ER (minimize human labeling time) . Graph insertion from the entities (eBay vs EBay)

Sentiment and social network analyses

Resolving entities with graph databases using Neo4j

Confirmed that data cleaning is the most time consuming - hard to generalize

Graphs are great for when there are relationships and nodes pointing to other nodes, as it provides a natural way of querying

Lessons Learned

The goals of each company is to increase the productivity. Productivity is based on the net income growth. We are comparing the net income productivity of Alibaba with its competitors that raise the following questions as our hypotheses that will be tested based on our analyses. We want to test if the effects of the services and sells on the productivity. Since we had a limited time, we are not able to answer many questions we have as it is requires more webscrping and data collections.

Questions?

UC San Diego