

A SYSTEMATIC COURSE PLANNER

INF 558 • Fan Pan, I-Hui Huang



- source: <https://open.hpi.de/courses/semanticweb2017>

TO START WITH....

Required Courses (20 units)

- INF 551 Foundations of Data Management Units: 4
- INF 552 Machine Learning for Data Informatics Units: 4
- INF 553 Foundations and Applications of Data Mining Units: 4

One of the following user interface/visualization classes:

- INF 554 Information Visualization Units: 4
- INF 555 User Interface Design, Implementation, and Testing Units: 4
- INF 556 User Experience Design and Strategy Units: 4

One of the following capstone classes:

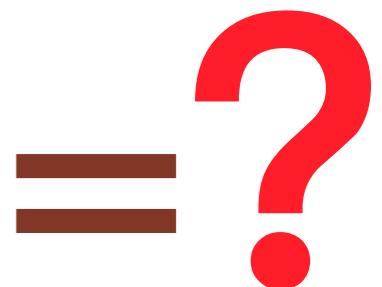
- INF 558 Building Knowledge Graphs Units: 4
- INF 560 Data Informatics Professional Practicum Units: 4

Elective Courses (8 units)

- CSCI 530 Security Systems Units: 4
- CSCI 544 Applied Natural Language Processing Units: 4
- CSCI 548 Information Integration on the Web Units: 4
- CSCI 550 Advanced Data Stores Units: 4
- CSCI 561 Foundations of Artificial Intelligence Units: 4
- CSCI 570 Analysis of Algorithms Units: 4
- CSCI 572 Information Retrieval and Web Search Engines Units: 4
- CSCI 587 Geospatial Information Management Units: 4



Upcoming Starts January 2, 2018	Current Self-Paced	Current Self-Paced
Statistics and Probability in Data Science using Python VERIFIED	Data Science for Executives Learn the foundations of...	Enabling Technologies for Data Science and...
Starting Soon Starts October 24, 2017	Current	Current
Microsoft Professional Program in Data Science Accelerate your career in o...	PURDUE VERIFIED	PURDUE VERIFIED
Current	Starting Soon Starts October 23, 2017	Upcoming Starts March 1, 2018
Big Data Analytics VERIFIED	Microsoft Algorithms and Data Structures VERIFIED	The Science of Partying VERIFIED



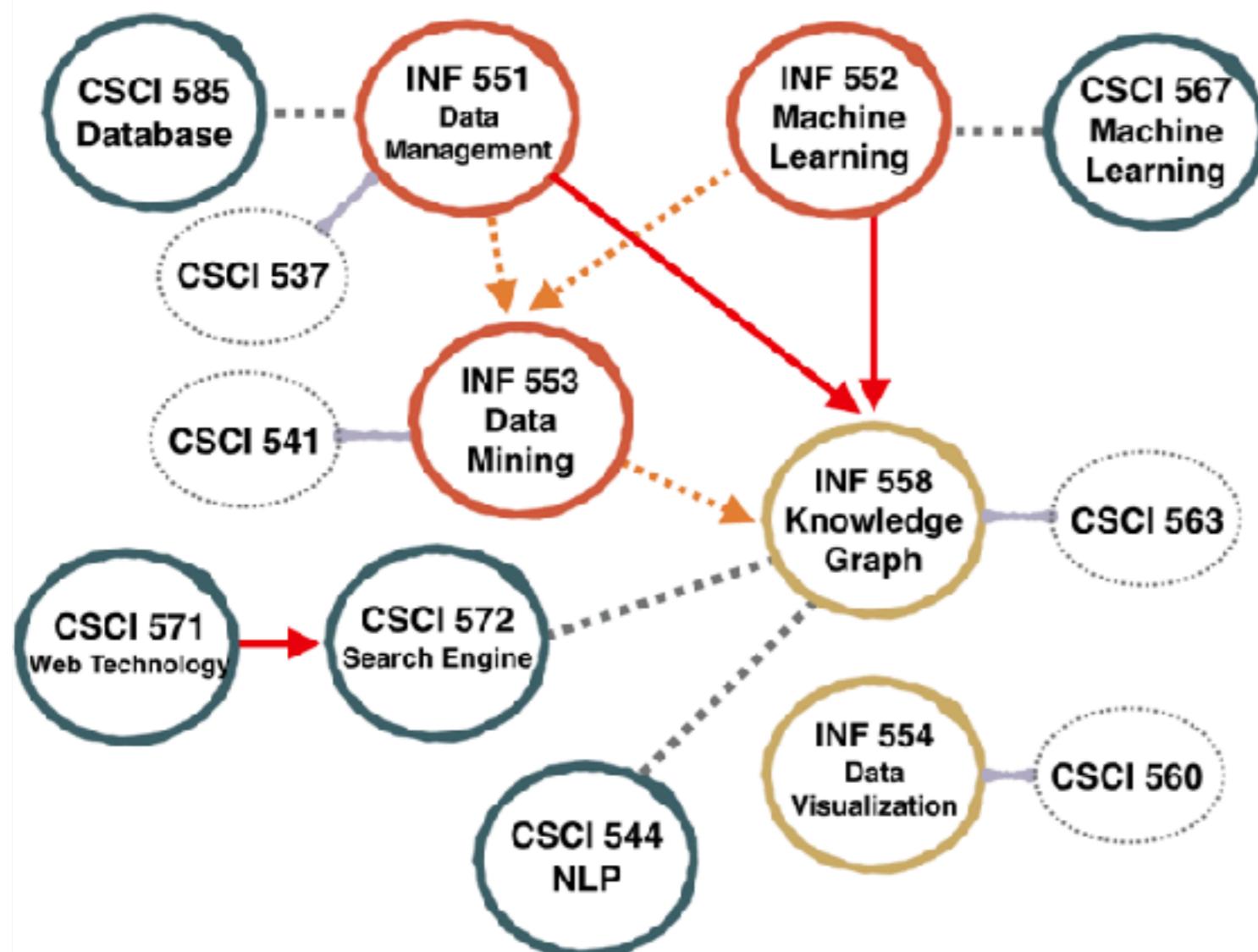
- Systematic, Well-designed

- Easy access, Self-paced,
Abundant resources



COURSE PLANNER

← USC DTIN Program

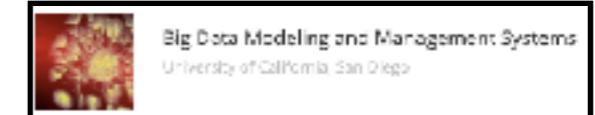


Data Informatics

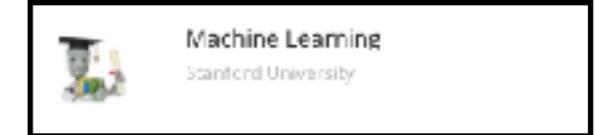


Course Plan

INF551 Data Management



INF552 Machine Learning



INF553 Data Mining



INF554 Data Visualization



INF558 Knowledge Graph



— cross listed

→ recommend preparation

····· text mining, inference

→ prerequisites

DATA SOURCES

USC Catalogue



coursera

 UDACITY

The Udacity logo features a stylized blue "U" icon followed by the word "UDACITY" in a black, sans-serif font.

 KHAN ACADEMY

The Khan Academy logo features a stylized tree with green leaves growing from a hand, followed by the words "KHAN ACADEMY" in a dark blue, sans-serif font.

INFORMATION RETRIEVAL I

USC Catalogue

This university-wide academic program provides a broadly integrative introduction to the breadth of knowledge you will need to consider yourself (and to be considered by other people) a generally well-educated person. This program is effective for all students entering USC in fall 2015 or later, or transfer students beginning college elsewhere at that time and subsequently transferring to USC. It requires eight courses in six Core Literacies, plus two courses in Global Perspectives (which may double-count with courses in the Core Literacies) and two courses in writing. See the General Education for more information.

Business Foundation Requirements

All students in the Leventhal School accounting undergraduate degree program take business foundation courses that focus on necessary analytic skills and theoretical knowledge in math, statistics, accounting and economics. Fundamental knowledge of the functional business disciplines and their integration completes the business core.

Accounting/Business Foundation Courses

BUAD 280 Introduction to Financial Accounting

BUAD 280 Introduction to Financial Accounting



Units: 3

Terms Offered: FaSpSm

Emphasis on understanding financial statements and the impact of business transactions on information presented to management and interested stakeholders.

Duplicates Credit in BUAD 250a, BUAD 250b, BUAD 285a, BUAD 286a, BUAD 286b,

BUAD 305

Instruction Mode: Lecture, Lab

Grading Option: Letter

• BUAD 281 Introduction to Managerial Accounting Units: 3

• BUAD 302T Communication Strategy in Accounting Units: 4

AJAX

base1 = 'http://catalogue.usc.edu/ajax/prev.....'

base2 = '&display_options=a:2:{s:8:~location~.....'

f = requests.get(base1+something+base2)

→ find pattern in HTML

to get target course page

INFORMATION RETRIEVAL II



Waterfall display

Refine your search

Availability	Count
Current	843
Starting Soon	235
Upcoming	138
Self-Paced	995
Archived	725

Subjects

Subject	Count
Architecture	18
Art & Culture	115
Biology & Life Sciences	135
Business & Management	333
Chemistry	43
Communication	75
Computer Science	484

SHOW MORE...

Courses & Programs

Type	Count
All Courses	1809
All Programs	105
Credit-Eligible	17
MicroMasters	41
Professional Certificate	35
Professional Education	64

Verified

Featured Courses

Course	Status
Introduction to Computer Science and Programmi...	Current
Introduction to Project Management	Self-Paced
Curriculum	Curriculum

Featured Programs

Program	Provider
LinuxFoundation® Introduction to Linux	LinuxFoundation
Professional Certificate Program	RTx
Soft Skills	Microsoft

coursera

Catalog > Arts and Humanities

Arts and Humanities

Active Filters: English

History | Music and Art | Philosophy

History

Course	Provider
Journey of the Universe: A Story for Our Times	Yale University
Age of Cathedrals	Yale University

CATEGORIES

- Arts and Humanities
- Business
- Computer Science
- Data Science
- Life Sciences
- Math and Logic
- Personal Development
- Physical Science and Engineering
- Social Sciences

coursera

Need user interaction

DATA EXTRACTION I



Beautiful Soup

by matching html tag

```
'li',{'class':'acalog-course'}
```

by matching XPATH

```
'//div[@class = "gsc-cursor-page"]'
```

by matching regex

```
'[A-Z][A-Z]+[0-9][0-9]+'
```

DATA EXTRACTION II

Program

id
type
url
name

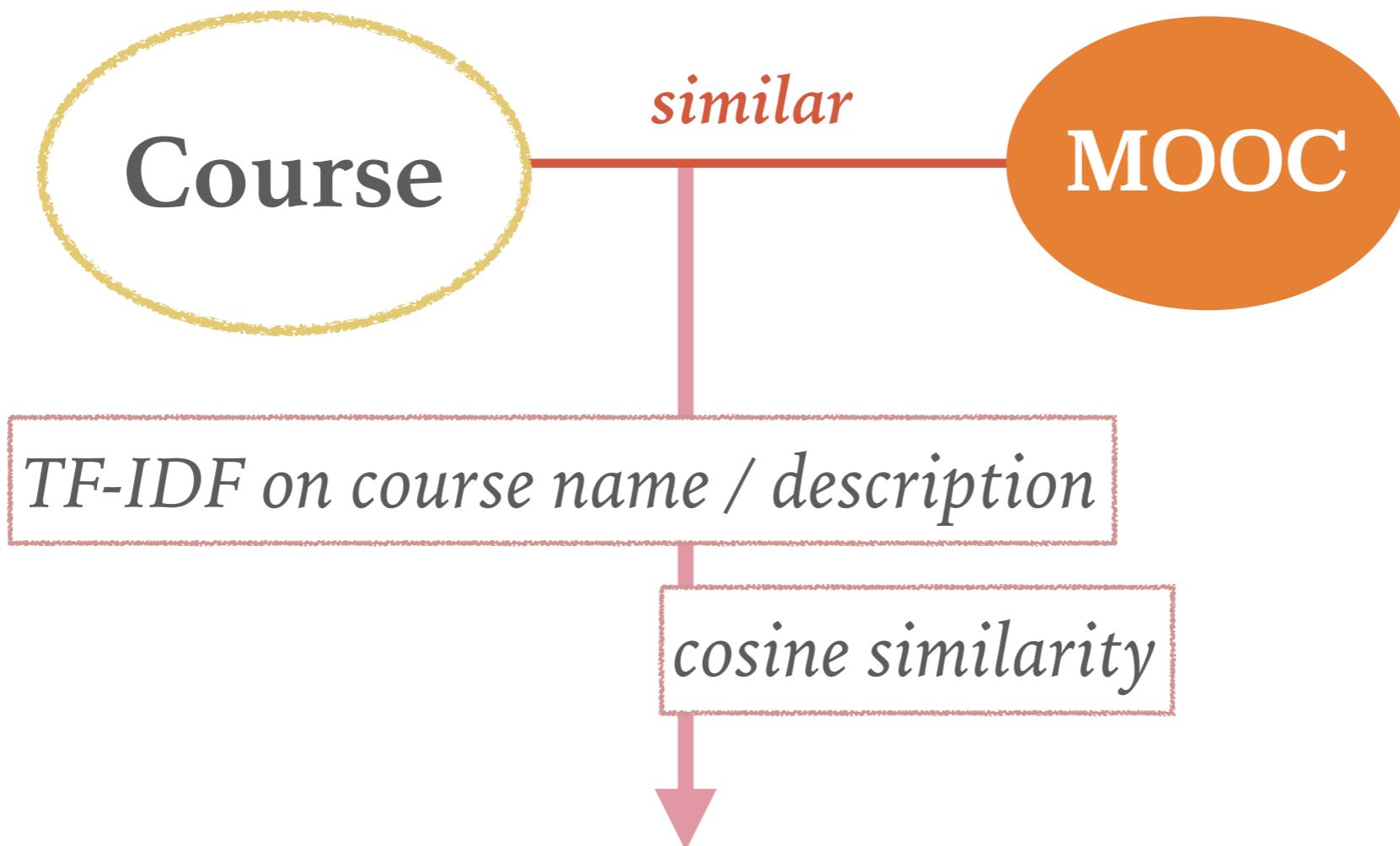
Course

id
name
description
prerequisite
preparation
crosslist
corequisite
duplicate

MOOC

id
name
description
subject
image
url
provenance
special_id

ENTITY LINKING I



$\text{similarity} = 0.5 * (\text{name similarity}) + 0.5 * (\text{description similarity})$

ENTITY LINKING II

Sampling 100 courses.....

topic matched	66
topic related	16
topic unrelated	18

GRAPH BUILDING I

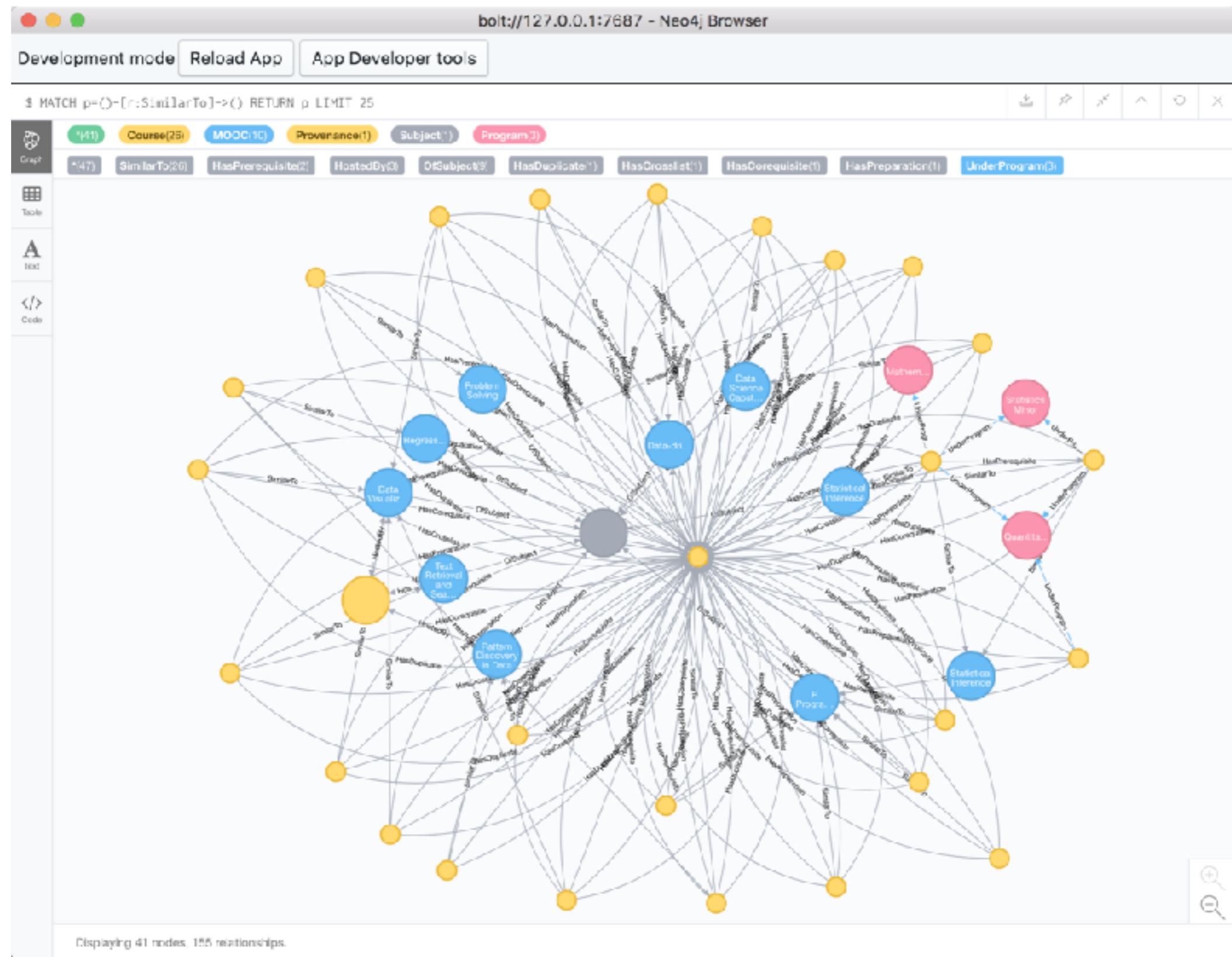
neo4j



LOAD CSV WITH HEADERS FROM...

```
    id:row.id,  
    name:row.name,  
    description:row.description})  
FOREACH (prerequisiteID in split(row.prerequisite,";") |  
    MERGE (pcourse:Course{id:prerequisiteID})  
    MERGE (course) - [:HasPrerequisite] -> (pcourse))  
FOREACH (p IN CASE row.similarity WHEN "" THEN []  
    MERGE (smooc:MOOC{id:split(p,"/")[0]})  
    MERGE (course) - [:SimilarTo{value:split(p,"/")[1]}] -> (smooc));  
...
```

GRAPH BUILDING II



QUERYING THE GRAPH

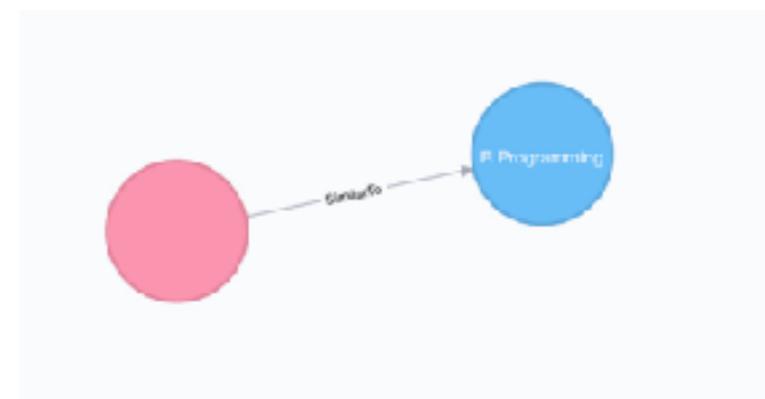
```
$ MATCH (n:Program{name:"Accounting (BS)"}) RETURN n
```

n

```
{  
    "name": "Accounting (BS)",  
    "id": "pgm00000",  
    "type": "Bachelor's Degree",  
    "url":  
        "http://catalogue.usc.edu/preview_program  
catoid=7&poId=6251&returnto=2081"
```

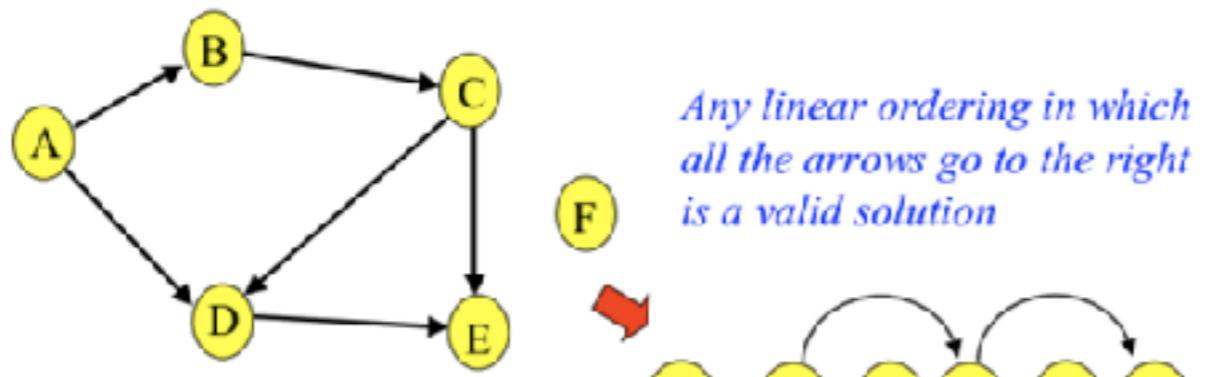
```
import pandas as pd  
for id in toposort[::-1]:  
    dic = {}  
    node = graph.data("Match (n:Course{id:'" + id + "'}) RETURN n")  
    coursename = graph.data("Match (n:Course{id:'" + id + "'}) RETURN n.name")[0]  
    coursedesc = graph.data("Match (n:Course{id:'" + id + "'}) RETURN n.description")[0]  
    mooc = graph.data("MATCH p = (n:Course{id:'" + id + "'})-[r:SimilarTo]->(m:MOOC) RETURN m")  
    image = graph.data("MATCH p = (n:Course{id:'" + id + "'})-[r:SimilarTo]->(m:MOOC) RETURN m")  
    url = graph.data("MATCH p = (n:Course{id:'" + id + "'})-[r:SimilarTo]->(m:MOOC) RETURN m.url")  
    print(mooc)  
    dic["code"] = id  
    dic["name"] = coursename["n.name"]  
    dic["mooc"] = ""  
    dic["url"] = ""  
    dic["image"] = ""
```

Cypher in neo4j

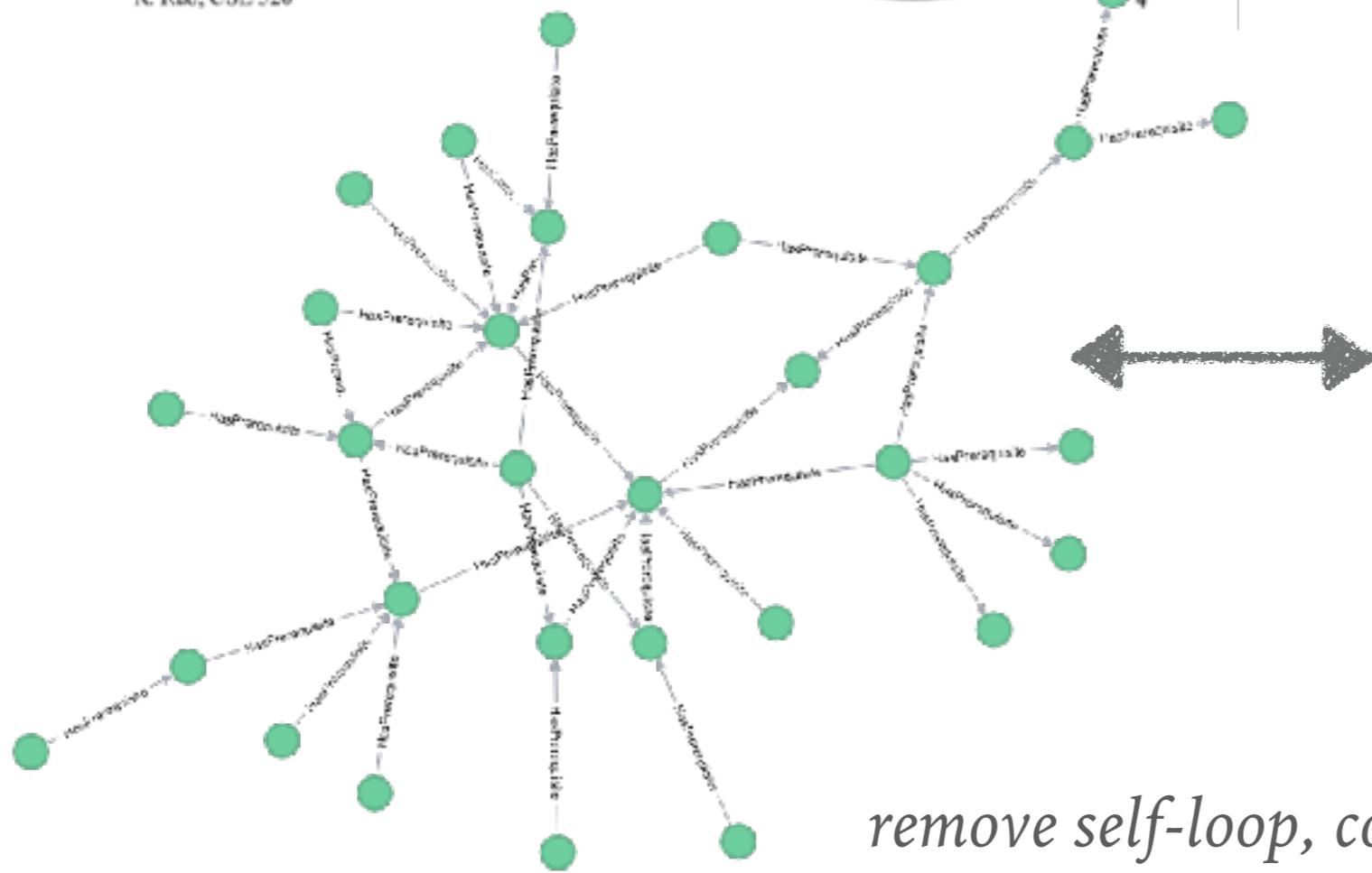


Py2neo

TOPOLOGICAL SORTING OF COURSES



R. Rao, CSE 326



*Sorted Program Order:
For Program of Accounting (BS)*

0	BUAD 280	BUAD 280 Introduction to Financial Accounting	Emphasis on understanding financial statements...	
1	MATH 118	MATH 118gx Fundamental Principles of Calculus	Derivatives ; extrema ; Definite Integral ; fu...	
2	MATH 125		MATH 125g Calculus I	Limits ; continuity , derivatives and applicat...
3	ECON 351	ECON 351x Microeconomics for Business	Development and business applications of : the...	
4	BUAD 308	BUAD 308 Business Finance	Financial problems of business enterprise ; fu...	
5	BUAD 281	BUAD 281 Introduction to Managerial Accounting	An emphasis on how accounting information , co...	
6	ACCT 370	ACCT 370 External Financial Reporting Issues	Understanding of decision-making , problem sol...	
7	ACCT 371	ACCT 371 Introduction to Accounting Systems	Understanding of technology and controls needs...	
8	ACCT 373	ACCT 373 Introduction to Auditing and Assurance	Exploration of the requisite skills and knowle...	
9	ACCT 377	ACCT 377 Valuation for Financial Statement Pur...	Explores Accounting Standards Codification (AS...	
10	ACCT 416	ACCT 416 Financial Reporting and Analysis	Examination of the role of financial statement...	
11	ACCT 462	ACCT 462 Detecting Fraudulent Financial Reporting	Understanding/identifying methods of fraudulen...	
12	ACCT 476	ACCT 476 Performance Measurement Issues	Introduction to understanding how management o...	
13	BUAD 302	BUAD 302T Communication Strategy in Accounting	Theory , practices , and techniques essential ...	
14	BUAD 307	BUAD 307 Marketing Fundamentals	Develops a managerial viewpoint in planning an...	
15	ACCT 477	ACCT 477 Intermediate Fair Value Issues in Acc...	Develops the ability to identify and understan...	
16	ACCT 478	ACCT 478 Accounting Systems Design	Explores the design of accounting systems . In...	
17	ACCT 473	ACCT 473 Financial Statement Auditing	Course builds on the background developed in AC...	
18	ACCT 475	ACCT 475 Systems Security and Audit	Issues related to the security , control , and...	
19	ACCT 483	ACCT 483 Internal Audit	Examination of internal audit/auditors and the...	
20	BUAD 304	BUAD 304 Organizational Behavior and Leadership	The role of leadership in business organizatio...	
21	BUAD 497	BUAD 497 Strategic Management	Examination of managerial decision-making , pl...	
22	ACCT 479	ACCT 479 Accounting Systems Development	Examines the fundamentals of accounting system...	
23	ACCT 374	ACCT 374 Introduction to Tax Issues	Basic tax principles , introduction to U . S ...	

MOOC COVERAGE OF USC COURSES

USC Catalog	881 programs, 16935 courses (5286 distinct)
Coursera	223 specializations, 2597 courses
Other Moocs	1409 Edx 1329 Khan 233 Udacity
covered by MOOC	779
cover ratio	14.7%

INTERACTIVE USER INTERFACE I

```
In [18]: display(searchbox)
```

X Search: Program Name

```
In [19]: display(button)
```

X See Detail!

```
In [ ]:
```

INTERACTIVE USER INTERFACE II

```
In [18]: display(searchbox)
```

✖ Search:

Accounting (BS)

Accounting Minor

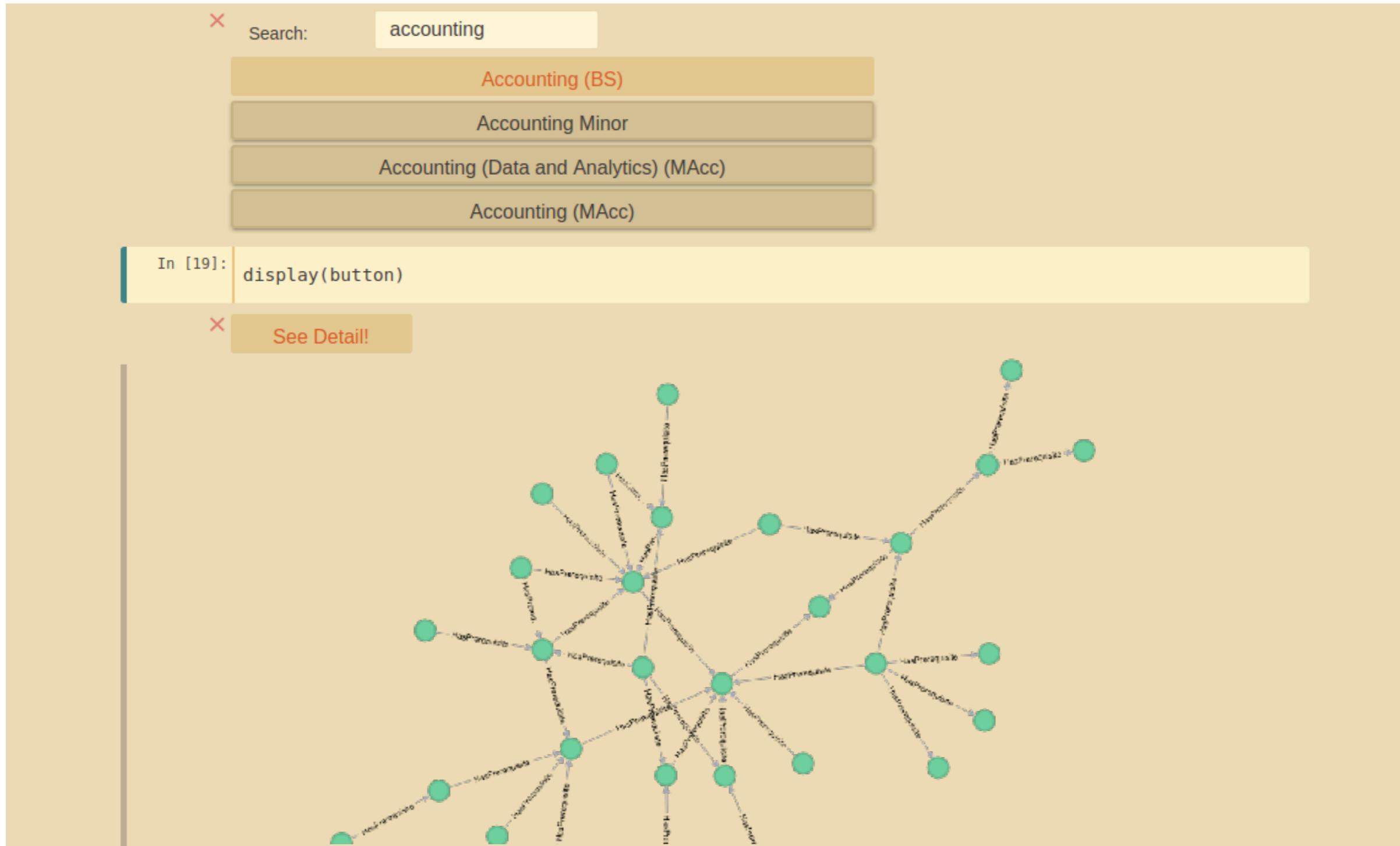
Accounting (Data and Analytics) (MAcc)

Accounting (MAcc)

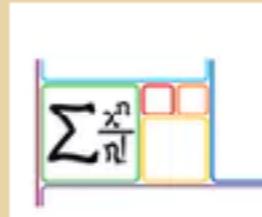
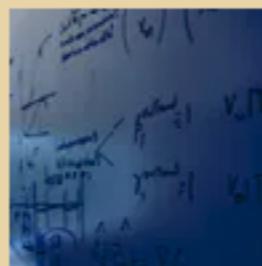
```
In [19]: display(button)
```

✖

INTERACTIVE USER INTERFACE III



INTERACTIVE USER INTERFACE IV

Code	Name	Image	MOOC
BUAD 280	BUAD 280 Introduction to Financial Accounting		Accounting: Principles of Financial Accounting
MATH 118	MATH 118gx Fundamental Principles of Calculus		Single Variable Calculus
MATH 125	MATH 125g Calculus I		Strategic Business Management - Microeconomics
BUAD 306	BUAD 306 Business Finance		Business English: Finance and Economics
BUAD 281	BUAD 281 Introduction to Managerial Accounting		Managerial Accounting: Cost Behaviors, Systems, and Analysis

CHALLENGES

- Some programs don't list full information of curriculum on university websites.
- Some rare topic of courses may not be available on MOOCs.
- Crawling dynamic webpages (Selenium, PhantomJS,Scrapy)
- Hard to learn and distinguish different course level.

