NLP-based Course Clustering and Recommendation

Kentaro Suzuki, Hyunwoo Park

December 10, 2009

Abstract

We have implemented NLP-based UC Berkeley course recommendation system by scoring similarity of courses and clustering courses based on course descriptions. In terms of course scoring, a simple tf-idf scheme and the inner product model is so powerful that we can easily find similar courses for each course. In terms of course clustering, the EM algorithm can effectively cluster courses into several clusters. Each cluster has one or some characteristic(s) and the information of which cluster a course belongs to is useful additional information for users to search the courses which they may be interested in because the cluster information may make up for a downside of a scalar representation of course similarity.

1 Introduction

When students register courses to enroll in, they usually depend on their department's formal advice or their friends' informal recommendation. It is partly because there is no legitimate resource outside with which students search the list of related courses. In other words, it is generally hard to discover a new course which one might be interested in taking only if one knows its existence.

Recognizing this problem, we decided to use Natural Language Processing (NLP) as a tool for generating a set of relevant courses for a given course entry. These lists are to be used for recommending courses regardless of the department the class is offered by.

2 Related Work

One related work to our project is online course catalog. UC Berkeley is currently providing online course schedule[1], but a user cannot search a course based on course description. Instead, meta information such as a department code or class time is the only available source to search with. This system is useful when a user knows some information about the course he/she is looking for. However, when a user is just trying to discover relevant courses, benefits of this system can be very limited. When it comes to cross-departmental course recommendation, such a system cannot provide a sufficient measure to find relevant courses.

3 Data, Features and Models

3.1 Corpus Construction

First, the entire course catalog should be gathered and constructed in order to be used as an NLP corpus to be analyzed. Another online course catalog search interface[2] provided by UC Berkeley was used to scrape course information. The following five information was collected for each course: department, department code, course number, title, description. After removing courses without description, 7004 courses remained.

[1] `http://schedule.berkeley.edu/`

[2] `http://sis.berkeley.edu/catalog/gcc_search_menu`

Table 1: Sample Course Raw Data

Dept Name	Information
Dept Code	INFO
Course No	256
Title	Applied Natural Language Processing
Description	This course examines the state-of-the-art in applied Natural Language Processing (also known as content analysis and language engineering), with an emphasis on how well existing algorithms perform and how they can be used (or not) in applications. Topics include part-of-speech tagging, shallow parsing, text classification, information extraction, incorporation of lexicons and ontologies into text analysis, and question answering. Students will apply and extend existing software tools to text-processing problems.

3.2 Lemmatization

We used course title and description as the target of our NLP analysis. We then tokenized title and description of each course by removing white spaces and punctuations, and making it lowercased. Therefore, the analysis is based on unigrams. One problem we faced after this process was that frequency of each word is actually quite low to be useful. There can be several ways to do so, and we decided to lemmatize each word using NLTK's WordNet morphy() function. Stemming with the Porter's stemmer was an alternative, but we thought it reduced feature too much. On the other side, some words appeared too much. We removed them using a list of stop words. [3]

Table 2: Sample Course Data After Lemmatization

Dept Name	Information
Dept Code	INFO
Course No	256
Title	apply natural language processing
Description	course examine state-of-the-art apply natural language processing also know content analysis language engineering emphasis how well exist algorithm perform how they can use not application topic include part-of-speech tag shallow parse text classification information extraction incorporation lexicon ontology into text analysis question answer student apply extend exist software tool text-processing problem

3.3 Features and Models

3.3.1 Features and Models for Scoring

In order to compute a score of similarity between courses, we use a vector space model and tf-idf weighting scheme. We use $\mathrm{tf}_{t,d}$, the frequency of term t in course description d, as term occurrence parameter, and $\log \frac{N}{\mathrm{df}_t}$, the logarithm document frequency, as document frequency, where N is the total number of courses and df_t is the number of courses t occurs in. tf-idf course vectors $(\mathrm{tf}_{1,d} \times \frac{\mathrm{N}}{\mathrm{df}_1}, \mathrm{tf}_{2,d} \times \frac{\mathrm{N}}{\mathrm{df}_2}, \cdots)$ are normarized and vector similarity is computed by inner product.This method is very common in the field of information retrieval[1]. In terms of scoring, we used course titles as well as course descriptions to improve the results. In that case, the weight between title and description was 1 to 1.

[3]The list of stop words are as follows: a, an, and, are, as, at, be, between, both, by, for, from, have, in, is, it, its, of, on, or, that, the, this, to, will, with.

3.3.2 Features and Models for Clustering

We adopted the bag of words approach to generate a feature for each course. If a course has a word, the element of the course vector will be true, and false otherwise. Namely, the course description representation vector is described as $< e_1, e_2, e_3, ...e_M >$, where $e_i = 1$ if term t_i appears in the course description, otherwise $e_t = 0$. Each boolean vector represents a course. For clustering, multi-gaussian Bernoulli model[2] is assumed. Namely, we assume that each course is generated from one of clusters and the probability of occurrence of a word in each course description is computed based on the distance between each course vector and the center of each cluster.We classify each course to the cluster with the maximum probability of generating the course.

In terms of clustering model, the course description representation vector is described as $< e_1, e_2, e_3, ...e_M >$, where $e_i = 1$ if term t_i appears in the course description, otherwise $e_t = 0$.

We use the EM algorithm to compute clusters[3]. However, in case we start to run the EM algorithm from a randomized initial set of clusters, the EM algorithm needs quite a few iterate computations to find the most appropriate set of clusters. In order to save computational time, we use K-means to find initial state of centroids of clusters.

4 Results

4.1 Scoring

Scoring result seems very good. We have already implemented a search-engine like interface[4] and you can refer to it. For example,"INFO 202 Information Organization and Retrieval" 's top 5 most similar courses are "INFO 245 Organization of Information in Collections", "INFO 211 Group and Organizational Approaches to Information Systems Use", "IND ENG C215 Analysis and Design of Databases", "IND ENG 215 Analysis and Design of Databases" and "INFO C258 Analysis and Design of Databases". These 5 courses are related to information organization and database, which are so closed to the contents of INFO 202. Also, "INFO 205 Information Law and Policy" 's top 5 most similar courses are "INFO 221 Information Policy", "INFO 203 Social and Organizational Issues of Information", "INFO 231 Economics of Information", "INFO 235 Cyberlaw" and "INFO 212 Information in Society". These 5 courses are related to information law and policy as well as social aspects and issues of information. Many students who are interested in INFO 205 are probably interested in these courses.

We examined a quantitive analysis about how good the scoring result was. In terms of I School graduate courses, the average ratio of the similarity of top 3 high similarity scores(i.e. the average of # of similar course/(# of similar courses+# of non-similar courses)) is 76%. Also, the average ratio of the similarity of top 5 high similarity scores is 69%, and when it gets to top 10, the average ratio becomes 57%. [Fig. 1] These scores show the result is useful to find similar courses.

Also we have implemented 2-D network visualization[5]. 2-D visualization was written in Processing and modified some codes from a book[4]. The visualization provides more informative features about network structures of courses that are omitted in terms of search-engine-interface representation. The screenshots of these implementations are included in Appendix A.

[4] http://data.oksure.net/nlp/

[5] http://people.ischool.berkeley.edu/~k_s_/nlp/index.html

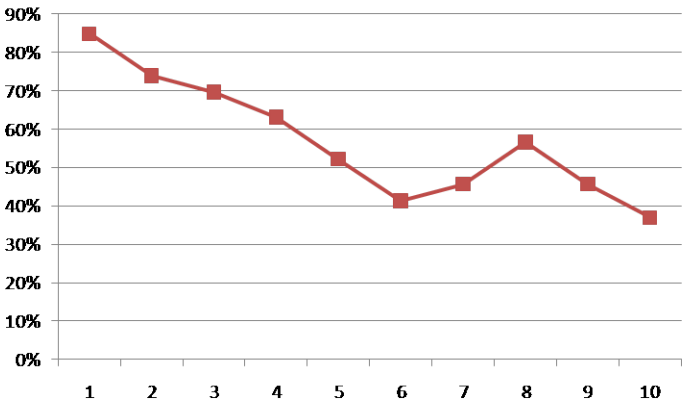


Figure 1: The average ratio of the number of similar courses(First top 10 results of I School graduate courses)

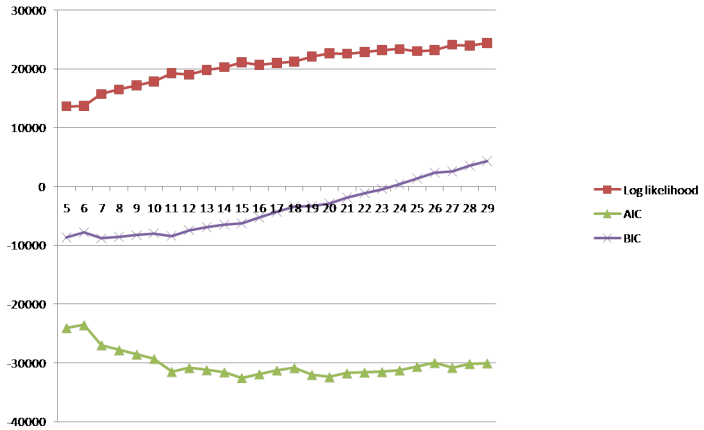


Figure 2: Trend of loglikelihood, AIC and BIC

4.2 Clustering

At first, we intended to compute clusters with all courses. However, we encountered errors such as a memory error and an error derived from inability of computing inverse matrix. The reason why these errors occur is because too many course vectors(i.e. 7004) and too big vocabulary size(i.e. 13457) often cause troubles to the EM algorithm computation with a home-use computer and the Python programming language.

Then, we reduced test data size, i.e. all courses of 5 close-related departments: Information, Computer Science, Mathematics, Statistics and Economics. The number of courses of the 5 departments are 458. We only use the terms that occurs 9 or more than 9 times in all course descriptions. The size of vocabulary is 323.

To find the appropriate number of clusters, we use two criteria: AIC(Akaike Information Criterion) and BIC(Bayesian Information Criterion). [Fig. 2] shows the transition of log-likelihood, AIC and BIC. In terms of AIC, the appropriate number of clusters is 7, and in terms of BIC, the appropriate number of clusters is 15.

When we use 7 clusters to classify 5 department courses, we can find each cluster has some characteristics.

- For example, Cluster 0 contains a lot of seminar courses such as "Freshman Seminar" and "Sophomore Seminar". In UC Berkeley, each department has so many seminar courses and these descriptions are very similar to each other. Cluster 0 also contains some economics courses such as "Political Economics" and "International Economics". These economics courses are less mathematical economics than "Econometrics" or "Micro Economics". We can describe Cluster 0 as "Seminar/Political Econ Cluster".
- Cluster 1 contains many Math department courses. These courses are more related to some particular fields of mathematics such as calculus, algebra and geometry. We can describe Cluster 1 as "Math(especially algebra, calculus and geometry) Cluster".
- Cluster 2 also contains many math courses, but the tendency is different from Cluster1. These math courses are related to more abstract mathematics, such as Lie Group theory, discrete mathematics and set theory. We can label this cluster as "Math(especially abstract mathematics) Cluster".
- Cluster 3 contains a lot of programming courses. We can label this cluster as "Programming course Cluster".
- Cluster 4 tends to contain I School or CS courses. One of the characteristics of the cluster is that the courses in the cluster tend to be more social/human/organization related courses such as "User Interface Design and Management". We may describe the cluster as "Info/CS(especially human/social aspects) Cluster".
- Cluster 5 is also Info/CS-related cluster, but tends to be more hardware/data-structural aspects. In other words, the cluster captures more "traditional" CS courses. We can describe the cluster as "traditional CS Cluster".
- Cluster 6 clearly captures probability-related/statistical courses. We can describe the cluster as "Probability/Stat Cluster".

Of course, there are several courses that cannot be suitable for the labels of clusters described above. 7 clusters may not be enough to precisely classify all courses. However, using too many clusters to classify courses will result in the situation such that "one cluster has only one course", which is meaningless. More detail results of clustering are attached in Appendix B.

We also analyze the result of course clustering based on 15 clusters. We can find 15 cluster is finer classification in detail. More detail results of clustering are attached as Appendix C.

These clustering information can be used as additional information of the result of similarity scoring. For example, the scoring result of "INFO256 Applied Natural Language Processing" contains a lot of courses generally considered non-related to NLP, such as "Metal Processing" and "Postbiblical Hebrew Text Analysis". However, we cannot differentiate these courses if we only use the similarity scoring with a scalar value of similarity. Clustering information can be used to reduce the list of similar courses by excluding totally different courses belonging to totally different clusters. Also, the cluster information can be used as an auto-generated tag for each course. Auto-generated tag can enable users easily to organize which courses the users want to see/hide.

5 Conclusion

We have implemented NLP-based UC Berkeley course recommendation system by scoring similarity of courses and clustering courses with course descriptions. In terms of course scoring, a simple tf-idf scheme and the inner product model is so useful that we can easily find similar courses for each course. Also, search

ID	Words that the cluster is most likely to generate	The cluster can be labeled as...	Samples of courses
0	topic, economics, economic, seminar, policy, department, semester, change, cover, politics, faculty, vary, political, market, special, offer ...	Seminar/Political Econ Cluster	Info24 Freshman seminar CS84 Sophomore Seminar Econ215A Political Economics Econ280B International Economics
1	theorem, function, equation, space, differential, algebra, integral, series, partial, geometry, surface, finite, riemann, algebraic, transformation,...	Math(esp. algebra, calculus and geometry) Cluster	Math 54: Linear Algebra and Differential Equations Math 222A:Partial Differential Equations
2	sequence, their, fields, field, lie, set, fall, begin, number, relations, price, class, labor, construction...	Math(esp. abstract math Cluster	Math 261B: Lie Group Math 235A: Theory of Sets
3	course, student, science, introduction, research, it, program, mathematical, instructor, how, may, but, provide, mathematics, who, work , unit, under,...	Programming course Cluster	CS9F:C++ for Programmers CS9G:JAVA for Programmers CS9H:Python for Programmers
4	design, information, group, study, computer, use, issue, management, organization, project, technology, development, implementation, new...	Info/CS(especially human/social aspects) Cluster	CS169: User Interface Design and Development INFO214: Needs and Usability Assessment
5	system, datum, programming, technique, language, structure, algorithm, network, advance, control, engineering, search, performance...	Traditional CS Cluster	CS61B:Data Structure INFO206: Distributed Computing Application and Infrastructure INFO257: Database Management
6	probability, estimation, random, variable, regression, statistical, markov, statistics...	Probability/Stat Cluster	STAT205B Probability Theory CS281A:Statistical Learning Theory

Figure 3: The summary of course clustering (7 clusters)

engine like interface and 2-D network model may help students to choose their interest courses. In terms of course clustering, the EM algorithm can automatically cluster courses into several clusters. Each cluster has one or some characteristic(s) and the information of which cluster a course belongs to is useful additional information for users to search the courses which they are interested in because the cluster information may make up for a downside of a scalar representation of course similarity. However, clustering costs a lot of time and memory to compute. All course clustering is difficult by home PC and with a script language like Python, so that we need more powerful PCs or sophisticated algorithms that reduces computational time and complexity. To expand our service, we can further develop a system (1) saving the courses a user has taken or is interested in, (2) recommending a set of courses based on the given list. In terms of clustering, we can literally expand the boundary by clustering courses from all departments instead of five of them.

References

[1] C. D. Manning and H. Schütze, *Foundation of Statistical Natural Language and Processing*,MIT Press, pp540-542

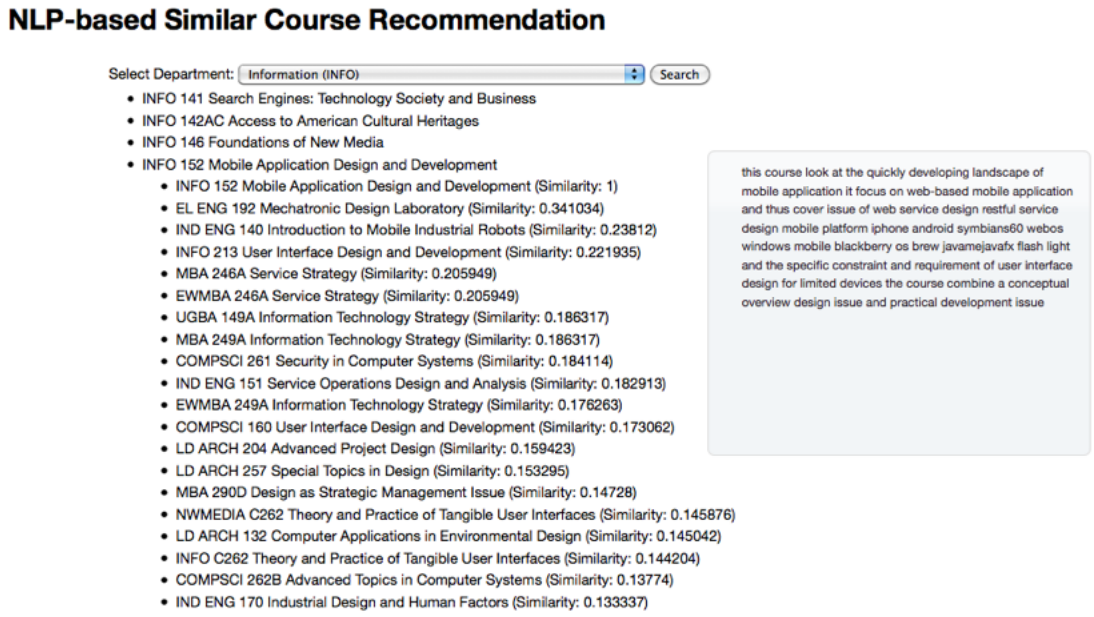
[2] C. D. Manning et al., *Introduction to Information Retrieval*, Cambridge University Press, pp245-248

[3] C. D. Manning et al., *Introduction to Information Retrieval*, Cambridge University Press, pp338-340

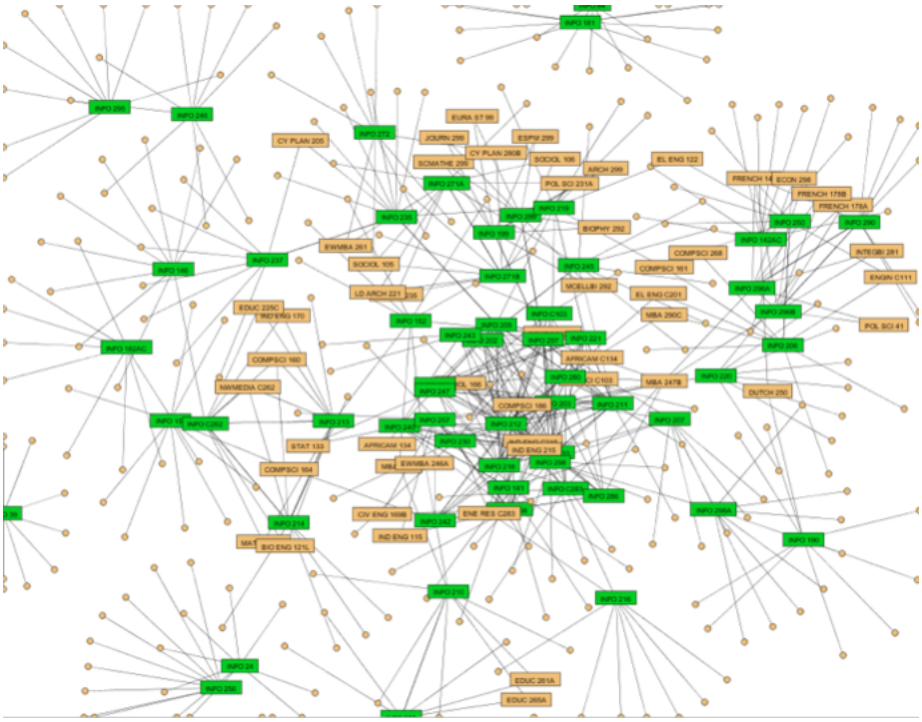
[4] Ben Fry, *Visualizing Data*, O'Reilly Media

A Application Screenshots

A.1 Recommendation Engine



A.2 2-D Network Visualization



Appendix B: Detailed Result of 7 cluster model

Cluster 0's courses(excerpt)

COMPSCI	24	Freshman Seminars
COMPSCI	39	Freshman/Sophomore Seminar
COMPSCI	84	Sophomore Seminar
ECON	2	Introduction to Economics--Lecture Format
ECON	24	Freshman Seminar
ECON	104	Advanced Microeconomic Theory
ECON	119	Psychology and Economics
ECON	131	Public Economics
ECON	138	Financial and Behavioral Economics
ECON	152	Wage Theory and Policy
ECON	215A	Political Economics
ECON	215B	Political Economics
ECON	219A	Foundations of Psychology and Economics
ECON	234A	Macroeconomic Finance
ECON	280B	International Economics
ECON	280C	International Economics
ECON	C102	Natural Resource Economics
ECON	C125	Environmental Economics
ECON	C215A	Political Economics
ECON	C215B	Political Economics
INFO	24	Freshman Seminar
INFO	39	Freshman/Sophomore Seminar
INFO	84	Sophomore Seminar
INFO	296A	Seminar
INFO	296B	Seminar
MATH	24	Freshman Seminars
MATH	84	Sophomore Seminar
MATH	197	Field Study
MATH	C290C	Topics in Fluid Mechanics
STAT	39	Freshman/Sophomore Seminar

Cluster 1's courses(excerpt)

MATH	1A	Calculus
MATH	16A	Analytic Geometry and Calculus
MATH	53M	Multivariable Calculus with Computers
MATH	54	Linear Algebra and Differential Equations
MATH	H54	Honors Linear Algebra and Differential Equations
MATH	54M	Linear Algebra and Differential Equations with Computers
MATH	104	Introduction to Analysis
MATH	105	Second Course in Analysis
MATH	128B	Numerical Analysis
MATH	170	Mathematical Methods for Optimization
MATH	189	Mathematical Methods in Classical and Quantum Mechanics
MATH	202A	Introduction to Topology and Analysis
MATH	202B	Introduction to Topology and Analysis
MATH	205	Theory of Functions of a Complex Variable
MATH	206	Banach Algebras and Spectral Theory
MATH	220	Introduction to Probabilistic Methods in Mathematics and the Sciences
MATH	221	Advanced Matrix Computations
MATH	222A	Partial Differential Equations
MATH	222B	Partial Differential Equations
MATH	228A	Numerical Solution of Differential Equations
MATH	228B	Numerical Solution of Differential Equations
MATH	258	Classical Harmonic Analysis
STAT	251	Stochastic Analysis with Applications to Mathematical Finance

Cluster 2's courses(excerpt)

MATH	55	Discrete Mathematics
MATH	113	Introduction to Abstract Algebra
MATH	235A	Theory of Sets
MATH	114	Second Course in Abstract Algebra
MATH	215B	Algebraic Topology
MATH	215A	Algebraic Topology
MATH	254B	Number Theory
MATH	254A	Number Theory
MATH	245A	General Theory of Algebraic Structures
MATH	261B	Lie Groups
MATH	261A	Lie Groups
MATH	256A	Algebraic Geometry
MATH	256B	Algebraic Geometry
MATH	227A	Theory of Recursive Functions
MATH	135	Introduction to the Theory of Sets
MATH	225B	Metamathematics
MATH	225A	Metamathematics
MATH	224B	Mathematical Methods for the Physical Sciences
MATH	224A	Mathematical Methods for the Physical Sciences
MATH	290	Seminars
MATH	127	Mathematical and Computational Methods in Molecular Biology
MATH	240	Riemannian Geometry
MATH	236	Metamathematics of Set Theory
MATH	142	Elementary Algebraic Topology
ECON	260B	Comparative Economics
ECON	260A	Comparative Economics
ECON	206	Mechanism Design and Agency Theory
ECON	209A	Theory and Application of Non-Cooperative Games
ECON	136	Financial Economics
ECON	136	Financial Economics
COMPSCI	172	Computability and Complexity

Cluster 3's courses(excerpt)

COMPSCI	3	Introduction to Symbolic Programming
COMPSCI	3L	Introduction to Symbolic Programming
COMPSCI	3S	Introduction to Symbolic Programming (Self-Paced)
COMPSCI	9A	Matlab for Programmers
COMPSCI	9B	Pascal for Programmers
COMPSCI	9C	C for Programmers
COMPSCI	9D	Scheme and Functional Programming for Programmers
COMPSCI	9F	C++ for Programmers
COMPSCI	9G	JAVA for Programmers
COMPSCI	9H	Python for Programmers
COMPSCI	47A	Completion of Work in Computer Science 61A
COMPSCI	47B	Completion of Work in Computer Science 61B
COMPSCI	47C	Completion of Work in Computer Science 61C
COMPSCI	61A	The Structure and Interpretation of Computer Programs
COMPSCI	C191	Quantum Information Science and Technology
COMPSCI	H196A	Senior Honors Thesis Research
COMPSCI	H196B	Senior Honors Thesis Research
ECON	C103	Introduction to Mathematical Economics
ECON	C142	Applied Econometrics and Public Policy
ECON	C181	International Trade
ECON	204	Mathematical Tools for Economics
ECON	C225	Workshop in Institutional Analysis
ECON	230B	Public Economics
ECON	241A	Econometrics
INFO	141	Search Engines: Technology, Society, and Business
INFO	181	Technology and Poverty
INFO	205	Information Law and Policy
INFO	209	Professional Skills Workshop
INFO	256	Applied Natural Language Processing
INFO	298A	Directed Group Work on Final Project
MATH	74	Transition to Upper Division Mathematics
MATH	C103	Introduction to Mathematical Economics
MATH	121A	Mathematical Tools for the Physical Sciences
MATH	121B	Mathematical Tools for the Physical Sciences
MATH	300	Teaching Workshop
STAT	C245E	Statistical Genomics
STAT	C245F	Statistical Genomics
STAT	296	Resources for Statistical Computing

Cluster 4's courses(excerpt)

COMPSCI	61C	Machine Structures
COMPSCI	150	Components and Design Techniques for Digital Systems
COMPSCI	160	User Interface Design and Development
COMPSCI	169	Software Engineering
COMPSCI	252	Graduate Computer Architecture
COMPSCI	289	Knowledge Representation and Use in Computers
COMPSCI	302	Designing Computer Science Education
ECON	222	Economics of Innovation
ECON	C222	Economics of Innovation
INFO	C103	History of Information
INFO	146	Foundations of New Media
INFO	152	Mobile Application Design and Development
INFO	182AC	Print, Literacy, and Power in America to 1900
INFO	202	Information Organization and Retrieval
INFO	203	Social and Organizational Issues of Information
INFO	210	The Information and Services Economy
INFO	214	Needs and Usability Assessment
INFO	216	Computer-Mediated Communication
INFO	218	Concepts of Information
INFO	237	Intellectual Property Law for the Information Industries
INFO	245	Organization of Information in Collections
INFO	246	Multimedia Information
INFO	247	Information Visualization and Presentation
INFO	C258	Analysis and Design of Databases
INFO	C262	Theory and Practice of Tangible User Interfaces
INFO	280	Information and Communication Technologies and Development: Context, Strategies and Impacts
INFO	C283	Information and Communications Technology for Development
INFO	286	Information, Innovation and International Development (I3D): Designing Rural Computing Applications
STAT	151A	Linear Modelling: Theory and Applications
STAT	151B	Linear Modelling: Theory and Applications

Cluster 5's courses(excerpt)

COMPSCI	61B	Data Structures
COMPSCI	C149	Introduction to Embedded Systems
COMPSCI	161	Computer Security
COMPSCI	162	Operating Systems and System Programming
COMPSCI	170	Efficient Algorithms and Intractable Problems
COMPSCI	C182	The Neural Basis of Thought and Language
COMPSCI	184	Foundations of Computer Graphics
COMPSCI	186	Introduction to Database Systems
COMPSCI	188	Introduction to Artificial Intelligence
COMPSCI	258	Parallel Processors
COMPSCI	261	Security in Computer Systems
COMPSCI	262A	Advanced Topics in Computer Systems
COMPSCI	262B	Advanced Topics in Computer Systems
COMPSCI	265	Compiler Optimization and Code Generation
COMPSCI	266	Introduction to System Performance Analysis
COMPSCI	C267	Applications of Parallel Computers
COMPSCI	268	Computer Networks
COMPSCI	270	Combinatorial Algorithms and Data Structures
COMPSCI	273	Foundations of Parallel Computation
COMPSCI	274	Computational Geometry
COMPSCI	285	Solid Free-Form Modeling and Fabrication
COMPSCI	286	Implementation of Data Base Systems
COMPSCI	287	Advanced Robotics
ECON	202A	Macroeconomic Theory
ECON	202B	Macroeconomic Theory
INFO	206	Distributed Computing Applications and Infrastructure
INFO	219	Privacy, Security, and Cryptography
INFO	242	XML Foundations
INFO	250	Computer-Based Communications Systems and Networks
INFO	257	Database Management
STAT	215A	Statistical Models: Theory and Application
STAT	215B	Statistical Models: Theory and Application
STAT	244	Statistical Computing
STAT	133	Concepts in Computing with Data

Cluster 6's courses(excerpt)

COMPSCI	70	Discrete Mathematics and Probability Theory
COMPSCI	C280	Computer Vision
COMPSCI	C281A	Statistical Learning Theory
COMPSCI	C281B	Advanced Topics in Learning and Decision Making
ECON	C110	Game Theory in the Social Sciences
ECON	140	Economic Statistics and Econonmetrics
ECON	141	Econometric Analysis
ECON	240B	Econometrics
MATH	C218A	Probability Theory
MATH	C218B	Probability Theory
MATH	C223A	Stochastic Processes
MATH	C223B	Stochastic Processes
STAT	20	Introduction to Probability and Statistics
STAT	21	Introductory Probability and Statistics for Business
STAT	25	Introduction to Probability and Statistics for Engineers
STAT	134	Concepts of Probability
STAT	C141	Statistics for Bioinformatics
STAT	C143	Introduction to Statistical Methods in Computational and Genomic Biology
STAT	200A	Introduction to Probability and Statistics at an Advanced Level
STAT	200B	Introduction to Probability and Statistics at an Advanced Level
STAT	204	Probability for Applications
STAT	C205A	Probability Theory
STAT	205B	Probability Theory
STAT	C205B	Probability Theory
STAT	206A	Stochastic Processes
STAT	C206A	Stochastic Processes
STAT	206B	Stochastic Processes
STAT	C206B	Stochastic Processes
STAT	210A	Theoretical Statistics
STAT	210B	Theoretical Statistics
STAT	212A	Topics in Theoretical Statistics
STAT	230A	Linear Models
STAT	C239A	The Statistics of Causal Inference in the Social Science
STAT	C241A	Statistical Learning Theory
STAT	C241B	Advanced Topics in Learning and Decision Making
STAT	C245A	Biostatistical Methods: Advanced Categorical Data Analysis
STAT	C245B	Biostatistical Methods: Survival Analysis and Causality
STAT	C245C	Biostatistical Methods: Computational Statistics with Applications in Biology and Medicine
STAT	C247C	Longitudinal Data Analysis
STAT	C249A	Censored Longitudinal Data and Causality
STAT	C249C	Multiple Testing and Loss Function Based Estimation: Applications in Biological Sciences
STAT	C261	Quantitative/Statistical Research Methods in Social Sciences
STAT	2	Introduction to Statistics
STAT	152	Sampling Surveys
STAT	135	Concepts of Statistics
STAT	150	Stochastic Processes
STAT	248	Analysis of Time Series

Appendix C: Result of 15 cluster model

	Words that the cluster is most likely to generate	The cluster may mean…
0	Politics, political, special, interest, institution, tool, setting, experience, growth, dynamic, standard, aspect, develop, alternative, supervise	Politics
1	Problem, semester, instructor, advance, case, current, determination, chosen, labor, each , see	Seminar
2	Theory, emphasis, cover, time, hypothesis, price, various, game, phd, modern, capital, preparation, choice, least	Econ
3	Sequence, set, geometry, fall, begin, number, algebraic , relations, class, operator, operations, logic,reform	Math(logic)
4	Policy, management, economy, resource, government, environmental, fiscal, environment, level, finance, production, american	Politics
5	Topic, theorem, group, issue, basic, such, algebra, their, fundamental, fields, representation, lie, additional, classification, surface, finite, transformation, manifold, etc, curve, construction, object	Math(fields thory)
6	Equation, space, differential, integral, series, partial, calculus, fourier, analytic, line, solution, vector, ordinary, integration, compact, matrix, boundary, value , euclidean, transform, nonlinear	Math(calculus, liner algebra)
7	System, introduction, computer, programming, language, problem, project, network, how, control, implementation, who, security, flow, computation, interface, memory, code, software, scheme, arithmetic , digital, array, protocol, self-paced, query, storage, know	CompSci

Cluster ID	Words that the cluster is most likely to generate	The cluster may mean…
8	Course, information, it, practice, market, not, but, empirical, focus, form, between, specific, role, these, both, explore, individual, international, behavior, present, can, write, apply, protection, introduce, trade, formal, developing, treatment, into	Econ
9	Student, seminar, department, faculty, vary, offer, may, require, work, unit, all, under, member, credit, take, must, opportunity, paper, limited, enrollment, campus, thesis	Seminar
10	Select, development, sophomore, division, freshman, teaching, lower	Teaching
11	Include, method, also, study, list, science, use, organization, technology, survey, property, intellectual, new, field, other, social , relate, theoretical, computational, evaluation, inference, using, decision, distribute, interaction, approach, history, technical, user, service, understanding, discussion, medium, learning	Social matter
12	Analysis, model, design, datum, technique, linear, structure, algorithm, statistics, statistical, public, regression, search, performance, testing, machine, planning, architecture, database, type, base, effects, experiment, eg, numerical, parallel, computing, selection	Statistics?
13	Application, economics, economic, research, mathematical, engineering, provide, health, intend, one, physical, which, graduate, section, second, country, question, area, services	Econ
14	Function, process, year, probability, estimation, change, markov, general, variable, random, concept, mathematic, distribution, convergence, expectation, content, measure, numbers, knowledge, limit, large, characteristic, riemann, central, further, sample, some, martingale, example, principle, andor, diffusion, chains, motion, real, strong, need, background, discrete, interval, brownian, poisson, gaussian, laws, metric, elementary	Probabilistic

Cluster 0's courses(excerpt)

MATH	197	Field Study
ECON	1	Introduction to Economics
ECON	100A	Economic Analysis--Micro
ECON	100A	Economic Analysis--Micro
ECON	151	Labor Economics
ECON	162	The Chinese Economy
ECON	182	International Monetary Economics
ECON	210A	Introduction to Economic History
ECON	210B	Topics in European Economic History
ECON	210C	Topics in American Economic History
ECON	215A	Political Economics
ECON	C215A	Political Economics
ECON	215B	Political Economics
ECON	C215B	Political Economics
ECON	224	Economics of Institutions
ECON	270C	Development Economics
ECON	C275A	Economic Demography
STAT	25	Introduction to Probability and Statistics for Engineers
STAT	97	Field Study in Statistics
STAT	197	Field Study in Statistics
STAT	240	Nonparametric and Robust Methods
COMPSCI	160	User Interface Design and Development
COMPSCI	264	Implementation of Programming Languages
INFO	213	User Interface Design and Development
INFO	245	Organization of Information in Collections
INFO	285	Design of Library Services

Cluster 1's courses(excerpt)

MATH	189	Mathematical Methods in Classical and Quantum Mechanics
MATH	191	Experimental Courses in Mathematics
MATH	251	Ring Theory
MATH	270	Hot Topics Course in Mathematics
MATH	271	Topics in Foundations
MATH	273	Topics in Numerical Analysis
MATH	274	Topics in Algebra
MATH	275	Topics in Applied Mathematics
MATH	276	Topics in Topology
MATH	277	Topics in Differential Geometry
MATH	278	Topics in Analysis
MATH	279	Topics in Partial Differential Equations
MATH	C290C	Topics in Fluid Mechanics
ECON	105	History of Economic Thought
ECON	124	Special Topics in Industrial Organization
ECON	163	Special Topics in Economic Systems
ECON	220C	Special Topics in Industrial Organization
ECON	234C	Financial Decision-Making in Firms
ECON	243	Special Topics in Econometrics
ECON	250A	Labor Economics
ECON	250B	Labor Economics
ECON	250C	Labor Economics
ECON	270D	Special Topics in Development
STAT	157	Seminar on Topics in Probability and Statistics
STAT	232	Experimental Design
COMPSCI	265	Compiler Optimization and Code Generation
COMPSCI	297	Field Studies in Computer Science
INFO	211	Group and Organizational Approaches to Information Systems Use
INFO	242	XML Foundations
INFO	C258	Analysis and Design of Databases
INFO	298	Directed Group Study

Cluster 2's courses(excerpt)

MATH	H104	Introduction to Analysis
MATH	H113	Introduction to Abstract Algebra
MATH	170	Mathematical Methods for Optimization
MATH	236	Metamathematics of Set Theory
ECON	101A	Economic Theory--Micro
ECON	101A	Economic Theory--Micro
ECON	126	Industrial Organization: Theory and Evidence
ECON	136	Financial Economics
ECON	136	Financial Economics
ECON	140	Economic Statistics and Econonmetrics
ECON	141	Econometric Analysis
ECON	201A	Economic Theory
ECON	201B	Economic Theory
ECON	202A	Macroeconomic Theory
ECON	202B	Macroeconomic Theory
ECON	209A	Theory and Application of Non-Cooperative Games
ECON	234A	Macroeconomic Finance
ECON	236C	Capital and Economic Growth
ECON	240A	Econometrics
ECON	240B	Econometrics
STAT	153	Introduction to Time Series
STAT	155	Game Theory
STAT	230A	Linear Models
STAT	248	Analysis of Time Series
INFO	240	Principles of Information Retrieval

Cluster 3's courses(excerpt)

MATH	1B	Calculus
MATH	32	Precalculus
MATH	55	Discrete Mathematics
MATH	115	Introduction to Number Theory
MATH	116	Cryptography
MATH	118	Fourier Analysis, Wavelets, and Signal Processing
MATH	123	Ordinary Differential Equations
MATH	125A	Mathematical Logic
MATH	127	Mathematical and Computational Methods in Molecular Biology
MATH	130	The Classical Geometries
MATH	135	Introduction to the Theory of Sets
MATH	136	Incompleteness and Undecidability
MATH	151	Mathematics of the Secondary School Curriculum I
MATH	203	Asymptotic Analysis in Applied Mathematics
MATH	204	Ordinary Differential Equations
MATH	215A	Algebraic Topology
MATH	215B	Algebraic Topology
MATH	224A	Mathematical Methods for the Physical Sciences
MATH	224B	Mathematical Methods for the Physical Sciences
MATH	225A	Metamathematics
MATH	225B	Metamathematics
MATH	227A	Theory of Recursive Functions
MATH	229	Theory of Models
MATH	235A	Theory of Sets
MATH	239	Discrete Mathematics for the Life Sciences
MATH	250B	Multilinear Algebra and Further Topics
MATH	254A	Number Theory
MATH	254B	Number Theory
MATH	256A	Algebraic Geometry
MATH	256B	Algebraic Geometry
MATH	290	Seminars
ECON	181	International Trade
ECON	260A	Comparative Economics
ECON	260B	Comparative Economics
COMPSCI	271	Randomness and Computation

Cluster 4's courses(excerpt)

ECON	C3	Introduction to Environmental Economics and Policy
ECON	100B	Economic Analysis--Macro
ECON	100B	Economic Analysis--Macro
ECON	101B	Economic Theory--Macro
ECON	101B	Economic Theory--Macro
ECON	113	American Economic History
ECON	121	Industrial Organization and Public Policy
ECON	123	Government Regulation of Industry
ECON	C125	Environmental Economics
ECON	131	Public Economics
ECON	138	Financial and Behavioral Economics
ECON	157	Health Economics
ECON	161	Economics of Transition: Eastern Europe
ECON	C171	Economic Development
ECON	206	Mechanism Design and Agency Theory
ECON	207A	Mathematical Economics
ECON	220A	Industrial Organization
ECON	230A	Public Economics
ECON	230C	Public Sector Microeconomics
ECON	236A	Aggregate Economics
ECON	236B	Aggregate Economics
ECON	280A	International Economics
ECON	280B	International Economics
STAT	21	Introductory Probability and Statistics for Business
STAT	243	Introduction to Statistical Computing
COMPSCI	9E	Productive Use of the UNIX Environment
COMPSCI	250	VLSI Systems Design
COMPSCI	262A	Advanced Topics in Computer Systems
COMPSCI	273	Foundations of Parallel Computation
COMPSCI	284	Computer-Aided Geometric Design and Modeling
COMPSCI	287	Advanced Robotics
COMPSCI	288	Artificial Intelligence Approach to Natural Language Processing
INFO	142AC	Access to American Cultural Heritages
INFO	207	Analysis of Information Systems
INFO	219	Privacy, Security, and Cryptography
INFO	220	Management of Information Systems and Services
INFO	221	Information Policy
INFO	230	Economic Methods for Decision Making
INFO	231	Economics of Information
INFO	235	Cyberlaw
INFO	247	Information Visualization and Presentation
INFO	250	Computer-Based Communications Systems and Networks
INFO	271A	Quantitative Research Methods for Information Systems and Management
INFO	272	Qualitative Research Methods for Information Systems and Management
INFO	297	Field Study in Information

Cluster 5's courses(excerpt)

MATH	113	Introduction to Abstract Algebra
MATH	114	Second Course in Abstract Algebra
MATH	142	Elementary Algebraic Topology
MATH	208	C*-algebras
MATH	209	Von Neumann Algebras
MATH	214	Differentiable Manifolds
MATH	219	Dynamical Systems
MATH	240	Riemannian Geometry
MATH	241	Complex Manifolds
MATH	242	Symplectic Geometry
MATH	245A	General Theory of Algebraic Structures
MATH	249	Algebraic Combinatorics
MATH	250A	Groups, Rings, and Fields
MATH	252	Representation Theory
MATH	255	Algebraic Curves
MATH	257	Group Theory
MATH	261A	Lie Groups
MATH	261B	Lie Groups
MATH	265	Differential Topology
STAT	210A	Theoretical Statistics
STAT	210B	Theoretical Statistics
STAT	244	Statistical Computing
COMPSCI	172	Computability and Complexity
COMPSCI	184	Foundations of Computer Graphics
COMPSCI	276	Cryptography

Cluster 6's courses(excerpt)

MATH	16B	Analytic Geometry and Calculus
MATH	53	Multivariable Calculus
MATH	H53	Honors Multivariable Calculus
MATH	53M	Multivariable Calculus with Computers
MATH	54	Linear Algebra and Differential Equations
MATH	H54	Honors Linear Algebra and Differential Equations
MATH	54M	Linear Algebra and Differential Equations with Computers
MATH	104	Introduction to Analysis
MATH	105	Second Course in Analysis
MATH	110	Linear Algebra
MATH	121A	Mathematical Tools for the Physical Sciences
MATH	121B	Mathematical Tools for the Physical Sciences
MATH	126	Introduction to Partial Differential Equations
MATH	128A	Numerical Analysis
MATH	128B	Numerical Analysis
MATH	140	Metric Differential Geometry
MATH	141	Elementary Differential Topology
MATH	143	Elementary Algebraic Geometry
MATH	152	Mathematics of the Secondary School Curriculum II
MATH	153	Mathematics of the Secondary School Curriculum III
MATH	160	History of Mathematics
MATH	172	Combinatorics
MATH	185	Introduction to Complex Analysis
MATH	187	Senior Level Analysis
MATH	202A	Introduction to Topology and Analysis
MATH	202B	Introduction to Topology and Analysis
MATH	205	Theory of Functions of a Complex Variable
MATH	206	Banach Algebras and Spectral Theory
MATH	212	Several Complex Variables
MATH	220	Introduction to Probabilistic Methods in Mathematics and the Sciences
MATH	221	Advanced Matrix Computations
MATH	222A	Partial Differential Equations
MATH	222B	Partial Differential Equations
MATH	228A	Numerical Solution of Differential Equations
MATH	228B	Numerical Solution of Differential Equations
MATH	258	Classical Harmonic Analysis
COMPSCI	278	Machine-Based Complexity Theory

Cluster 7's courses(excerpt)

COMPSCI	3	Introduction to Symbolic Programming
COMPSCI	3L	Introduction to Symbolic Programming
COMPSCI	3S	Introduction to Symbolic Programming (Self-Paced)
COMPSCI	9A	Matlab for Programmers
COMPSCI	9B	Pascal for Programmers
COMPSCI	9C	C for Programmers
COMPSCI	9D	Scheme and Functional Programming for Programmers
COMPSCI	9F	C++ for Programmers
COMPSCI	9G	JAVA for Programmers
COMPSCI	9H	Python for Programmers
COMPSCI	47A	Completion of Work in Computer Science 61A
COMPSCI	47B	Completion of Work in Computer Science 61B
COMPSCI	61A	The Structure and Interpretation of Computer Programs
COMPSCI	61C	Machine Structures
COMPSCI	150	Components and Design Techniques for Digital Systems
COMPSCI	152	Computer Architecture and Engineering
COMPSCI	161	Computer Security
COMPSCI	162	Operating Systems and System Programming
COMPSCI	164	Programming Languages and Compilers
COMPSCI	169	Software Engineering
COMPSCI	186	Introduction to Database Systems
COMPSCI	188	Introduction to Artificial Intelligence
COMPSCI	252	Graduate Computer Architecture
COMPSCI	260	User-Interfaces to Computer Systems
COMPSCI	263	Design of Programming Languages
COMPSCI	268	Computer Networks
COMPSCI	282	Algebraic Algorithms
COMPSCI	289	Knowledge Representation and Use in Computers
INFO	155	Introduction to High-Level Programming
INFO	257	Database Management

Cluster 8's courses(excerpt)

MATH	74	Transition to Upper Division Mathematics
MATH	300	Teaching Workshop
ECON	2	Introduction to Economics--Lecture Format
ECON	104	Advanced Microeconomic Theory
ECON	119	Psychology and Economics
ECON	152	Wage Theory and Policy
ECON	C181	International Trade
ECON	204	Mathematical Tools for Economics
ECON	219A	Foundations of Psychology and Economics
ECON	230B	Public Economics
ECON	280C	International Economics
STAT	212A	Topics in Theoretical Statistics
STAT	296	Resources for Statistical Computing
COMPSCI	47C	Completion of Work in Computer Science 61C
INFO	152	Mobile Application Design and Development
INFO	181	Technology and Poverty
INFO	182AC	Print, Literacy, and Power in America to 1900
INFO	205	Information Law and Policy
INFO	209	Professional Skills Workshop
INFO	216	Computer-Mediated Communication
INFO	246	Multimedia Information
INFO	256	Applied Natural Language Processing
INFO	280	Information and Communication Technologies and Development: Context, Strategies and Impacts
INFO	286	Information, Innovation and International Development (I3D): Designing Rural Computing Applications
INFO	298A	Directed Group Work on Final Project

Cluster 9's courses (excerpt)

MATH	24	Freshman Seminars
MATH	84	Sophomore Seminar
MATH	H90	Honors Undergraduate Seminar in Mathematical Problem Solving
MATH	301	Undergraduate Mathematics Instruction
MATH	303	Professional Preparation: Supervised Teaching of Mathematics
ECON	24	Freshman Seminar
ECON	98	Directed Group Study
ECON	115	The World Economy in the Twentieth Century
ECON	122	Industrial Organization Seminar
ECON	137	Aggregate Economics Seminar
ECON	153	Labor Economics Seminar
ECON	173	Economic Development Seminar
ECON	183	International Economic Seminar
ECON	190	Seminar on Topics in Economics
ECON	H195B	Senior Honors Thesis
ECON	197	Field Studies
ECON	198	Directed Group Study
ECON	209B	Theory and Application of Non-Cooperative Games: II
ECON	218	Seminar in Psychology and Economics
ECON	219C	Topics in Psychology and Economics
ECON	219D	Experimental Economics
ECON	235	Financial Economics Seminar
ECON	291	Departmental Seminar
ECON	295	Survey of Research in Economics
ECON	299	Supervised Independent Study and Research
ECON	301	GSI Practicum
STAT	260	Topics in Probability and Statistics
STAT	272	Statistical Consulting
STAT	601	Individual Study for Master's Candidates
COMPSCI	24	Freshman Seminars
COMPSCI	84	Sophomore Seminar
COMPSCI	H196A	Senior Honors Thesis Research
COMPSCI	H196B	Senior Honors Thesis Research
INFO	24	Freshman Seminar
INFO	84	Sophomore Seminar
INFO	141	Search Engines: Technology, Society, and Business
INFO	190	Special Topics in Information
INFO	199	Individual Study
INFO	290	Special Topics in Information
INFO	295	Doctoral Colloquium
INFO	296A	Seminar
INFO	296B	Seminar
INFO	299	Individual Study

Cluster 10's courses(excerpt)

MATH	49	Supplementary Work in Lower Division Mathematics
MATH	98	Supervised Group Study
MATH	196	Honors Thesis
MATH	198	Directed Group Study
MATH	253	Homological Algebra
MATH	295	Individual Research
MATH	299	Reading Course for Graduate Students
MATH	600	Individual Study for Master's Students
ECON	172	Case Studies in Economic Development
ECON	H195A	Senior Honors Thesis
ECON	196	Special Topics in Economics
ECON	199	Supervised Independent Study and Research
ECON	220B	Industrial Organization
ECON	241B	Econometrics
ECON	251	Seminar in Labor Economics
ECON	270A	Development Economics
ECON	270B	Development Economics
ECON	275B	Selected Topics in Economic Demography
ECON	296	Special Topics in Economics
ECON	298	Directed Group Study for Graduates
STAT	39	Freshman/Sophomore Seminar
STAT	98	Directed Group Study
STAT	198	Directed Study for Undergraduates
STAT	278B	Statistics Research Seminar
STAT	298	Directed Study for Graduate Students
STAT	300	Professional Preparation: Teaching of Probability and Statistics
COMPSCI	39	Freshman/Sophomore Seminar
COMPSCI	61BL	Data Structures and Programming Methodology
COMPSCI	61CL	Machine Structures
COMPSCI	98	Directed Group Study
COMPSCI	99	Individual Study and Research for Undergraduates
COMPSCI	194	Special Topics
COMPSCI	198	Directed Group Studies for Advanced Undergraduates
COMPSCI	199	Supervised Independent Study
COMPSCI	294	Special Topics
COMPSCI	298	Group Studies Seminars, or Group Research
COMPSCI	299	Individual Research
COMPSCI	300	Teaching Practice
COMPSCI	301	Teaching Techniques for Computer Science
COMPSCI	399	Professional Preparation: Supervised Teaching of Computer Science
INFO	39	Freshman/Sophomore Seminar

Cluster 11's courses(excerpt)

ECON	C102	Natural Resource Economics
ECON	222	Economics of Innovation
ECON	C222	Economics of Innovation
STAT	C239A	The Statistics of Causal Inference in the Social Science
STAT	C241A	Statistical Learning Theory
STAT	C241B	Advanced Topics in Learning and Decision Making
STAT	C261	Quantitative/Statistical Research Methods in Social Sciences
COMPSCI	C149	Introduction to Embedded Systems
COMPSCI	C182	The Neural Basis of Thought and Language
COMPSCI	C195	Social Implications of Computer Technology
COMPSCI	261	Security in Computer Systems
COMPSCI	C280	Computer Vision
COMPSCI	C281A	Statistical Learning Theory
COMPSCI	C281B	Advanced Topics in Learning and Decision Making
COMPSCI	302	Designing Computer Science Education
INFO	C103	History of Information
INFO	146	Foundations of New Media
INFO	203	Social and Organizational Issues of Information
INFO	210	The Information and Services Economy
INFO	212	Information in Society
INFO	214	Needs and Usability Assessment
INFO	218	Concepts of Information
INFO	237	Intellectual Property Law for the Information Industries
INFO	C262	Theory and Practice of Tangible User Interfaces
INFO	C283	Information and Communications Technology for Development

Cluster 12's courses(excerpt)

ECON	241A	Econometrics
STAT	133	Concepts in Computing with Data
STAT	C141	Statistics for Bioinformatics
STAT	C143	Introduction to Statistical Methods in Computational and Genomic Biology
STAT	151A	Linear Modelling: Theory and Applications
STAT	151B	Linear Modelling: Theory and Applications
STAT	215A	Statistical Models: Theory and Application
STAT	215B	Statistical Models: Theory and Application
STAT	C245A	Biostatistical Methods: Advanced Categorical Data Analysis
STAT	C245B	Biostatistical Methods: Survival Analysis and Causality
STAT	C245C	Biostatistical Methods: Computational Statistics with Applications in Biology and Medicine
STAT	C247C	Longitudinal Data Analysis
STAT	C249C	Multiple Testing and Loss Function Based Estimation: Applications in Biological Sciences
COMPSCI	61B	Data Structures
COMPSCI	170	Efficient Algorithms and Intractable Problems
COMPSCI	258	Parallel Processors
COMPSCI	262B	Advanced Topics in Computer Systems
COMPSCI	266	Introduction to System Performance Analysis
COMPSCI	C267	Applications of Parallel Computers
COMPSCI	270	Combinatorial Algorithms and Data Structures
COMPSCI	274	Computational Geometry
COMPSCI	285	Solid Free-Form Modeling and Fabrication
COMPSCI	286	Implementation of Data Base Systems
INFO	202	Information Organization and Retrieval
INFO	206	Distributed Computing Applications and Infrastructure
INFO	271B	Quantitative Research Methods for Information Systems and Management

Cluster 13's courses(excerpt)

MATH	1A	Calculus
MATH	16A	Analytic Geometry and Calculus
MATH	C103	Introduction to Mathematical Economics
MATH	H110	Linear Algebra
MATH	H185	Introduction to Complex Analysis
ECON	C103	Introduction to Mathematical Economics
ECON	C110	Game Theory in the Social Sciences
ECON	C142	Applied Econometrics and Public Policy
ECON	154	Economics of Discrimination
ECON	C175	Economic Demography
ECON	219B	Applications of Psychology and Economics
ECON	C225	Workshop in Institutional Analysis
ECON	244	Applied Econometrics
ECON	C270A	Microeconomics of Development
ECON	C275B	Aging: Economic and Demographic Aspects
STAT	135	Concepts of Statistics
STAT	C245E	Statistical Genomics
STAT	C245F	Statistical Genomics
STAT	C249A	Censored Longitudinal Data and Causality
COMPSCI	C191	Quantum Information Science and Technology
INFO	243	Document Engineering and Information Architecture

Cluster 14's courses(excerpt)

MATH	C218A	Probability Theory
MATH	C218B	Probability Theory
MATH	C223A	Stochastic Processes
MATH	C223B	Stochastic Processes
STAT	2	Introduction to Statistics
STAT	20	Introduction to Probability and Statistics
STAT	131A	Statistical Inferences for Social and Life Scientists
STAT	134	Concepts of Probability
STAT	150	Stochastic Processes
STAT	152	Sampling Surveys
STAT	200A	Introduction to Probability and Statistics at an Advanced Level
STAT	200B	Introduction to Probability and Statistics at an Advanced Level
STAT	204	Probability for Applications
STAT	C205A	Probability Theory
STAT	205B	Probability Theory
STAT	C205B	Probability Theory
STAT	206A	Stochastic Processes
STAT	C206A	Stochastic Processes
STAT	206B	Stochastic Processes
STAT	C206B	Stochastic Processes
STAT	251	Stochastic Analysis with Applications to Mathematical Finance
COMPSCI	70	Discrete Mathematics and Probability Theory
COMPSCI	174	Combinatorics and Discrete Probability