

**2015.01**

# **Bigdata** **Machine Learning Platform**

**BICube™**

<http://www.bicube.co.kr>

## BICube™ Architecture

## BICube

Manipul Mashup Preprocess Classify Cluster Associate Bioinfo Visual

## Machine Learning Engine

Mesos

Zookeeper

Yarn  
(2.0)

Spark  
SQL

Spark  
Stream

Mahout  
ML

GraphX

Hive  
Pig

MapReduce

Spark

## HDFS (Hadoop Distributed File System)

Sqoop

Flume

Storm

Kafka

Akka

Authentication

CubeManager

RDB

RDB Mashup

SNS Crawler

Log Crawler



## BI Cube™ Algorithms

MultiLinearRegression, KoreanMorpheme  
Logistic Regression, SVM, Adaboost, Neural Network  
RandomForest, NaiveBayssian, Timeseries Analysis  
Kmeans, LDA, HMM, PCA, SVD, CRF, TDA  
Covariance, Correlation, ROC, LSA, EigenFace Image  
CF Recommender, ConfusionMatrix, AUC  
Bagging, Wagging, ArcX4, LogitBoost, ModestAdaboost  
SAMME, BinaryClibration, IsonicCalibration...  
Discriminant Adaptive Nearist Neighbor  
Hierachical agglomerative clustering  
Kernel Recursive Least Square, Radial Base Network  
Local scaled density base clustering, EMGaussianMixture  
Sequential minimum optimization, Dual coordinate Descent  
T-Test, F-Test, G-Test, Chi-square-Test Anova etc....

## Latent Dirichlet Allocation on Hadoop

## Document-Topic Matrix Viewer

Cluster		MatrixViewer													
Editable		Header Size : 6 Font Size : 6													
Docs	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15
doc0	칠거민	집안형편	호연	독실	반벌터리	농업학교	어린이날	자명종	신사참배	영탁검	미계지	인분	전업	아동교육	상업
doc1	설	하길	아토	오인해	진한	따름	즐거워하	체류기간	스커트	군대군대	논개	논바닥	되뇌이	가래	시건방지
doc2	쫓수	허기증	방새	실무	파차	손기정	건너뛰	다음번	무의촌	심장마비	큰살	한자말	새신랑	법석	수험생
doc3	고개	끼	피	치	남	사라	시절	위로	세상	나가	소년	가면	사람은	통해	밖
doc4	출달음 치	불력	하변	회중	증	예상대로	킥터	버금가	깎곤	무너뜨리	변자인	이달	홀러딩	형하	다세대
doc5	기백만	고용주	본원	독재정치	일률	구토	영원불변	결단성	출혈	전취	칠투칠미	인내심	인류문화	급습급습하	부적합
doc6	정하	밀어붙이	베드민턴	묘사력	본전	고쳐지	배부로	셀	과외공부	탕	안성	노소	타구	람이	사관생도
doc7	동자	복여	사퇴	기회주의	칠권	배시간	재고	토칸	생산성	국민주의	완결성	포화	4프롤레타리아	몰레바워	코헨
doc8	곡식	농부	오래가	판서	외고	변변이	대면	마땅히	인지	대틀	썩름	답안	병석	국무	학내
doc9	워지	사생아	소용돌이	지름	정설	모순구조	유즈	사회규정	규칙	물자	무정부상태	대자연	브란덴부르크	대변혁	호평
doc10	어때요	넉넉히	상관없이	외동딸	아치	라틴	세일본	토목점	에마틴	걸히	후듯	두더지	숫	칠하	화투놀이
doc11	고향길	머리채	들까불	버들	끈끈하	정해년	시답	중낮	어지럼증	다리섬	가벼워지	외침소리	가시내	분연히	웅송그리
doc12	으며	송의	들격대	설욕	덕장	쓰라림	기초교육	원구	차들	이강수	이적행위	국난	머리통	각송	진입
doc13	격동색	유성	가수	인사드리	후탁	청자	귀인	여의주	대거리	결연히	팔파집	성덕	사시사철	화살대	헬죽
doc14	합의제	물건	반다	숨어들	주파	허사	사급	웅성	독선	방심	주라	풍경화	아지	군력	살머시
doc15	중기	전광석화	지방대	구세	마장	수항	가엾이	승리자	사시나무	안전지대	유방	쇄도	수초	문름	너비
doc16	수목화	소세지	두레	광주진압	랭기	독일수상	보급자리	통행금지	이기주의	육식	민주기	세원	사회학자	연화	우하
doc17	동경대	박진감	로씨	기서	다스려지	유치홍	마르세이 유	여하	더블린	서구열강	생성과정	마사	홍문관	부속학교	가오루
doc18	집산	식음	음악	박장현	격자생존	담판	시종관	김경하	한울	외래	전라남도	이용태	확실시	항일독립운동	이태인
doc19	반골	고린도전 서	오인	할랄	꽃피우	심벌	독산물	박정수	눈시울	서정각	철석	인스	이종우	정신박악아	고단

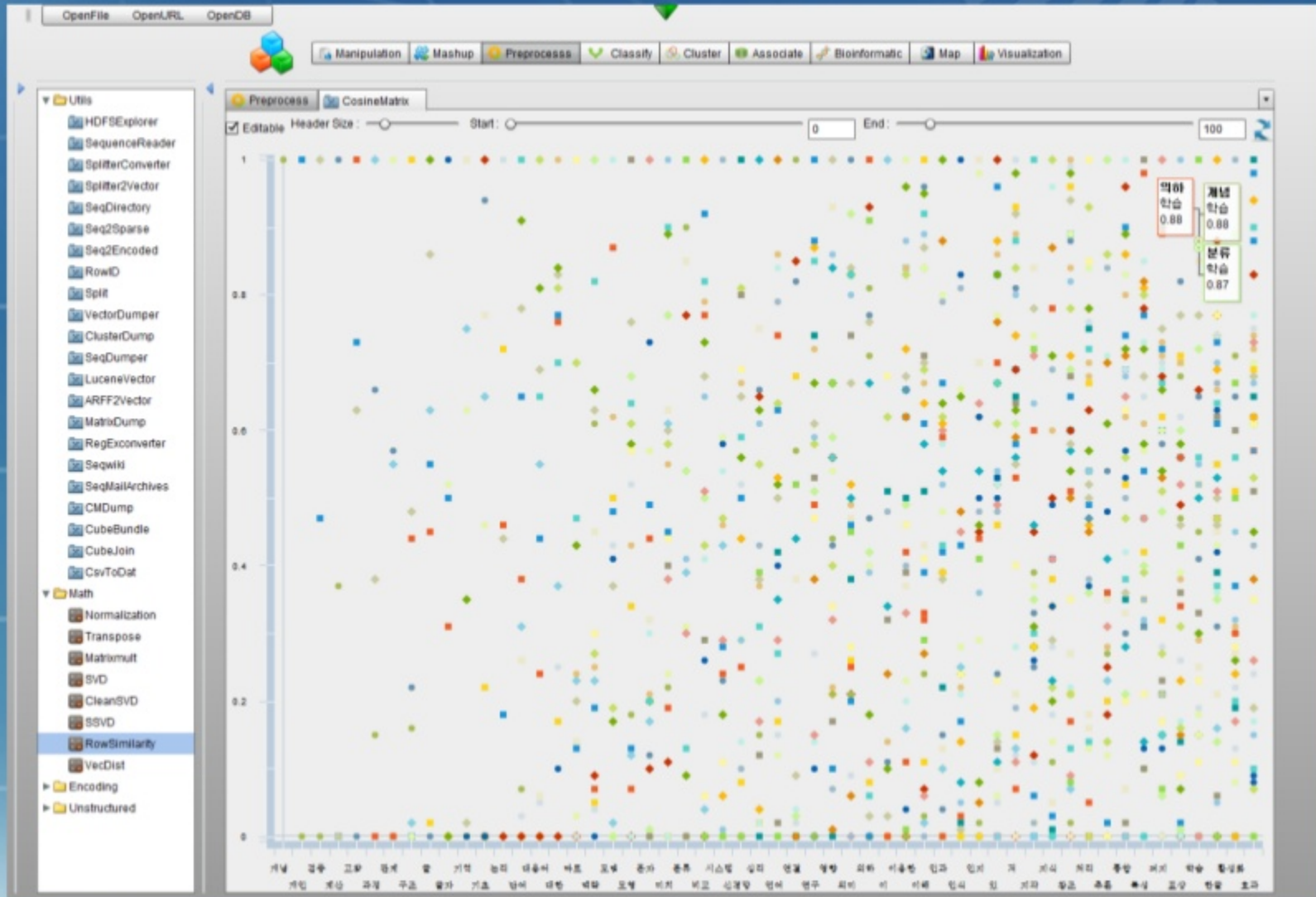


## Latent Dirichlet Allocation on Hadoop

## Word-Topic Matrix Viewer

Cluster	MatrixViewer	MatrixViewer																		
Editable Header Size : 7 Font Size : 7																				
Topics	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13	W14	W15	W16	W17	W18	W19	W20
topic1	시호	보습	인간학	비합리성	개고	선심	불일치	인식능력	진보주의	방매	유명해지	산악	박동	법신문	자연필서	발업	위협성	보죽	정산	류즈
topic2	제한군인	뒷자리	두병	송담	대통령학	부리내리	황보	육군대신	우우측손	정양	원사	언론인	박화성	출려농	관복	군국주의	한송이	경신	자이탈구	구인회
topic3	이와	강제	거역	거력	일그러지	싸구려	노론자위	철철	나론	구름장	아닌제야	밀려가	군말	풀이가	결지	만용	한탄	전성기	호각적세	낼새나
topic4	군말	시계	급소	제해	호제	독전	무기력	선수권	알설	중립대륙	무순	무서리	역전승	아쉬워하	통경	거수	협박	절경	굴뚝거리	로상
topic5	기찰	글지	원	경종	오복	내력	차마물	입초	추장	깜짝거리	복변	밀실	견곤	창의문	제손	형심환	비도덕	앗	으아	다오
topic6	제현	한년	국외자	만고	파제예프	속담	봉인	죽맛후	블라디보	헛간	황년	미즈노	제라농	고속도로	막힘없	토자	증식	폐독	숙적	교조주
topic7	도마위	보통교육	조직망	다방면	조세	수혜자	내과의사	변역관	연간	합작	포용성	양감	등급	개혁안	난폭	처방전	메일	자금문제	서구문명	학위취
topic8	남국	화강암	1시골	질레	핏방울	라춤	풍뎡이	나라시	수직	고요히	월계수	푸른빛	대통령	비이	내용	여우	무지개빛	나비	오보	돈
topic9	1시골	월계수	드니	화강암	겨우살이	수직	나뭇등결	덜히	애도	푸른빛	백옥	나라시	라춤	질레	풍뎡이	양탄자	꼭포수	여우	원예	남국
topic10	화강암	비이	나뭇등결	1시골	대통령	빗줄기	수직	양탄자	푸른빛	드니	무지개빛	오보	월계수	꼭포수	푸드덕거	애도	남국	밀림	질레	덜히
topic11	업연히	김동리	미선	송화	동택	기노	방인근	9	구인회	독백	1관복	정음사	손쉽게	정정의	손문	시민생활	복사꽃	산불	지면관계	눈가루
topic12	남국	풍뎡이	나비	경비원	불식	이래저래	문명세계	전성	괴값	양탄자	어업	불신감	합작	유용하	중국학	다되	관변	사업주	심대	건의서
topic13	나뭇등결	1시골	파해치	척기	허상	공문서	김동리	보리스	영인	주범	1관복	국가저상	잔병	잡지사	급전	7만주국	선미	별견	이용진	기자생
topic14	여유만만	유악	정치생활	구속력	육사병	자기도취	복해	죽군	가두시위	방청	격론	사기관	에테트	미숙현	선전자	밀어닥치	창시자	프링크	의석	실효
topic15	전종	대중조직	고등교육	파본	기술면	남아프리카	태지우리	투자대상	영유	역동	강타	중국학	오콜라호	개인교습	이용가치	심사숙고	놀라워하	정치지도	매설	소관사
topic16	1시골	석천	황면	아전인수	동적	조복	은모	습관화	주사기	카르트	목청것	자가용	대중시	배자기	나귀	사춘기	작아지	김광섭	양모	완
topic17	1시골	백옥	비이	양탄자	나뭇등결	남국	벌지갑치	황금빛	청초	두문불출	복심	무심결에	관자	나신	궁실	돗자리	원	활형	별당	일투족
topic18	질레	드니	초록빛	수직	라춤	나라시	덜히	풍뎡이	고등교육	교육과정	합작	핏발	나카야마	중부	조가비	날라다루	개전	오버	택시	출판계
topic19	두병	급전	이의준	문화단체	의학전문	일편	방인근	규선	조한	경쟁심	차주	보완	자력경쟁	합박	보통일	전향	교원생활	대국출판	부도덕성	민족은
topic20	화강암	핏방울	겨우살이	무지개빛	고요히	비이	월계수	빗줄기	풍뎡이	라춤	나라시	푸드덕거	1시골	남국	질레	가위	케이	수직	백옥	코피

## Document Similarity Visualization





## Document Similarity Matrix

Manipulation Mashup Preprocess Classify Cluster Associate Bioinformatic Map Visualization

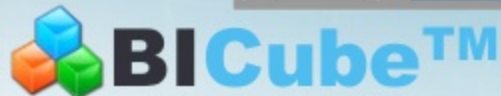
Preprocess CosineMatrix CosineMatrix

☒ Editable Header Size: 60 Start: 0 End: 100

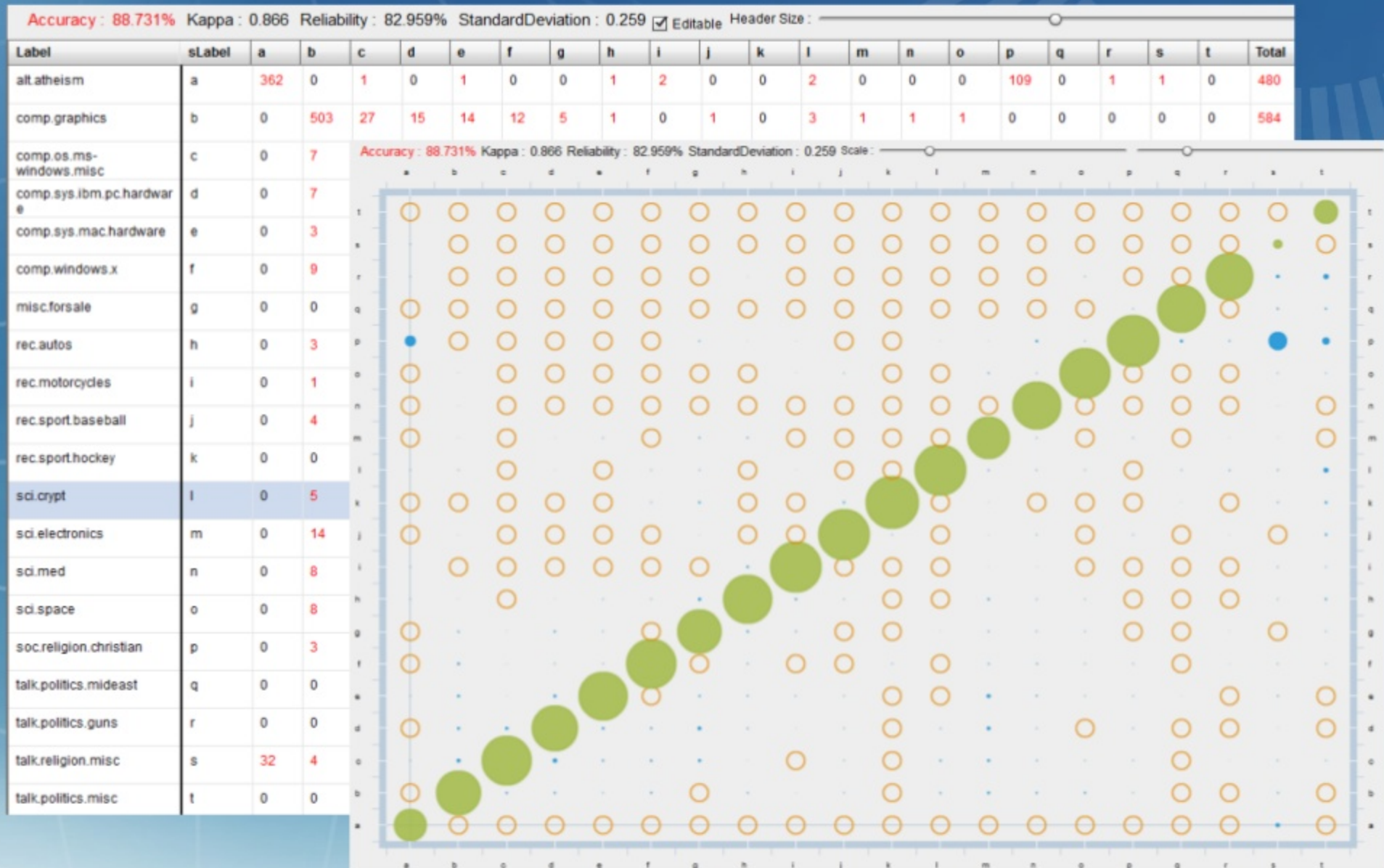
Docs	doc/c01	doc/c02	doc/c03	doc/c04	doc/c05	doc/c06	doc/c07	doc/c08	doc/c09	doc/c10	doc/c11	doc/c12	doc/c13	doc/c14	doc/c15	doc/c16	doc/c17	do...
doc/c01	1.0																	
doc/c02	0.9	1.0																
doc/c03	0.77	0.89	1.0															
doc/c04	0.0	0.0	0.0	1.0														
doc/c05	0.9	0.96	0.87	0.0	1.0													
doc/c06	0.57	0.81	0.53	0.0	0.74	1.0												
doc/c07	0.79	0.89	0.64	0.0	0.75	0.84	1.0											
doc/c08	0.38	0.0	0.0	0.05	0.0	0.0	0.08	1.0										
doc/c09	0.17	0.32	0.22	0.45	0.12	0.36	0.6	0.0	1.0									
doc/c10	0.33	0.6	0.41	0.0	0.66	0.86	0.47	0.0	0.0	1.0								
doc/c11	0.91	0.81	0.81	0.0	0.91	0.42	0.55	0.23	0.01	0.37	1.0							
doc/c12	0.14	0.0	0.0	0.0	0.0	0.0	0.15	0.63	0.0	0.0	0.0	1.0						
doc/c13	0.12	0.0	0.19	0.14	0.0	0.0	0.0	0.7	0.0	0.0	0.12	0.26	1.0					
doc/c14	0.32	0.41	0.57	0.0	0.61	0.24	0.0	0.0	0.0	0.59	0.65	0.0	0.0	1.0				
doc/c15	0.0	0.0	0.0	0.46	0.0	0.05	0.04	0.06	0.29	0.03	0.0	0.0	0.0	0.0	1.0			
doc/c16	0.02	0.0	0.0	0.32	0.01	0.15	0.06	0.0	0.19	0.19	0.05	0.0	0.0	0.0	0.98	1.0		
doc/c17	0.0	0.0	0.0	0.5	0.0	0.08	0.04	0.0	0.36	0.04	0.0	0.0	0.0	0.0	0.99	0.95	1.0	
doc/c18	0.03	0.0	0.0	0.15	0.05	0.16	0.0	0.0	0.0	0.27	0.09	0.0	0.0	0.09	0.92	0.98	0.88	1.0
doc/c19	0.53	0.63	0.28	0.0	0.45	0.79	0.91	0.05	0.68	0.38	0.2	0.29	0.0	0.0	0.26	0.25	0.31	0.17
doc/c20	0.62	0.73	0.46	0.0	0.77	0.85	0.61	0.0	0.0	0.86	0.53	0.15	0.0	0.42	0.0	0.11	0.0	0.22
doc/c21	0.27	0.0	0.09	0.0	0.0	0.0	0.03	0.9	0.0	0.0	0.1	0.63	0.89	0.0	0.0	0.0	0.0	0.0
doc/c22	0.4	0.3	0.55	0.31	0.36	0.0	0.1	0.23	0.29	0.0	0.62	0.0	0.45	0.46	0.03	0.01	0.0	0.0







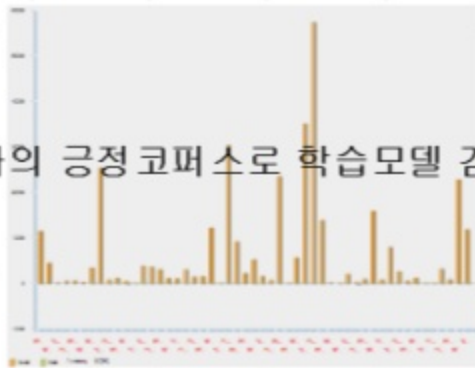
## Confusion Matrix for Accuracy Measure





# Latent Semantic Visualization

영화의 긍정 코퍼스로 학습모델 검증 결과 100%



호텔의 긍정 코퍼스로 학습모델 검증 결과 100%



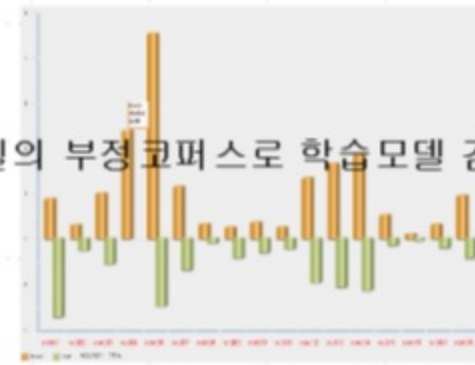
영화의 긍정 코퍼스로 학습모델 검증 결과 97%



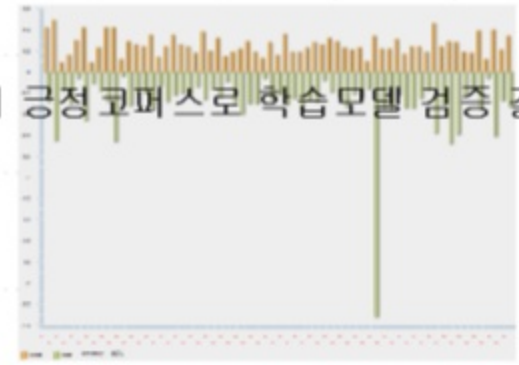
호텔의 긍정 코퍼스로 학습모델 검증 결과 100%



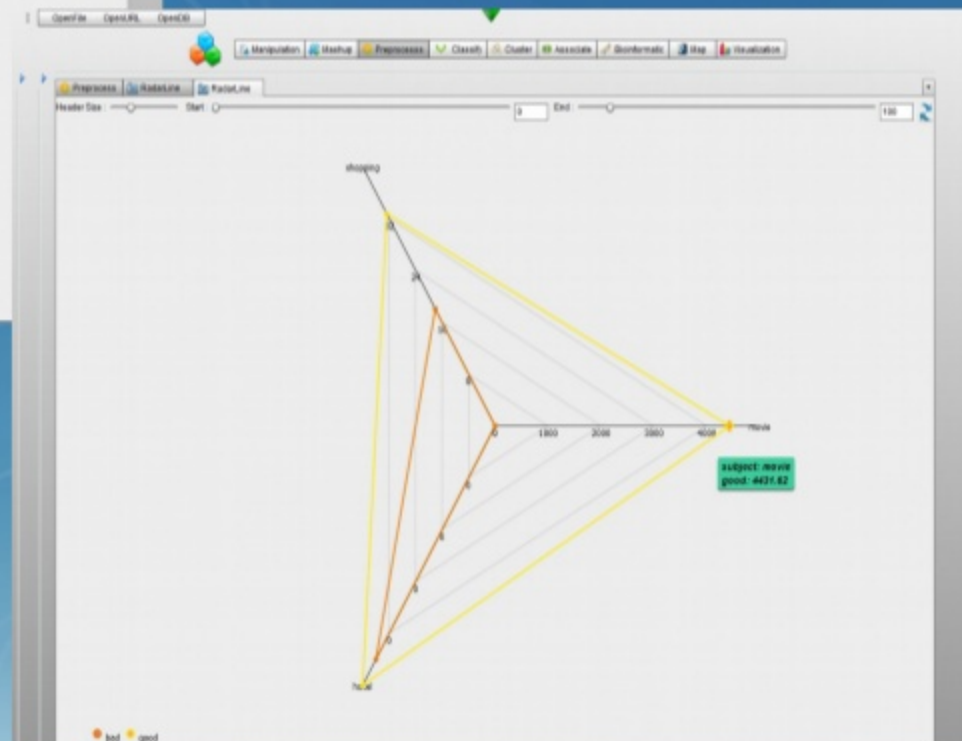
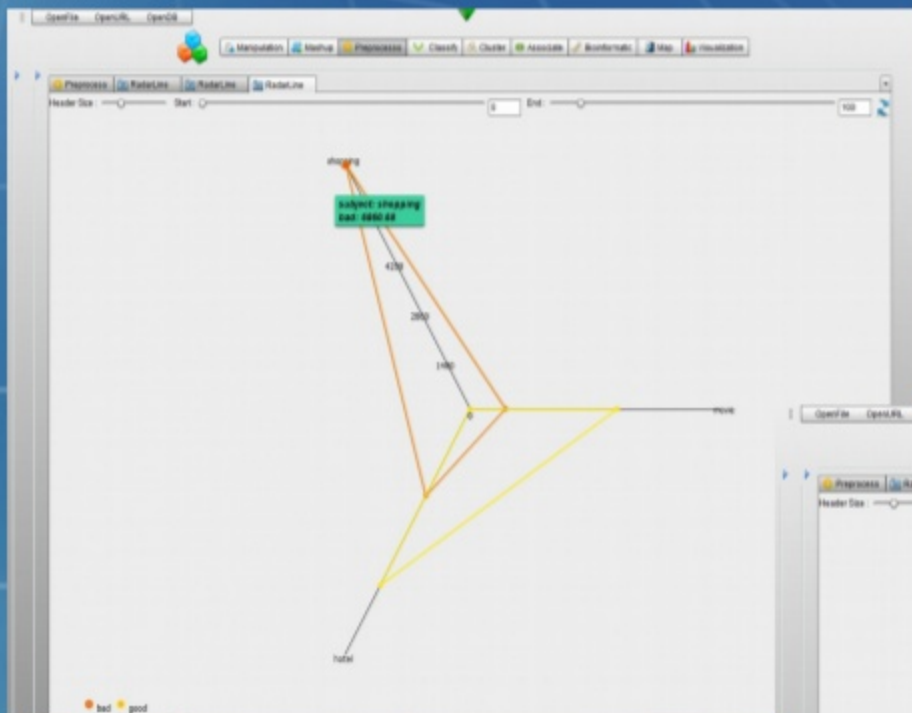
호텔의 부정 코퍼스로 학습모델 검증 결과 11%



영화의 긍정 코퍼스로 학습모델 검증 결과 49%

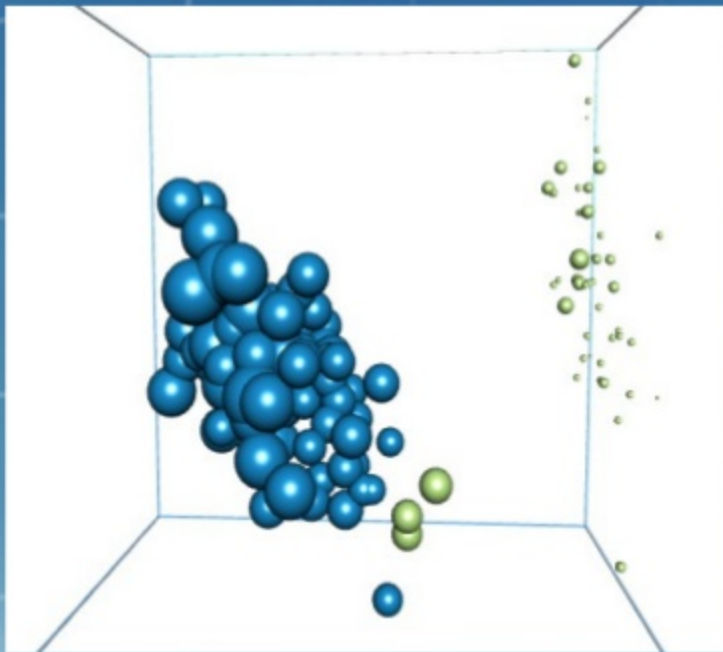
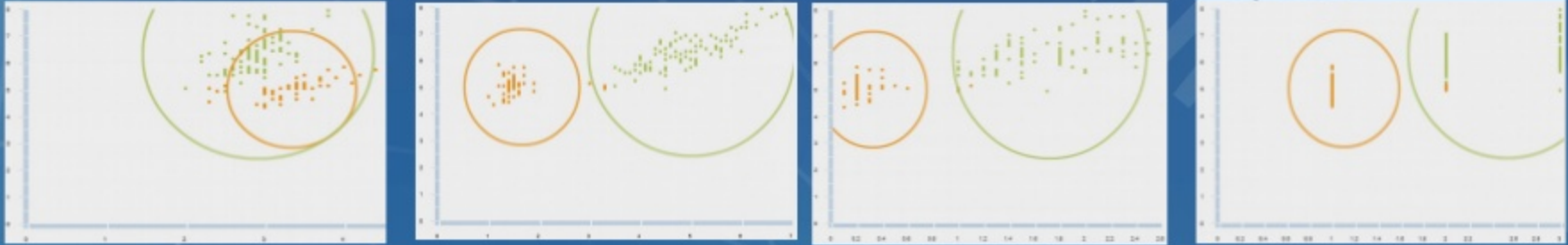


## Latent Semantic Visualization

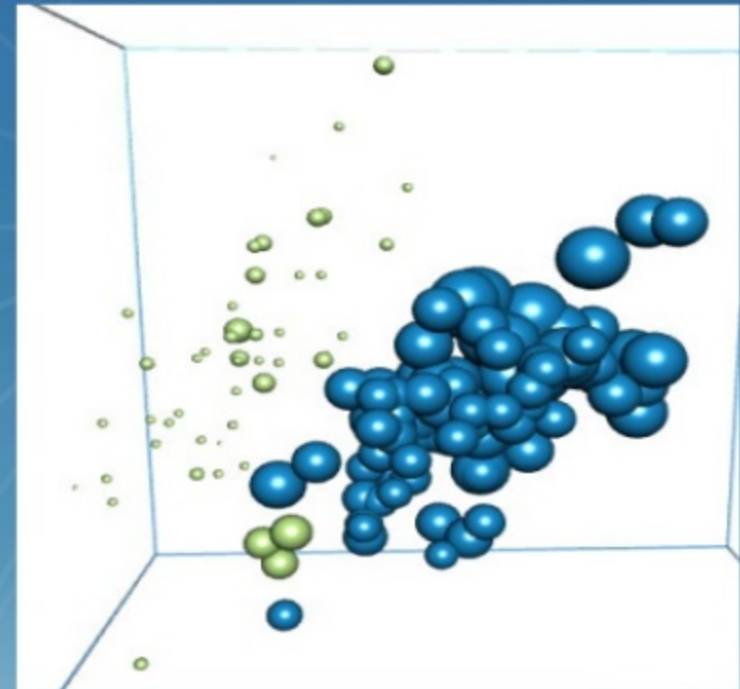




# Unstructured Document Clustering



3D Visualization on GPU



# Association Rule from MBA

Manipulation Mashup Preprocess Classify Cluster Associate Bioinformatic Map Visualization

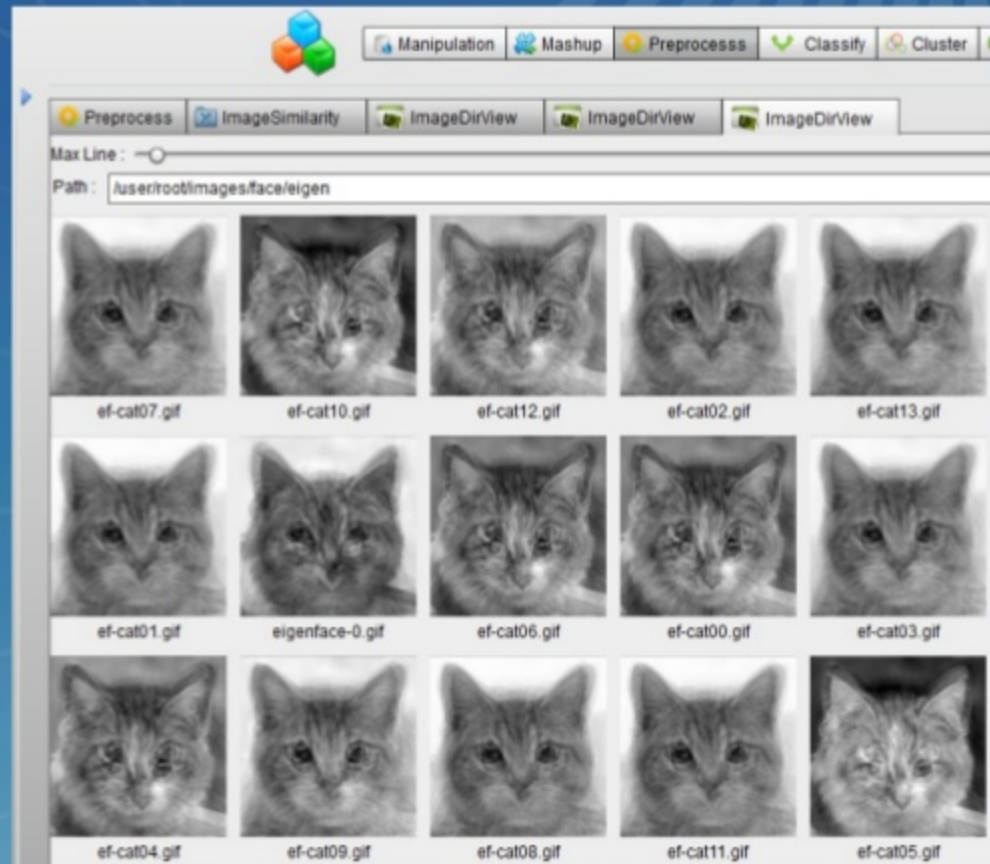
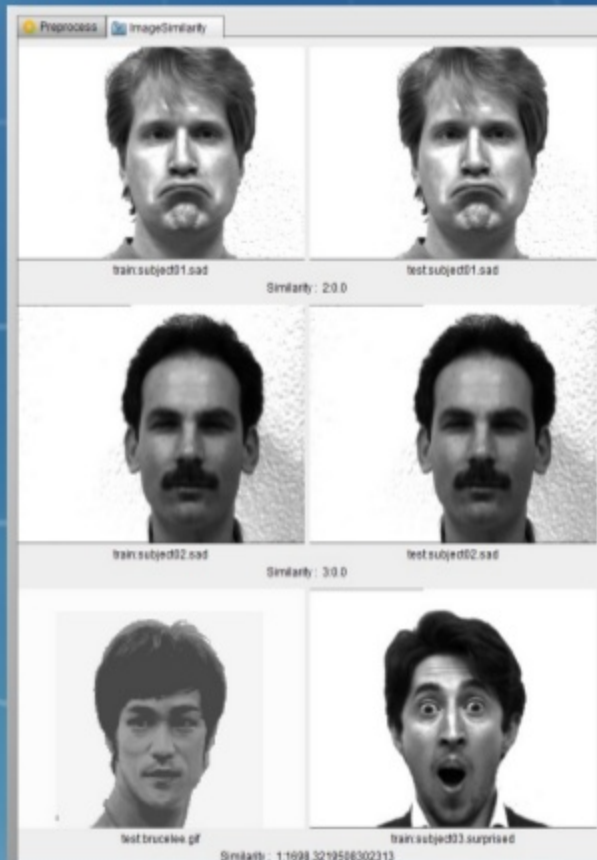
Associate ARules

FontSize: 14 Generation Transaction: 43 Editable

sourceitem	targetitem	support	confidence	lift	conviction
Hot Dogs	Hot Dog Buns	0.04188096987509184	0.7125	0.005654761904761905	3.1562470050793854
98pct. Fat Free Hamt	Hamburger Buns	0.048493754592211606	0.6804123711340206	0.00535757772546473	2.837050555805741
White Bread	Aspirin	0.04188096987509184	0.6263736263736264	0.0038665038665038664	2.3578899598046417
White Bread	Tomatoes	0.04041146216017634	0.6111111111111112	0.0037722908093278463	2.265351107379028
White Bread	Toothpaste	0.04775900073475386	0.6018518518518519	0.0037151348879743942	2.2126685234864927
Eggs	Toothpaste	0.04481998530492285	0.5648148148148148	0.0033821246396096697	2.0159144558913185
White Bread	Wheat Bread	0.04335047759000735	0.5619047619047619	0.0034685479129923573	2.010909497492253
2pct. Milk	Wheat Bread	0.0426157237325496	0.5523809523809524	0.0037072547139661233	1.98946331702284
Hot Dogs	Sweet Relish	0.0470242468772961	0.5517241379310345	0.004378762999452655	2.0242468772961058
2pct. Milk	Toothpaste	0.04335047759000735	0.5462962962962963	0.0036664180959482978	1.9627824678732624
White Bread	Onions	0.0426157237325496	0.5321100917431193	0.003284630195945181	1.882857184019824
White Bread	98pct. Fat Free Hambu	0.04922850844966936	0.5275590551181102	0.0032565373772722853	1.8647195689444032
White Bread	Potato Chips	0.05143277002204261	0.5263157894736842	0.003248862897985705	1.85982529186056
Eggs	Sweet Relish	0.04481998530492285	0.5258620689655172	0.003148874664464175	1.850297241333244
Eggs	Potatoes	0.045554739162380606	0.5254237288135594	0.0031462498731350858	1.8485882229453137
Eggs	Wheat Bread	0.04041146216017634	0.5238095238095238	0.003136583974907328	1.8423218221895663
Hamburger Buns	98pct. Fat Free Hambu	0.048493754592211606	0.5196850393700787	0.00535757772546473	1.933583069343901
Eggs	Cola	0.04041146216017634	0.5188679245283019	0.003106993560049712	1.823399749319272
White Bread	Cola	0.04041146216017634	0.5188679245283019	0.0032028884230142093	1.83103542666148
2pct. Milk	Cola	0.04041146216017634	0.5188679245283019	0.003482335063948335	1.8508881877512209
2pct. Milk	Onions	0.041146216017634095	0.5137614678899083	0.0034480635428852905	1.8314502377552575
Sweet Relish	Hot Dogs	0.0470242468772961	0.5079365079365079	0.004378762999452655	1.8590457680547985
Eggs	Onions	0.04041146216017634	0.5045871559633027	0.00302147997582816	1.7708384357906766
Eggs	Potato Chips	0.048493754592211606	0.49624060150375937	0.0029715006078069425	1.7414982398806846
White Bread	Potatoes	0.0426157237325496	0.4915254237288136	0.003034107553881565	1.732574087680627
Eggs	98pct. Fat Free Hambu	0.045554739162380606	0.4881889763779528	0.002923287283700316	1.714101622110439
2pct. Milk	Sweet Relish	0.041146216017634095	0.4827586206896552	0.003239990742883592	1.7216752387950038
Eggs	2pct. Milk	0.05216752387950037	0.47651006711409394	0.002853353695293976	1.675860509806138
White Bread	2pct. Milk	0.05143277002204261	0.4697986577181208	0.002899991714309388	1.6615760935276556
Potato Chips	98pct. Fat Free Hambu	0.04335047759000735	0.4645669291338583	0.0034929844295778817	1.6851363616717812
Eggs	White Bread	0.055106539309331376	0.46296296296296297	0.0027722333111554667	1.633585852187793
Eggs	Hot Dogs	0.0426157237325496	0.4603174603174603	0.002756391977948864	1.6255780784025584



## Image Analysis from Hadoop



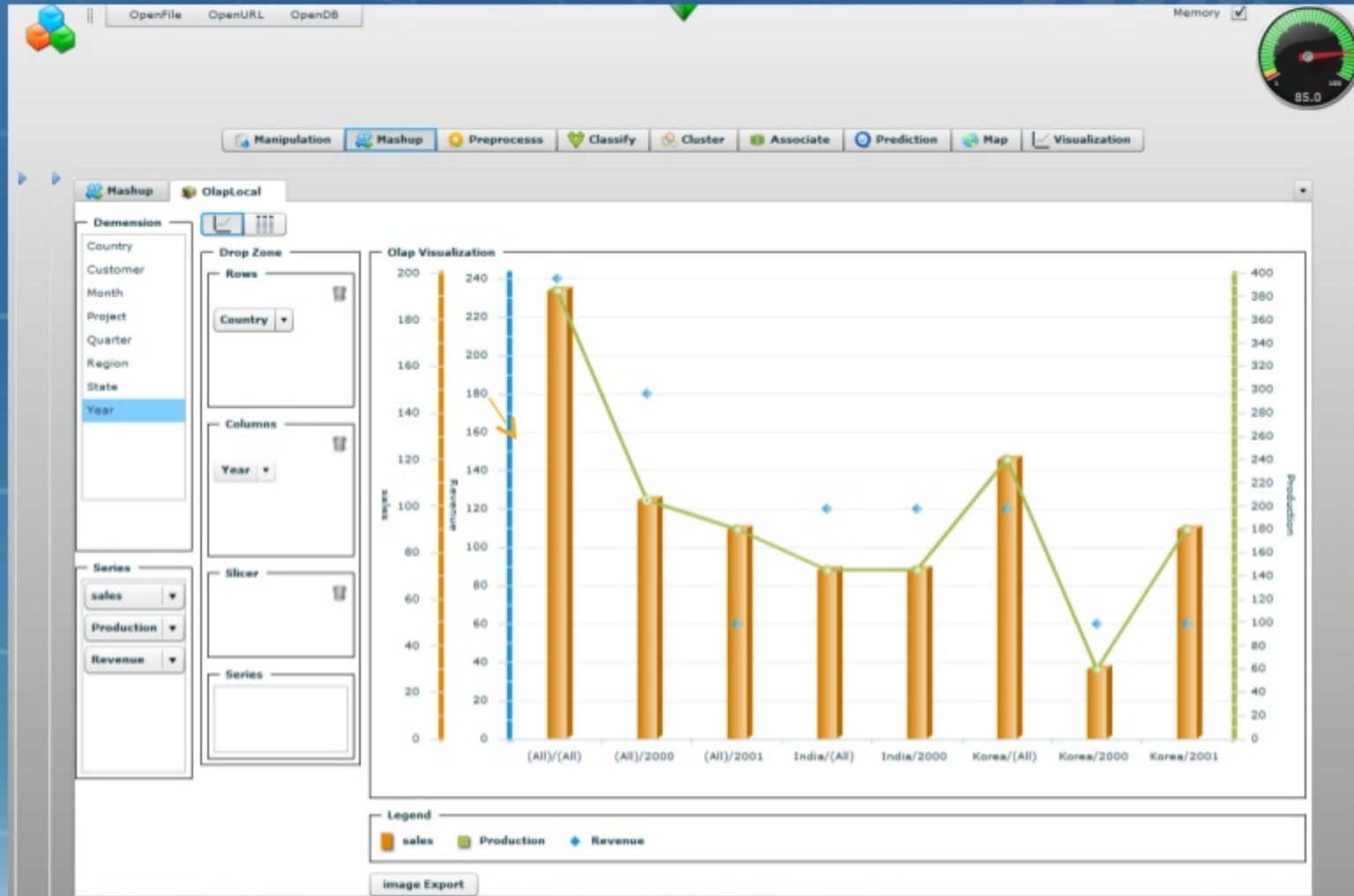
Computer Vision on Hadoop

## Face Recognition from Hadoop

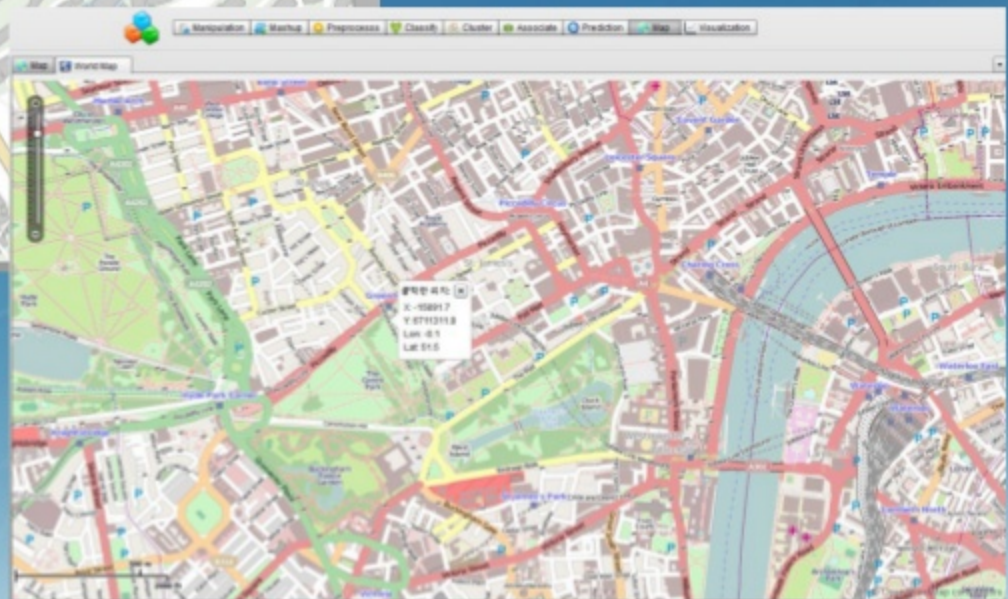
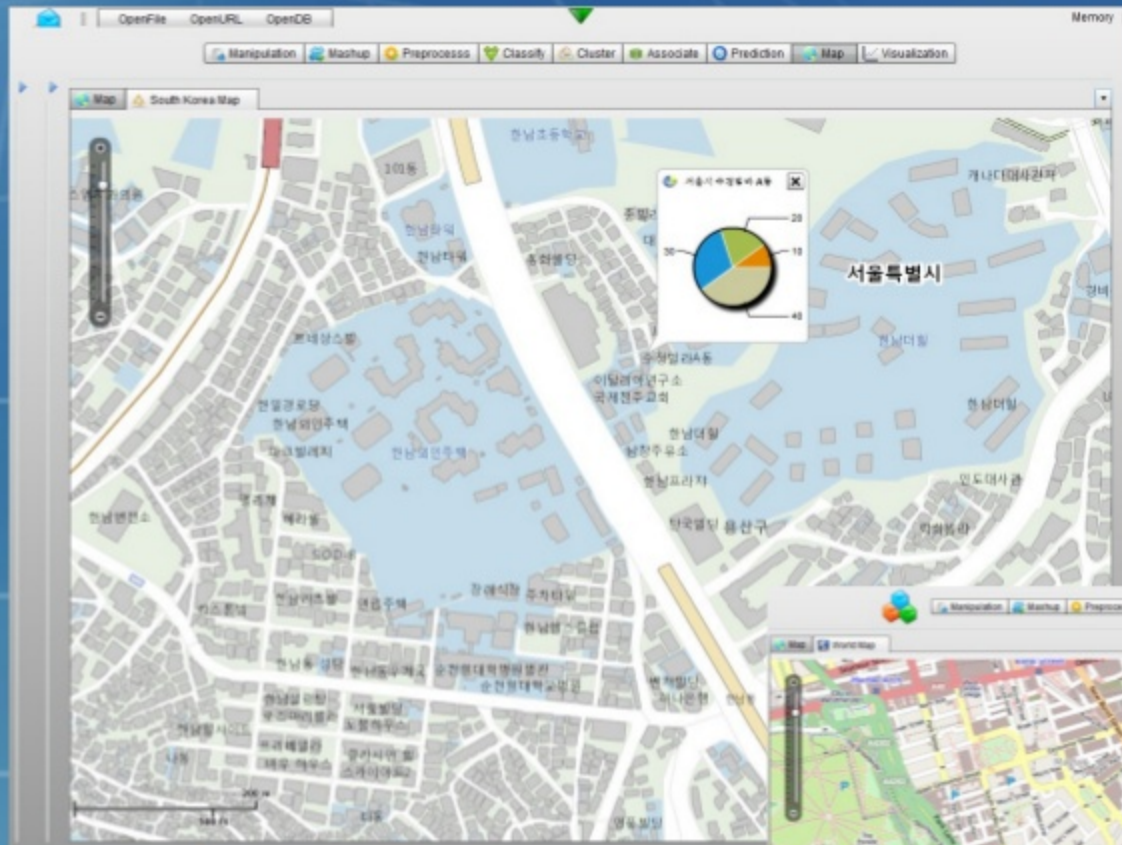




## OLAP Analysis from Data Warehouse



## GEO Analysis for location data





## PDF Parser on Hadoop HDFS

**Singular Value Decomposition Tutorial**

Kirk Baker  
March 29, 2005

**Contents**

key	value
1 Introduction	2
2 Points and Space	2
3 Vectors	3
4 Matrices	3
4.1 Matrix Notation	4
5 Vector Terminology	5
5.1 Vector Length	5
5.2 Vector Addition	6
5.3 Scalar Multiplication	6
5.4 Inner Product	6
5.5 Orthogonality	6
5.6 Normal Vector	7
5.7 Orthonormal Vectors	7
5.8 Gram-Schmidt Orthonormalization Process	7

ROUGH DRAFT - USE AT OWN RISK: suggestions kbaker@ling.osu.edu

7 Singular Value Decomposition 14

7.1 Example of Full Singular Value Decomposition ..... 15

7.2 Example of Reduced Singular Value Decomposition ..... 21

8 References 23

1 Introduction

Most tutorials on complex topics are apparently written by very smart people whose goal is to use as little space as possible and who assume that their readers already know almost as much as the author does. This tutorial's not like that. It's more a manifesto for the rest of us. It's about the mechanics of singular value decomposition, especially as it relates to some techniques in natural language processing. It's written by someone who knew zich about

## Stock Market Analysis using ANN

- Recurrent Neural Network
- Feedforward Neural Network(Elman)
- Counterpropagation Neural Network
- Hopfield Neural Network
- Boltzman Neural Network
- Neat Neural Network
- Som Neural Network
- Art 1 Neural Network
- Bam Neural Network
- Adaline Neural Network

## Trade Market Prediction

