

Course Overview

406.424 Internet Applications

Jonghun Park

jonghun@snu.ac.kr

**Dept. of Industrial Eng.
Seoul National University**

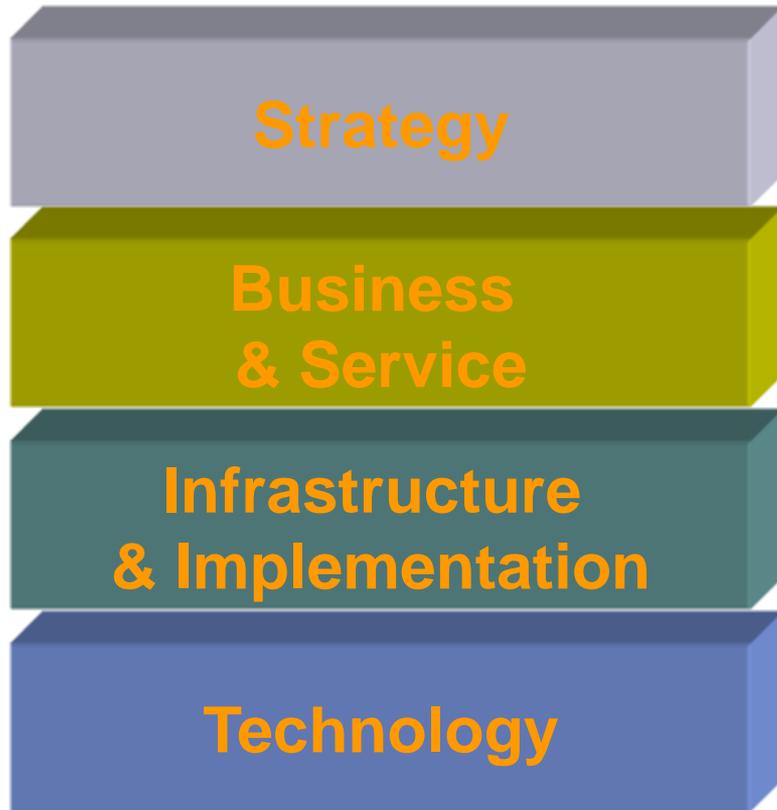
9/1/2010



internet services: historical perspective



internet business: a layered view



a lot of celebrities



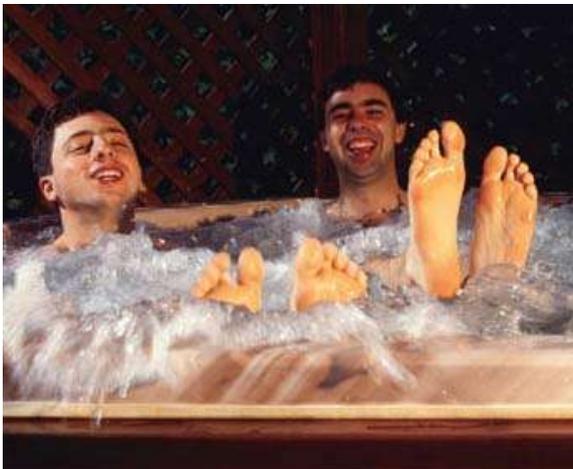
frontiers in internet services (as of 2007)

기 업	성 명	연 령	출 신 지	국 적	출 신 학 교	전 공	성 별
구글	세르게이 브린	34	러시아	미국	스탠포드	컴퓨터	남
구글	래리 페이지	34	미국	미국	스탠포드	컴퓨터	남
구글	에릭 슈미트	52	미국	미국	버클리	컴퓨터 전자	남
디그 닷컴	케빈 로스	30	미국	미국	네바다주립	컴퓨터	남
마이스페이스	크리스 드웬프	41	미국	미국	남가주	경영학	남
아마존	제프 벅조스	43	미국	미국	프린스턴	컴퓨터, 전자공학	남
애플	스티브 잡스	52	미국	미국	리드대 중퇴	물리학, 문학	남
야후	제리 양	39	대만	미국	스탠포드	전자공학	남
야후	테리 세멜	64	미국	미국	롱아일랜드	회계학	남
워드프레스	매트 뮐렌웨그	23	미국	미국	옥스던	정치학, 철학	남
위키피디아	지미 웨일즈	41	미국	미국	인디애나	재무	남
위키피디아	래리 세인저	39	미국	미국	오하이오 주립	철학박사	남
유튜브	스티브 첸	29	대만	미국	일리노이	컴퓨터	남
유튜브	채드 헐리	30	미국	미국	인디애나	미술	남
이베이	피에르 오미디어	40	프랑스	미국	터프츠	컴퓨터 공학	남
이베이	맥 휘트먼	51	미국	미국	프린스턴	경제학	여
주스트	니콜라스 천스트룀	41	스웨덴	스웨덴	을살라	물리학 경영학	남
딴오라일리	딴오라일리	52	아일랜드	미국	하버드	고전문학	남
페이스북	마이크 주커버그	22	미국	미국	하버드 중퇴	심리학	남
플리커	스튜어트 버터필드	34	캐나다	캐나다	캠브리지	철학	남



current winner: google

- founders: Larry Page and Sergey Brin
 - both of their fathers were professors
 - PhD advisor: Rajeev Motwani
- mission: organize the world's information and make it universally accessible and useful
 - “don’t be evil”



The Anatomy of a Large-Scale Hypertextual Web Search Engine

Sergey Brin and Lawrence Page

Computer Science Department,
Stanford University, Stanford, CA 94305, USA
sergey@cs.stanford.edu and page@cs.stanford.edu

Abstract

In this paper, we present Google, a prototype of a large-scale search engine which makes heavy use of the structure present in hypertext. Google is designed to crawl and index the Web efficiently and produce much more satisfying search results than existing systems. The prototype with a full text and hyperlink database of at least 24 million pages is available at <http://google.stanford.edu>. To engineer a search engine is a challenging task. Search engines index tens to hundreds of millions of web pages involving a comparable number of distinct terms. They answer tens of millions of queries every day. Despite the importance of large-scale search engines on the web, very little academic research has been done on them. Furthermore, due to rapid advance in technology and web proliferation, creating a web search engine today is very different from three years ago. This paper provides an in-depth description of our large-scale web search engine -- the first such detailed public description we know of to date. Apart from the problems of scaling traditional search techniques to data of this magnitude, there are new technical challenges involved with using the additional information present in hypertext to produce better search results. This paper addresses this question of how to build a practical large-scale system which can exploit the additional information present in hypertext. Also we look at the problem of how to effectively deal with uncontrolled hypertext collections where anyone can publish anything they want.

Keywords

World Wide Web, Search Engines, Information Retrieval, PageRank, Google

1. Introduction

(Note: There are two versions of this paper -- a longer full version and a shorter printed version. The full version is available on the web and the conference CD-ROM.)

The web creates new challenges for information retrieval. The amount of information on the web is growing rapidly, as well as the number of new users inexperienced in the art of web research. People are likely to surf the web using its

<http://www7.scu.edu.au/programme/fullpapers/1921/com1921.htm>

1

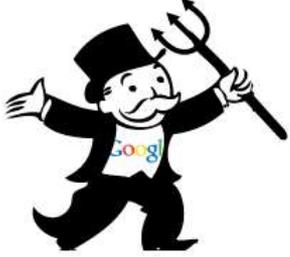
526.80

+0.37 (0.07%)

Pre-market: **529.95** +3.15 (0.60%)

Range	523.48 - 531.75	Mkt cap	167.51B	Shares	317.98M
52 week	289.45 - 629.51	P/E	25.70	Beta	1.13
Open	527.42	Div/yield	-	Inst. own	62%
Vol / Avg.	100.00/3.42M	EPS	20.49		



GO TO DEADPOOL 404 Error - File Not Found	890 SHARES EBAY	150 SHARES GOOGLE	860 SHARES PAYPAL	960 SHARES SKYPE	900 SHARES SOCIAL PLATFORM	920 SHARES FACEBOOK	800 SHARES BOX.NET	FOLLOW INSTRUCTIONS ON TOP CARD I'm Feeling Lucky	800 SHARES LINKEDIN	GOOGLE FOUNDATION • GIVE 75 SHARES TO RANDOM GOOD CAUSE
SALES FORCE.COM 300 SHARES	AOI 300 SHARES	google.org FOLLOW INSTRUCTIONS ON TOP CARD	NEWS CORP 300 SHARES	DEVELOPER PLATFORM 200 SHARES 	I'm Feeling Lucky FOLLOW INSTRUCTIONS ON TOP CARD	YAHOO 350 SHARES	GET IN ACCIDENT • PAY 800 SHARES FOR GOOGLE HEALTH	MICROSOFT 350 SHARES	LOOPT 60 SHARES	COLLECT 300 SHARES OF THE FRANKS & YOUNG 
DOUBLE CLICK 200 SHARES	GRAND CENTRAL 180 SHARES	google.org FOLLOW INSTRUCTIONS ON TOP CARD	URCHIN 180 SHARES	DESKTOP PLATFORM 200 SHARES 	YOUTUBE 160 SHARES	SKETCHUP 140 SHARES	GOOGLE AMERICA 150 SHARES	KEYHOLE 140 SHARES	WRITELY 120 SHARES	JOTSPOT 100 SHARES
 <h1>Googolopoly</h1> <p>Welcome to the game of Googolopoly, where the goal is to organize all of the world's information. To achieve this you can buy internet properties or build them yourself. Once you have built a product or acquired a company, you can invest in developers and servers.</p> <p>Brought to you by Box.net</p>										
JUST VISITING 404 Error - File Not Found				BLOGSPOT 100 SHARES	MOBILE PLATFORM 200 SHARES 	INCOME TAX • Pay your lobbyists in Washington DC 10% or 800 Shares	SCRIBD 60 SHARES	google.org FOLLOW INSTRUCTIONS ON TOP CARD		



and lots of startups



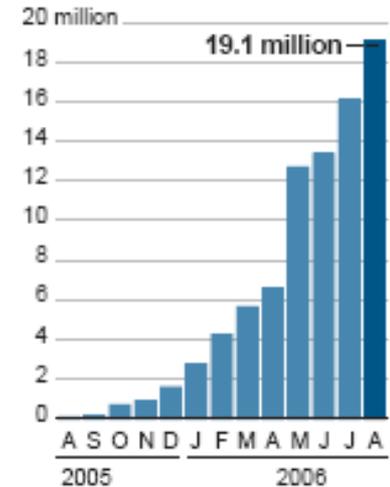
one happy ending: youtube.com

The screenshot shows an MSNBC news article. The main headline is "Google buys YouTube for \$1.65 billion" in red text. Below it is a sub-headline: "Search giant's purchase of video sharing service biggest in its history". The article is attributed to AP Associated Press and is dated October 10, 2006. The text describes Google's acquisition of YouTube for \$1.65 billion, highlighting it as the most expensive purchase made by Google during its eight-year history. A small inset image shows the YouTube website interface at the time, with the "Launch" button visible. A "Story continues below" link is at the bottom of the article.

On the 'Tube'

The number of visitors to YouTube.com skyrocketed in the past year.

Monthly visitors to YouTube



SOURCE: comScore

AP

- monthly UV: 16M (at the time of acquisition)
-> valuation: \$103 per user
- one of exit strategies for startups



current winner in korea



- 1  371억회
- 2  85억회
- 3  32억회
- 4  22억회
- 5  20억회

8월 한달간 검색쿼리 기준
(전 세계 610억회)

편집=서명덕 기자

NHN 임원 보유주식 평가액

2007. 10.8 종가, 단위:주, %, 억원

이름	보유주식수	지분율	평가총액
이해진	2,454,883	5.11	6333.5
최휘영	73,900	0.15	190.6
김범수	896,504	1.87	2312.9
허훈	15,030	0.03	38.7
김정호	270,800	0.56	698.6
천양현	37,000	1.01	95.5
김기섭	8,970	0.02	23.1
부경훈	8,970	0.02	23.1
허진호	8,970	0.02	23.1
김희숙	300,000	0.72	774.0
문태식	261,147	0.66	673.5
오승환	452,479	0.94	1167.3
유영복	8,970	0.02	23.1
문장원	60,000	0.14	154.8
이승원	2,919	0.02	7.5
김병관	178,970	0.37	461.7
허민	28,003	0.06	72.1
최병엽	16,500	0.03	42.5
김근회	1,383	0	3.5
황성순	75,000	0.2	193.5
김창근	17,042	0.04	43.9
박상진	10,000	0	25.8
이준호	1,145,400	4.93	2955.1



happy endings in korea



The Next Big Thing: Mobile Web



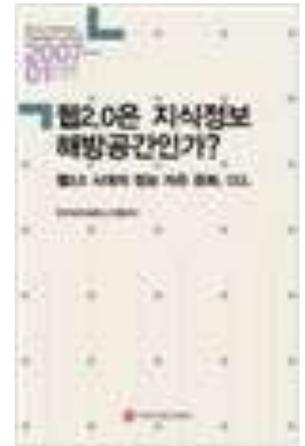
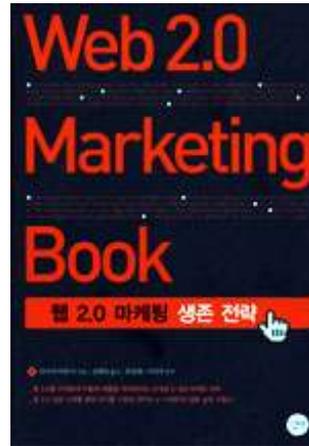
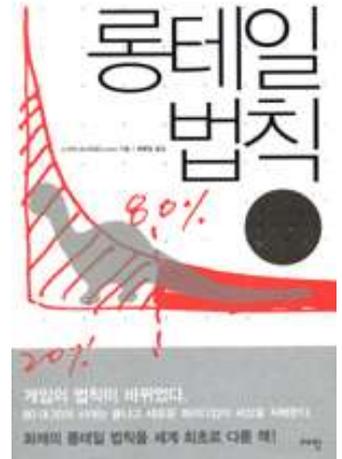
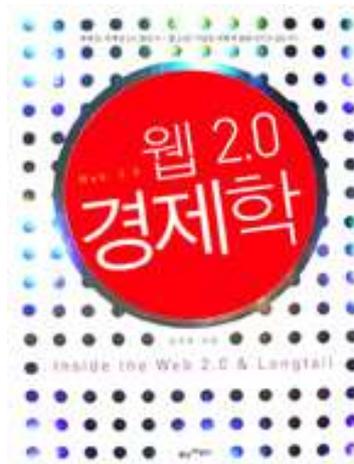
◆애플 아이폰

◆구글 G1

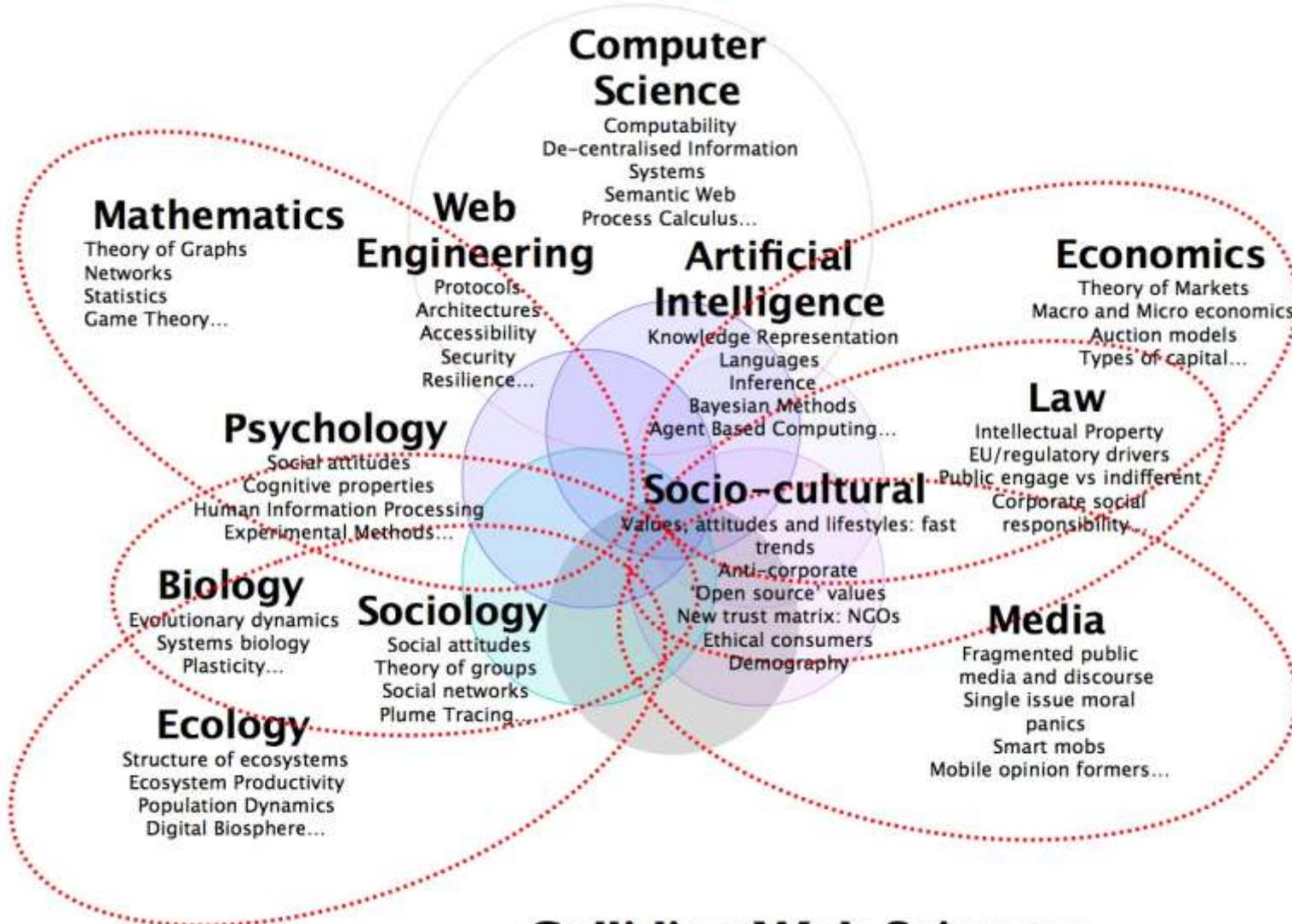
◆삼성전자 T옴니아

Starting 2010!





Web Science >>> CS



Colliding Web Sciences



course objectives

- understanding **internet** and **web** from technology as well as business points of view
- learning **information retrieval** theory
 - core theory for WWW and information systems research
 - based on probability & statistics
- focus on **technical issues**
 - to make your ideas realizable
 - beyond the ppt business models
 - to give you competitive edges in the market with “no entrance barrier”
 - to make you imagine better



search engine example: FeedMil.com

web 2.0

surprising well known

feedmil
a "long tail" feed search engine

web 2.0 feeds

results 1 - 30 of 17,400

feed type: **all feeds** sort by: **feedmil rank** filter by: **english** **latest filters**

Social Networking and Web 2.0

Viral Marketing, Web 2.0, Video Sharing, Social Networking and More

topic: **google**
popularity: authority: activity: frequency: 11.2 posts per week

GigaOM

Business, Internet, Technology & Strategy

topics: **aster data system, gigaom, initial public offering, mobile internet devices, obama for america, private equity group, sponsor, terms of service, venture capital firm, windows mobile marketplace**
popularity: authority: activity: frequency: 60.0 posts per week

M Mashable!

Social software and social networking 2.0.

topics: **oscar o brien, dallas fort worth, ding dong ditch, dog a life, facebook, google image search, google street view, natural language processing, pirate bay trial, twitter**
popularity: authority: activity: frequency: 107.7 posts per week

ReadWriteWeb

topics: **blackberry enterprise server, google, google friend connect, search, time to learn, twitter, user, wall st journal, wall street journal, web office suite**
popularity: authority: activity: frequency: 73.7 posts per week

topic significance



無名な 有名な

feedmil
リアルタイムフィード検索

ホットペッパー feeds

results 1 - 10

hot now **じゃんで★ちーにんぐ**

feed type: **blogs** microblogs podcasts public media social media

エディソンチャン 流矢画像 まとめ

エディソンチャン 流矢画像 まとめ
writer: **ホットペッパー**, **ホットペッパー**
popularity: - updated: 1 day ago - frequency: 8.8 entries / week

青春に悔いなし!同級生卒業記念画像 無料 ダウンロード

青春に悔いなし!同級生卒業記念画像 無料 ダウンロード
writer: **ホットペッパー**
popularity: - updated: 1 day ago - frequency: 13.1 entries / week

SVM日記ブログ

話題やニュースにSVMがふよやく日記通信ブログ
writer: **ホットペッパーの本0**
popularity: - updated: 1 day ago - frequency: 0.7 entries / week

ガンダムvsガンダム iso

ガンダムvsガンダム iso
writer: **ホットペッパー**
popularity: - updated: 1 day ago - frequency: 8.5 entries / week

ドクターコバ ラジオ 影乃かなみ

ドクターコバ ラジオ 影乃かなみ
writer: **ホットペッパー**
popularity: - updated: 1 day ago - frequency: 9.5 entries / week

ホットペッパー

ホットペッパー(Hot Pepper)ってご存知ですか?無料クーポン付きフリーペーパーです。
writer: **真人ドスケベGALS 完全版+潮吹きで撃沈!レスで悶絶編-[DVD]**
popularity: - updated: 1 week ago - frequency: 0.1 entries / week

what's new?

1

FriendFeed - masateru mentioned:
'神戸ダイニング 「area3」(FocMoo by ホットペッパー(HotPepper))' (5 hours ago)

bigweatのブックマーク published:
'神戸ダイニング 「area3」(FocMoo by ホットペッパー(HotPepper))' (5 hours ago)

Twitter / mina8823 mentioned:
'mina8823: ホットペッパーめくりながら、なぜかぐるなびのサイトと照らし合わせて何度も「んん?」ってなりながら店探してたwwwどんくさず」' (13 hours ago)

Videos related to 「さゆ 「アイドリグ大好き」」 published:
'モーニング娘。 ホットペッパーのCM 初編' (10 hours ago)

Videos related to 「グイフハウスCM〜ジャック魔」 published:
'プリズンブレイク+ホットペッパー-総集編' (18 hours ago)

bookmark published:
'京博館賞物展 竹取姫めぐりFocMoo by ホットペッパー(HotPepper)' (21 hours ago)



search engine example: LiveK.com



2010동계올림픽에 대한 실시간 검색중

최선순 | 인기순

27초전 우리의 피겨여왕 **김연아** 선수- 청와대 행사만 하시고, ...
우리의 피겨여왕 **김연아** 선수- 청와대 행사만 하시고, 꼭 쉬고 가시기를- 대통령 이 토리노로 가면 안되겠나?! 우리에게 더 소중한 사람이라고-!!**김연아** 선수에
Twitter / cylee624 2010.3.1 (월) 오후 8:07



59초전 RT @chanjin 정말 청와대 행사 때문에 **김연아** 선수가 1박 ...
RT @chanjin 정말 청와대 행사 때문에 **김연아** 선수가 1박 2일로 들어오는 건가요? @saunakim http://tl.gd/corha
Twitter / bluerose78 2010.3.1 (월) 오후 8:06



1분전 우리 국민들은 **김연아**가 피곤하고 지친 몸으로 청와대 ...
우리 국민들은 **김연아**가 피곤하고 지친 몸으로 청와대 끌려가서 얼굴마담 노릇하다가 정신없이 토론토로 날아가는거보다 꼭 쉬고 힘차게 훈련 다시 시작하는걸
Twitter / lion_cat 2010.3.1 (월) 오후 8:06



1분전 이것이 이번 동계올림픽 공식 메달 집계라는 ^^ 메달의 ...
이것이 이번 동계올림픽 공식 메달 집계라는 ^^ 메달의 가치는 색깔 순이 아니라능. ㅎㅎ http://yfrog.com/2enntj
Twitter / ParkKD 2010.3.1 (월) 오후 8:05



2분전 정말 청와대 행사 때문에 **김연아** 선수가 1박 2일로 들어오는 ...

LIVE K 아이폰앱 출시
Available on the iPhone
App Store

찾아주세요 >

- 1 김연아 -
- 2 소치 -
- 3 김연아 동계올림픽 -
- 4 소치 동계올림픽 -
- 5 한일 사이버전쟁 -
- 6 김연아 비방 -
- 7 국가대표 +
- 8 김연아 갈라쇼 -
- 9 삼일절 폭주족 -
- 10 은지원 유해판정 -

내 관심 이슈 >

2010동계올림픽 x

박종현 교수 x



recommendaion engine example

PixCow 멀티미디어 정보마당

로그아웃 | 회원정보 | 마이페이지 | 고객센터

카테고리 | BIZ센터 | 오늘본영상 | 스크랩영상 | 구매영상 | 즐겨찾기 | 통합검색 | SEARCH

[공지] PixCow Dance Club Grand Open | 더보기+ | drburix02님 | 치즈 : 0조각 | 관리 | My Shop | 동영상업로드 | 치즈충전

Web Award Korea 2007 동영상 부문 대상 (선정기관: 플어워드위원회) | Marketplace | Multimedia Business | UCC

Mini Shop drburix02님께 추천하는 영상 목록입니다

SEARCH Mini Shop 검색어를 입력하세요

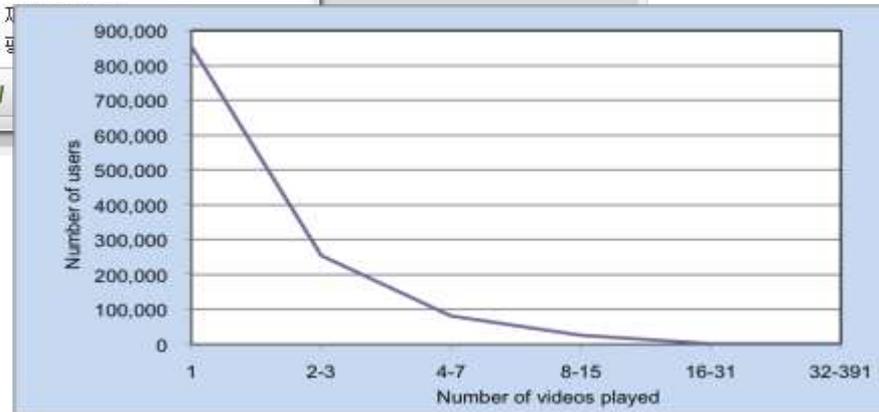
+ 관련 영상

과일 예쁘게 깎는 방법 [1 / 20]

전체보기+

영상보기

과일 예쁘게 깎는 방법
 등록자 : 밝은향기
 등록일 : 2006-08-02 11:06:18
 재생시간 : 00:11:10



let's get started

- course home page: <http://imlab.snu.ac.kr/>
- some comments on the course syllabus
- TA
 - Mr. Yongwook Shin (Ph.D. student at information management lab, dept. of IE)
 - Email: yongwook.shin@gmail.com
- survey questions
 - web & internet experience
 - search engine experience
 - fluency in web programming languages and platforms
 - fluency in XML



term project (35%)

- topic: search service implementation using open source search engine
- project description: distributed on March 18, 2010
- team composition: March 16, 2010
 - up to 3 people
- system design (5%): April 20 & 22, 2010
 - design presentation
 - report
- system implementation (30%): June 8 & 10, 2010
 - final presentation
 - report
 - demo
 - search quality



homework #0 (not scored)

- create a twitter account
- follow the tweets from
 - TechCrunch
 - rww
 - mashable

