

INF3800/INF4800

Søketeknologi

2012.01.16

Aleksander Øhrn

aleksaoh@ifi.uio.no

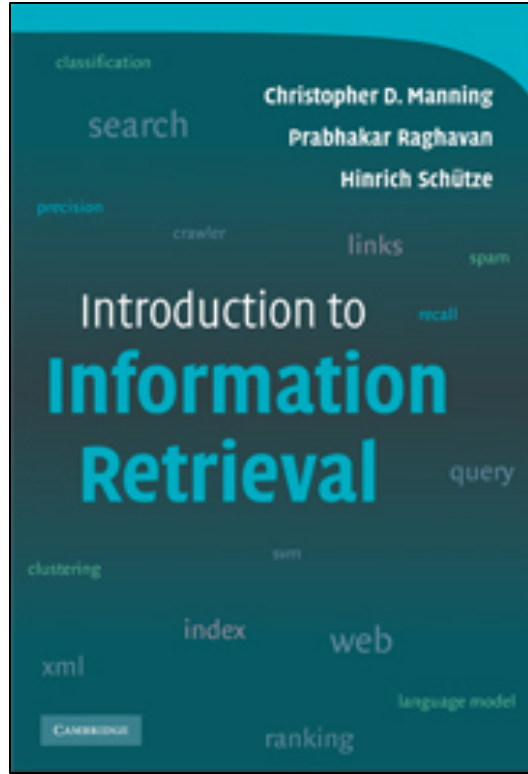
Pierre Lison

plison@ifi.uio.no

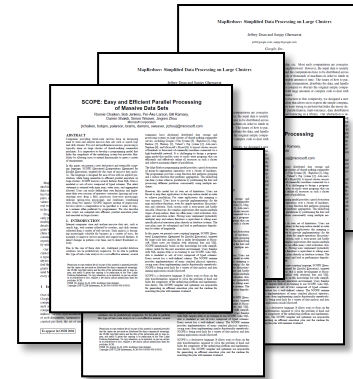
<http://www.uio.no/studier/emner/matnat/ifi/INF3800/v12/>

Pensum

<http://nlp.stanford.edu/IR-book/information-retrieval-book.html>



+




Øvinger

DRAFT

Algoritme*

X



Eksamen



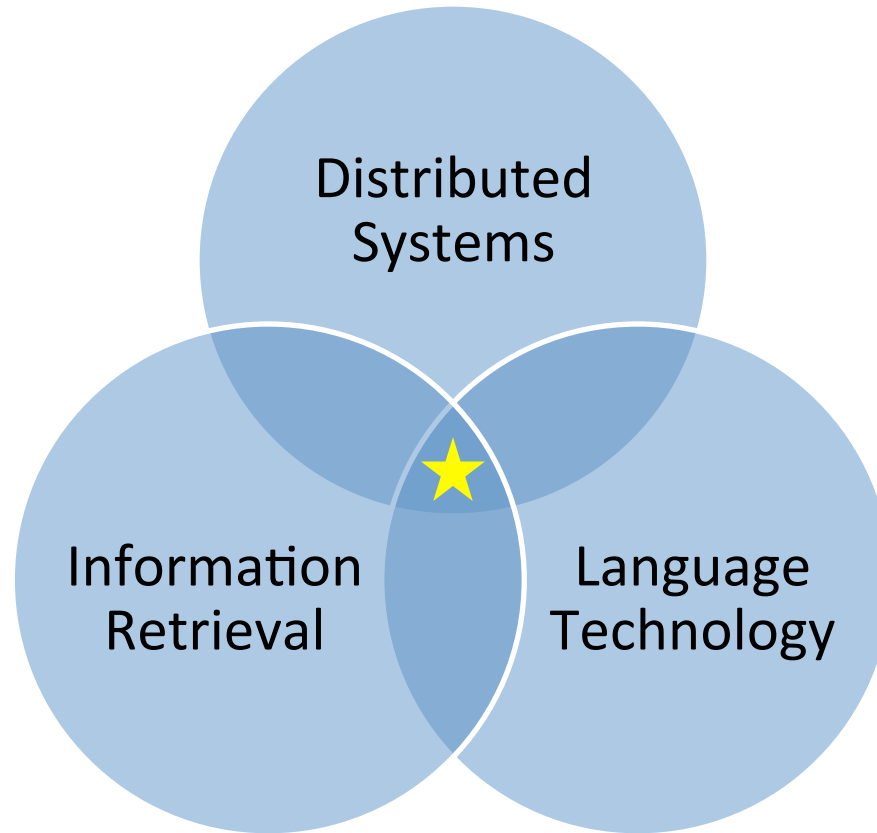
12. juni 2012

14:30

4 timer

Introduksjon

The Sweetspot



Web Search

Plan 9 from Outer Space - Bing - Windows Internet Explorer

http://www.bing.com/search?q=Plan+9+from+Outer+Space&FORM=RSFD4

Plan 9 from Outer Space

ALL RESULTS 1-19 of 4,030,000 results

Images of Plan 9 from Outer Space

Videos of Plan 9 from Outer Space

Plan 9 from Outer Space - Wikipedia, the free encyclopedia

Plan 9 from Outer Space (originally titled as Grave Robbers from Outer Space) is a 1959 science fiction / horror film written, and directed by Edward D. Wood, Jr. Synopsis · Cast · History and development · Mistakes

Plan 9 from Outer Space (1959)

Aliens resurrect dead humans as zombies and vampires to stop human kind from creating the Solaranite (a sort of sun-driven bomb).

PLAN 9

Enter Site Copyright © 2009 Darkstone Entertainment. All rights reserved. plan9movie.com

Videos of Plan 9 from Outer Space

plan 9 from outer space (part one) Dailymotion 10:01

plan 9 from outer space (part three) Dailymotion 10:07

plan 9 from outer space (part two) Dailymotion 9:59

Plan 9 from outer space part 8 Dailymotion 6:21

elvis costello - Google Search - Windows Internet Explorer

http://www.google.com/search?hl=en&q=elvis%20costello

elvis costello

News results for elvis costello

Costello's poses nothing short of sublime - 3 hours ago

Not only is **Elvis Costello** one of rock 'n' roll's revered misfits, Sunday's gig was also dedicated to raising money for the Fort Edmonton Foundation and its ...

Spectacle Elvis Costello with ...

Official site with Island Records includes news, biography, sound and video clips, appearance schedule, "Ask Elvis" feature, message board and desktop ...

Elvis Costello - Wikipedia, the free encyclopedia

Declan Patrick MacManus (born 25 August 1954), known by the stage name **Elvis Costello**, is an English singer-songwriter of Irish heritage. ...

The Elvis Costello Home Page - The Elvis Costello Home Page

This is the site to visit for everything concerning **Elvis Costello**. Extensive information about forthcoming events, complete concert listing, discography of ...

Image results for elvis costello - Report images

Elvis Costello on MySpace Music - Free Streaming MP3s, Pictures ...

MySpace Music profile for **Elvis Costello**. Download **Elvis Costello** Rock // music singles, watch music videos, listen to free streaming mp3s, & read Elvis ...

Elvis Costello Guide - Discography | Setlists | Songs | Lyrics ...

Elvis Costello Guide - Complete Costello resource with discography, set lists, web sites list, pictures, photos, recordings, CDs, books, movies, videos, ...

Elvis Costello - Discover music, videos, concerts, & pictures at ...

Watch videos & listen to **Elvis Costello**. Alison, Pump N' In & more... plus 46 pictures. Da'lan

lost - Yahoo! Search Results - Windows Internet Explorer

http://search.yahoo.com/search_yftA0geu4zKFLyOQ4m5tX1y0a7p=lost&fr2=sb-top&fr=yfp-t

lost

U.S. markets make up lost ground

NEW YORK, Feb. 9 (UPI) -- U.S. markets swung higher Tuesday, recovering ground lost during a three-week trend that has investors concerned a 10 percent downward... full story

Lost party at The Knitting Factory - New York Post - 1 hour ago

Got Lost plans tonight? Also: Dharma love thrives in Peru - USA Today - 2 hours ago

TV Review: 'Lost,' 'Kate' expectations - Entertainment Weekly - 9 hours ago

Lost - ABC

Official ABC site for **Lost**, the survival drama telling the story of plane crash survivors who find themselves stranded on a mysterious desert island ...

ABC.com - Lost - Episode Guide

Get caught up on the ABC show **Lost**. Read a full recap of the "LA X (Parts 1 and 2)" from Season 6 of the show.

Lost - Wikipedia

Lost is an American serial drama television series. It follows the lives of plane crash survivors on a mysterious tropical island, after a commercial passenger jet flying between Sydney ...

Lost - IMDb

Cast and crew information about **Lost**, the ABC TV drama, with plot outline, trivia, and user comments.

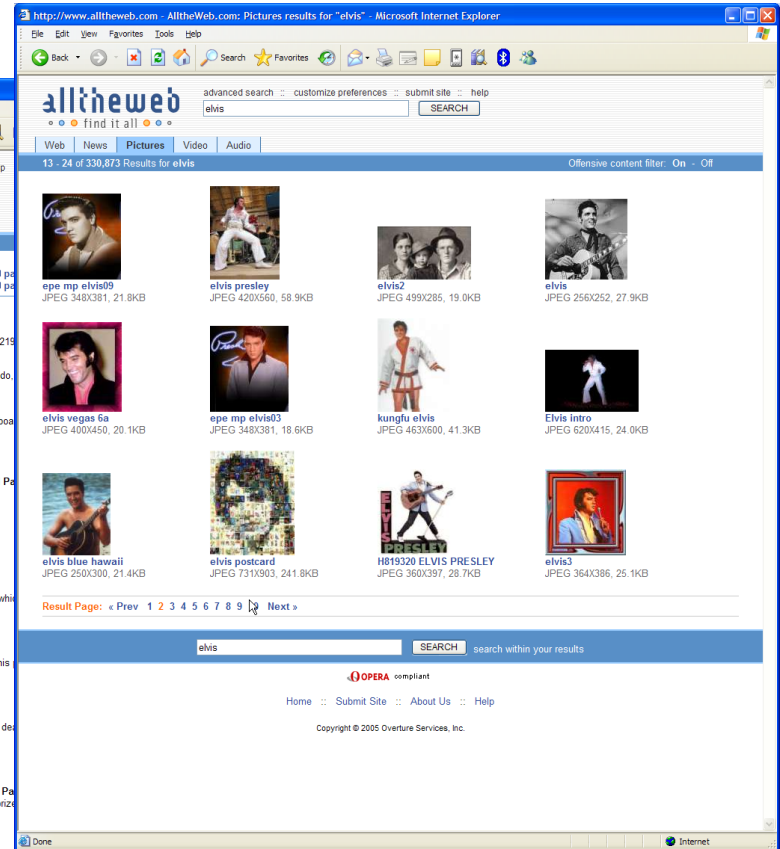
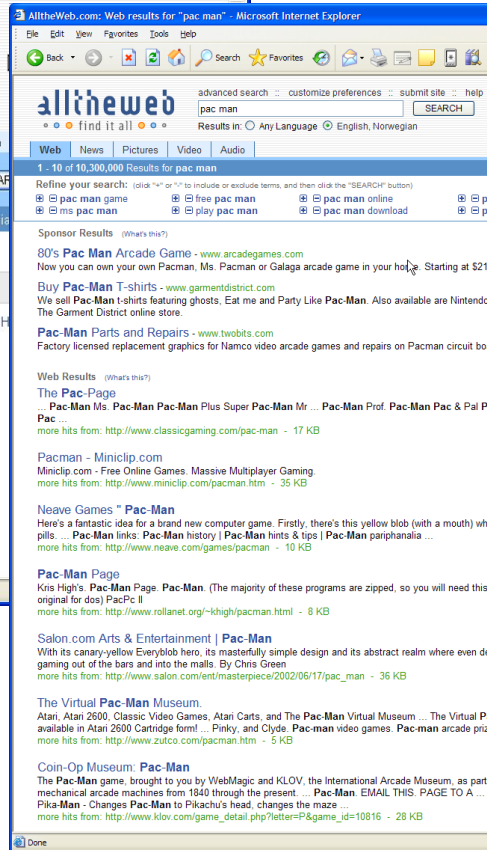
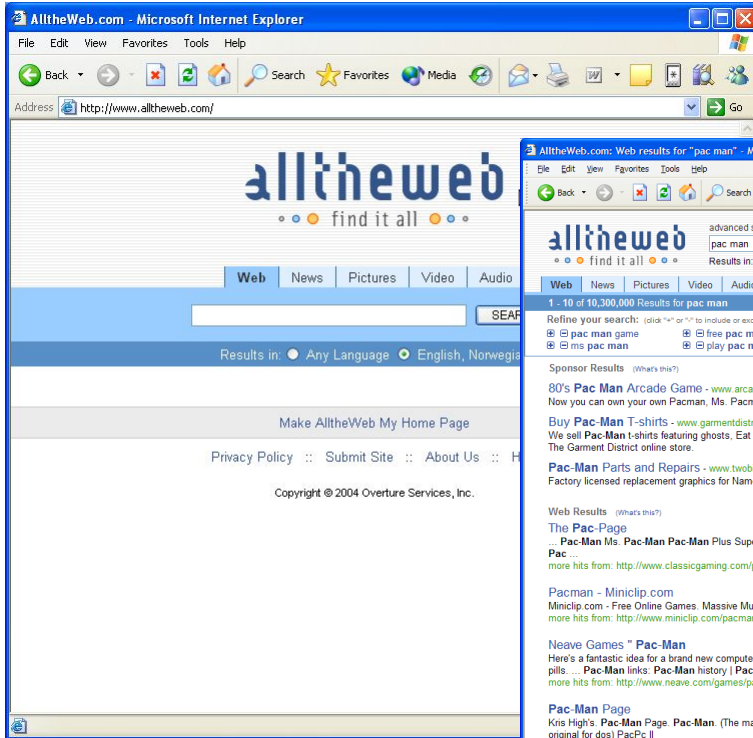
Lostpedia

Encyclopedia wiki fan site keeps track of mysteries, facts, and theories surrounding the ABC TV series, **LOST**. With spoilers, news, and discussion forums.

Lost - TV.com

alltheweb.com

1999-2003



Enterprise Search

Much more than intranets



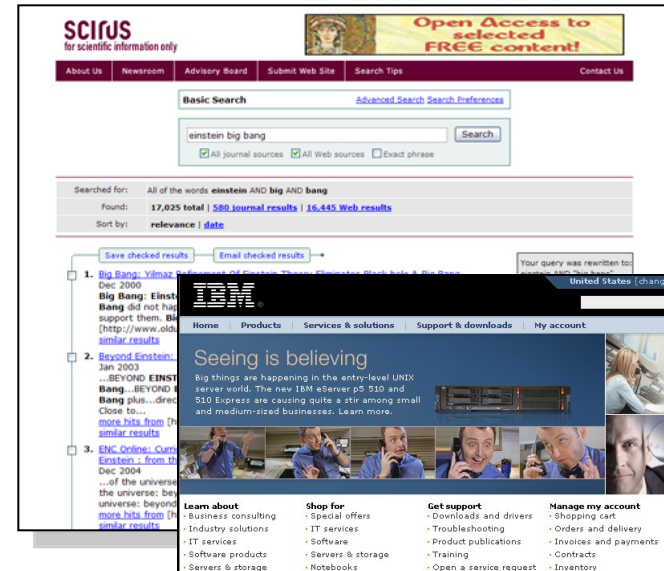
The FIRSTGOV.gov website is the official web portal of the U.S. Government. It features a prominent search bar at the top with the text "Enter Search Term(s)". Below the search bar, there are navigation links for "Home", "About Us", "Site Index", "Help", "Español", and "Other Languages". The main content area is divided into sections for "Citizens", "for Businesses and Nonprofits", "for Federal Employees", and "Government-to-Government". A sidebar on the left lists "By Organization" with links to various agency indexes and "Reference Center" with links to data and statistics, government forms, and laws. A "Citizens: Get It Done Online!" section lists various services like finding government benefits, applying for grants, and renewing driver's licenses.



The careerbuilder.com website is a major online job search platform. It features a navigation bar with links for "Home", "Find Jobs", "Post Resumes", "My CareerBuilder", "Advice & Resources", and "Career Fairs". A central banner asks "WHO ARE THESE MONKEYS?" and lists three steps for finding jobs: 1. Find Jobs, 2. Set Up Job Alerts, and 3. Post a Resume. Below this, there are sections for "QUICK JOB SEARCH" with input fields for keywords, city, and state, and "JOB ALERTS" for finding specific jobs. A "RESUMES" section offers tips for creating effective resumes. The footer lists "MOST POPULAR JOB CATEGORIES" such as Accounting, Contract/Recruiting, Executive Recruitment, Information Technology, Manufacturing, Nonprofit, Retail, and Sales & Marketing.



The Gule Sider website is a Norwegian e-commerce and information portal. It features a search bar at the top with the text "Gule Sider®" and "Telefonkatalogen™". Below the search bar, there are links for "Nummersøk", "Kart", "SMS", and "Logg inn". The main content area includes a "Kartskole/Ruteplanlegger" section with a map and a "Sammenlign og finn laveste pris" section for comparing prices for products like Canon Exus 30 and Apple iPod Photo 4GB. A "Nettbutikker i Norge" section lists various online retailers. The footer contains contact information and a "Nettbutikk i Norge" section.



The SCIRUS website is a scientific information search engine. It features a navigation bar with links for "About Us", "Newstern", "Advisory Board", "Submit Web Site", "Search Tips", and "Contact Us". A "Basic Search" section has a search bar with the text "einsteinst big bang" and options for "Journal sources", "All Web sources", and "Exact phrase". Below the search bar, there are sections for "Searched for:", "Found:", and "Sort by:". The "Found:" section shows "17,025 total | 580 journal results | 16,445 Web results". A list of search results is displayed, including "Big Bang: Yilmaz", "Big Bang: Einstein", "Big Bang: Beyond Einstein", and "ENC Online: Current Events".



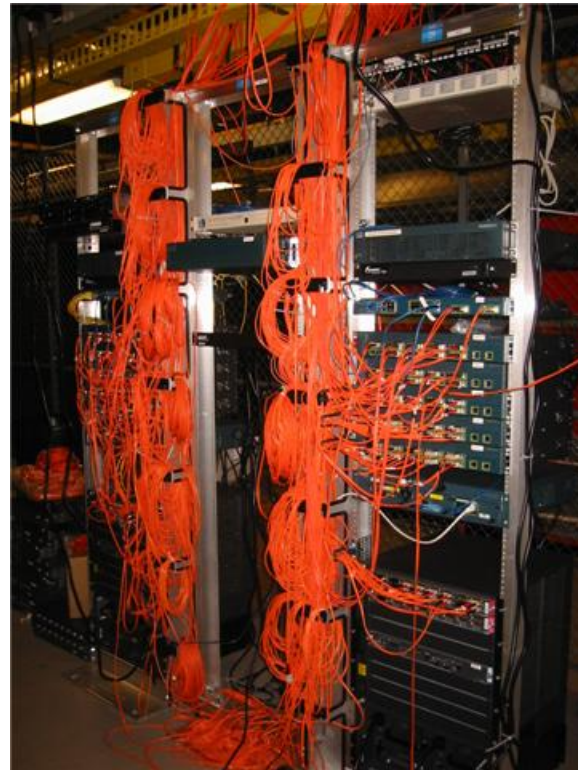
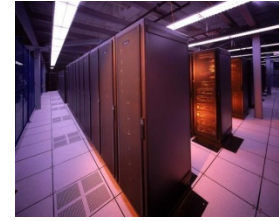
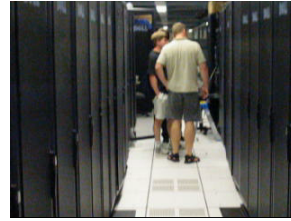
The Rakuten website is a Japanese e-commerce platform. It features a navigation bar with links for "楽天市場" and "楽天モバイル". Below the navigation bar, there are sections for "楽天市場" with various product listings, "楽天モバイル" with service information, and "my Rakuten" with user account information. The footer contains contact information and a "THE HELP DESK IS OPEN" section.

Order your AIX upgrade online. It's never been easier to upgrade your operating system.

THE HELP DESK IS OPEN → Learn more.

Data Centers

alltheweb.com 2000



Data Centers

Microsoft 2010



<http://www.youtube.com/watch?v=K3b5Ca6lzqE>

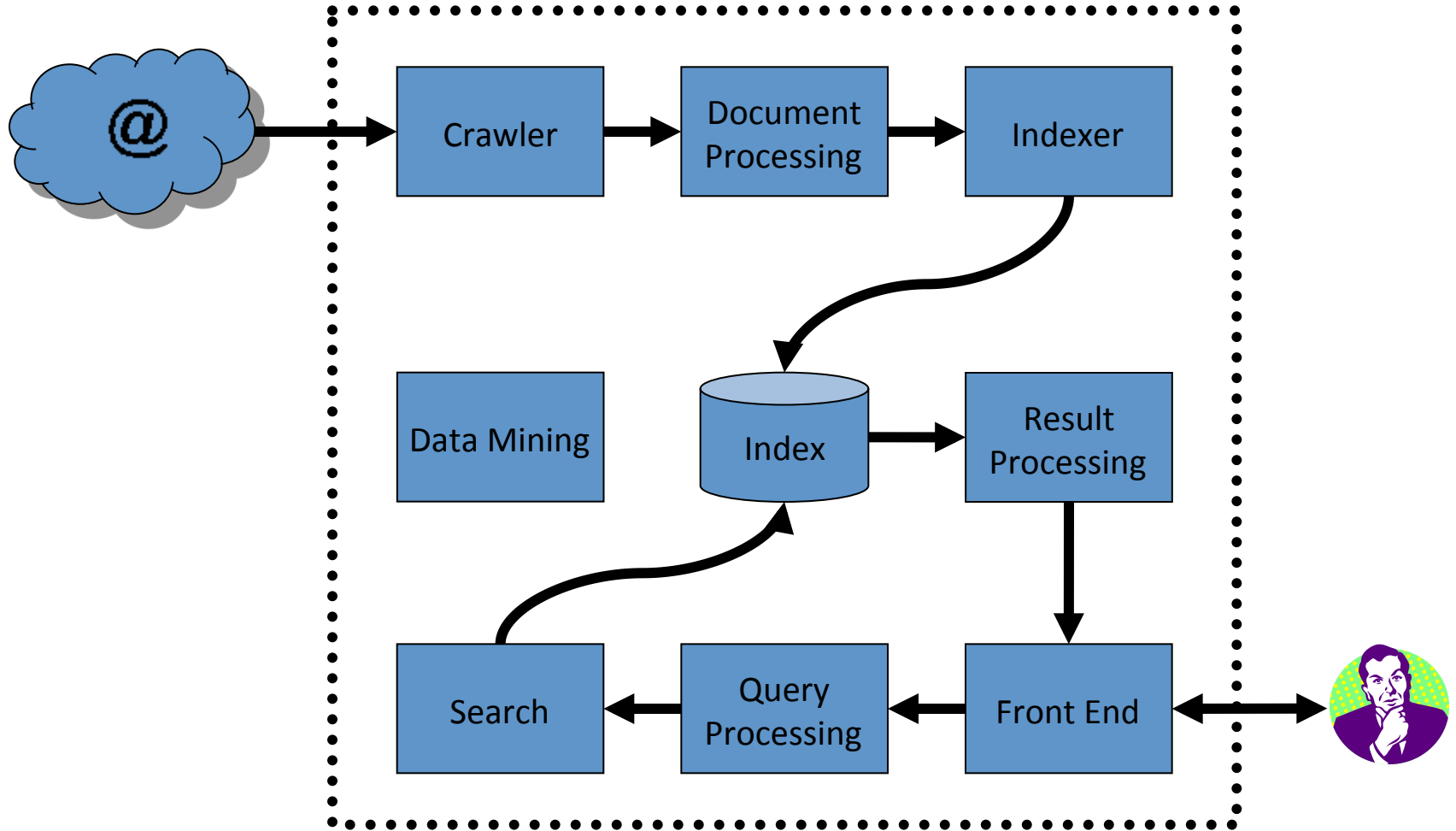
A screenshot of a YouTube search results page. The search bar contains the text "microsoft data centers". Below the search bar, there are several search results. The first result is titled "Microsoft Data Centers" and is a sponsored link from "it-service-managementtechweb.com". The second result is titled "Microsoft OS Cloud Windows Azure Data Center - Google and Amazon" and has a video thumbnail showing a server rack. The third result is titled "Microsoft Generation 4.0 Data Center Vision" and has a video thumbnail showing a data center interior. The page also includes navigation links for "Home", "Videos", and "Channels", and a "Search options" link.



<http://www.youtube.com/watch?v=PPnoKb9fTka>

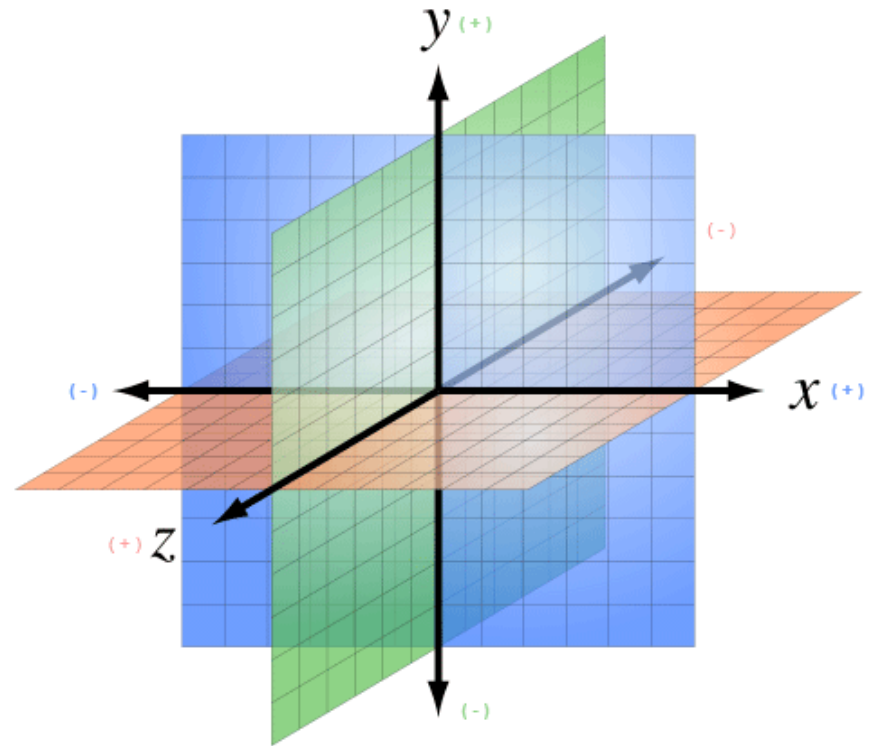
Search Platform Anatomy

The 50,000 Foot View



Scaling

- **Content Volume**
 - How many documents are there?
 - How large are the documents?
- **Content Complexity**
 - How many fields does each document have?
 - How complex are the field structures?
- **Query Traffic**
 - How many queries per second are there?
 - What is the latency per query?
- **Update Frequency**
 - How often does the content change?
- **Indexing Latency**
 - How quickly must new data become searchable?
- **Query Complexity**
 - How many query terms are there?
 - What is the type and structure of the query terms?



Scaling



Query Traffic

Scale through replicating the partitions



Content Volume

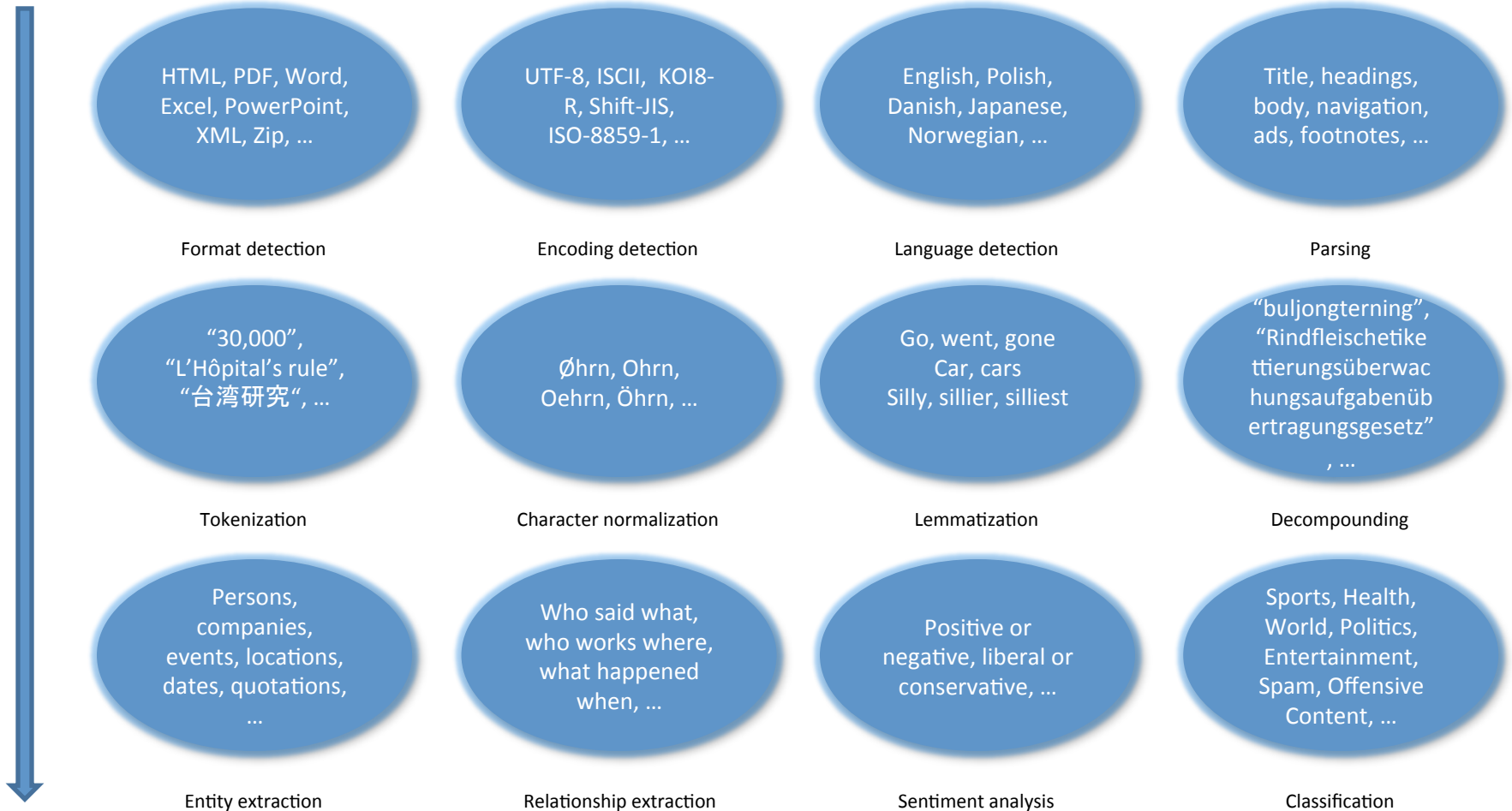
Scale through partitioning the data



Crawling The Web



Processing The Content



Creating The Index



Word	Document	Position
tea	4	22
	4	32
	4	76
	8	3
teacart	8	7
teach	2	102
	2	233
	8	77
teacher	2	57

Processing The Query

"LED TVs between
\$1000 and \$2000"

"I am looking for
fish restaurants
near Majorstua"



"hphotos-snc3
fbcdn"

"brintney speers
pics"

"23445 + 43213"

Searching The Content

Introduction to Information Retrieval | Sec. 2.3

Recall basic merge

- Walk through the two postings simultaneously, in time linear in the total number of postings entries

2 → 8 ← [2 → 4 → 8 → 41 → 48 → 64 → 128] Brutus
[1 → 2 → 3 → 8 → 11 → 17 → 21 → 31] Caesar

If the list lengths are m and n , the merge takes $O(m+n)$ operations.

Can we do better?
Yes (if index isn't changing too fast).

Introduction to Information Retrieval | Sec. 2.3

Augment postings with skip pointers (at indexing time)

41 → 128
2 → 4 → 8 → 41 → 48 → 64 → 128

11 → 31
1 → 2 → 3 → 8 → 11 → 17 → 21 → 31

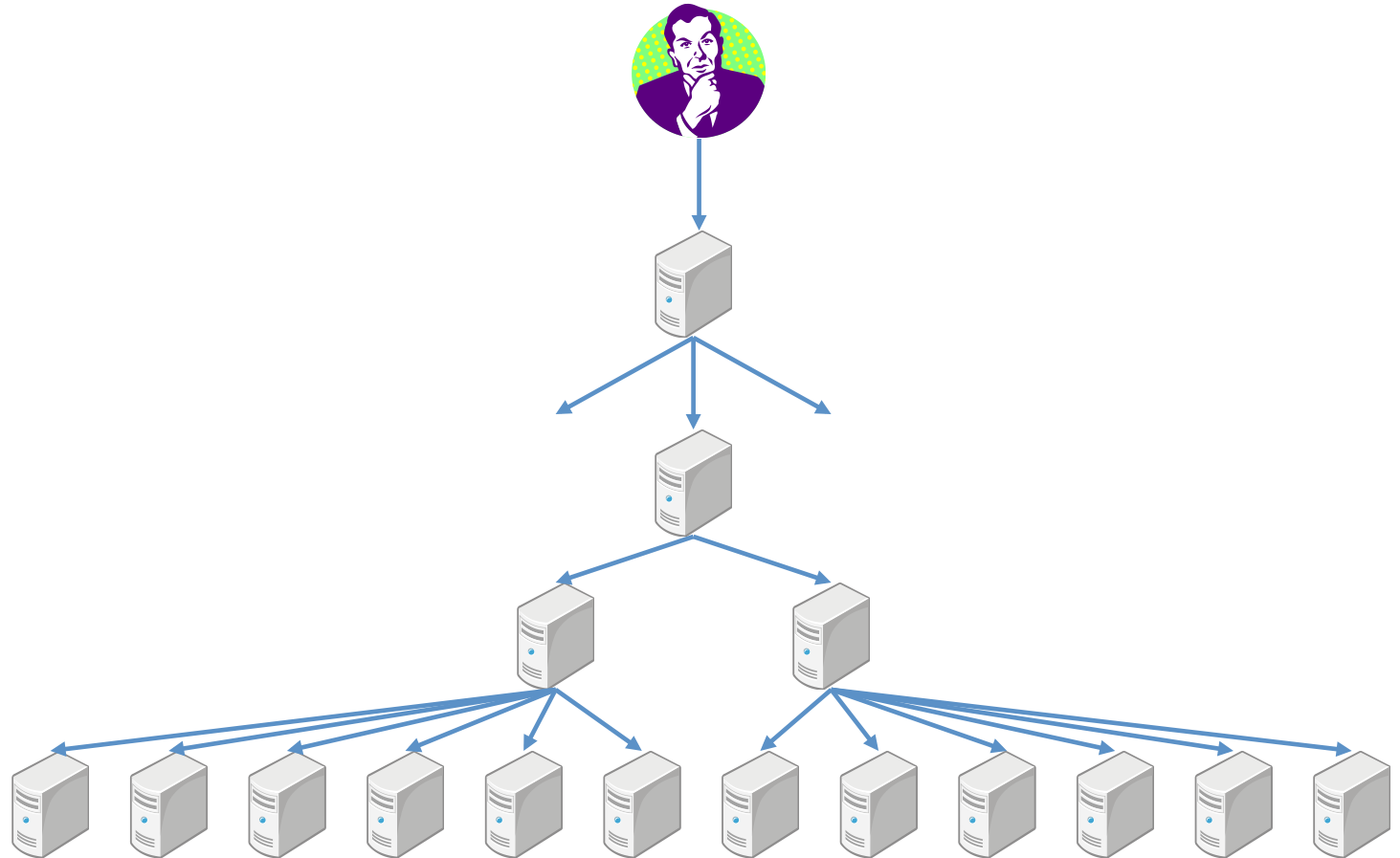
- Why?
- To skip postings that will not figure in the search results.
- How?
- Where do we place skip pointers?

<http://www.stanford.edu/class/cs276/handouts/lecture2-dictionary.pdf>



Assess relevancy as we go along

Searching The Content



Federation
Query processing
Result processing

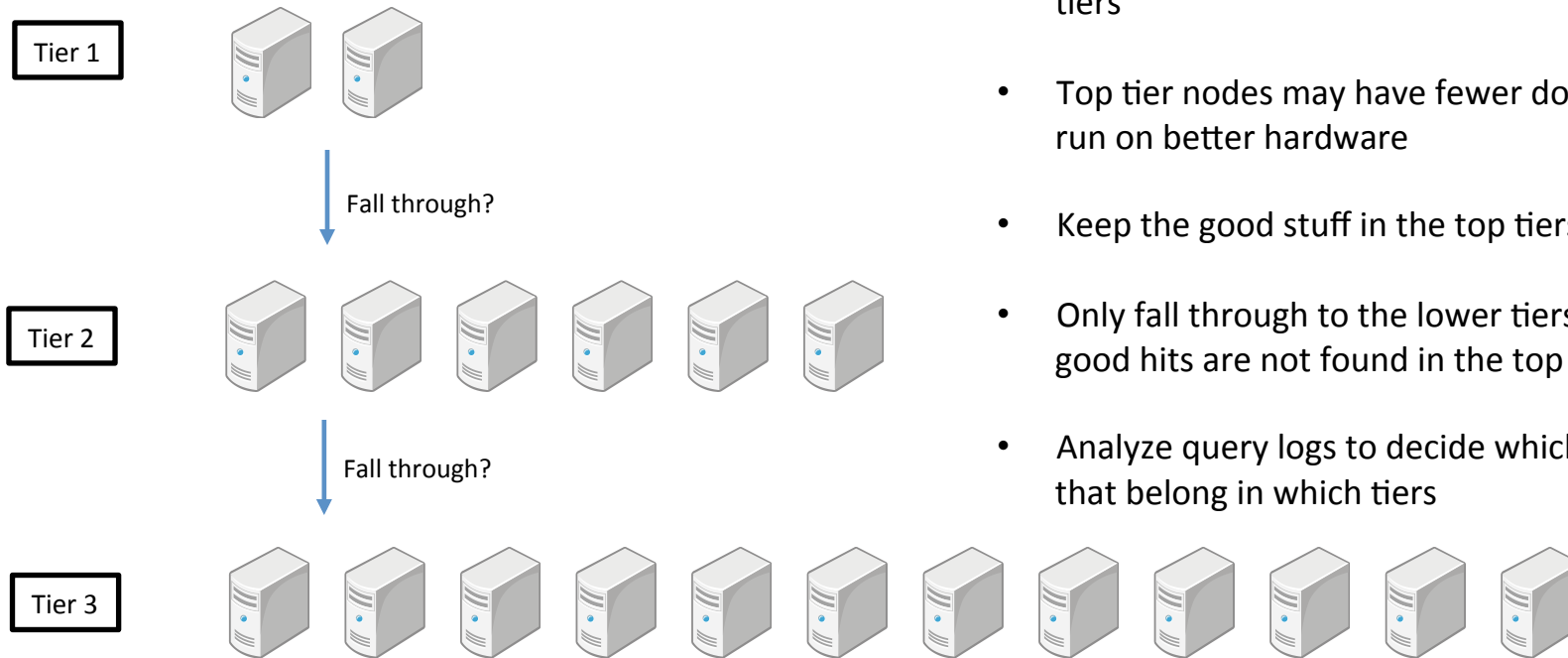
Dispatching
Merging

Searching
Caption generation

“Divide and conquer”

Searching The Content

Tiering

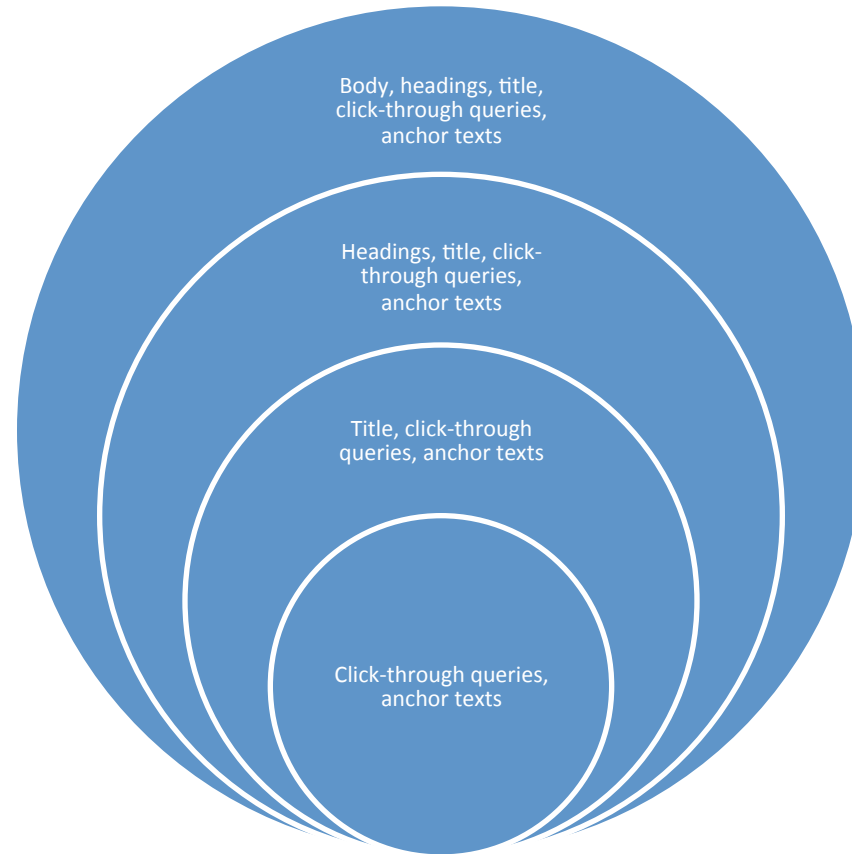


- Organize the search nodes in a row into multiple tiers
- Top tier nodes may have fewer documents and run on better hardware
- Keep the good stuff in the top tiers
- Only fall through to the lower tiers if not enough good hits are not found in the top tiers
- Analyze query logs to decide which documents that belong in which tiers

“All search nodes are equal, but some are more equal than others”

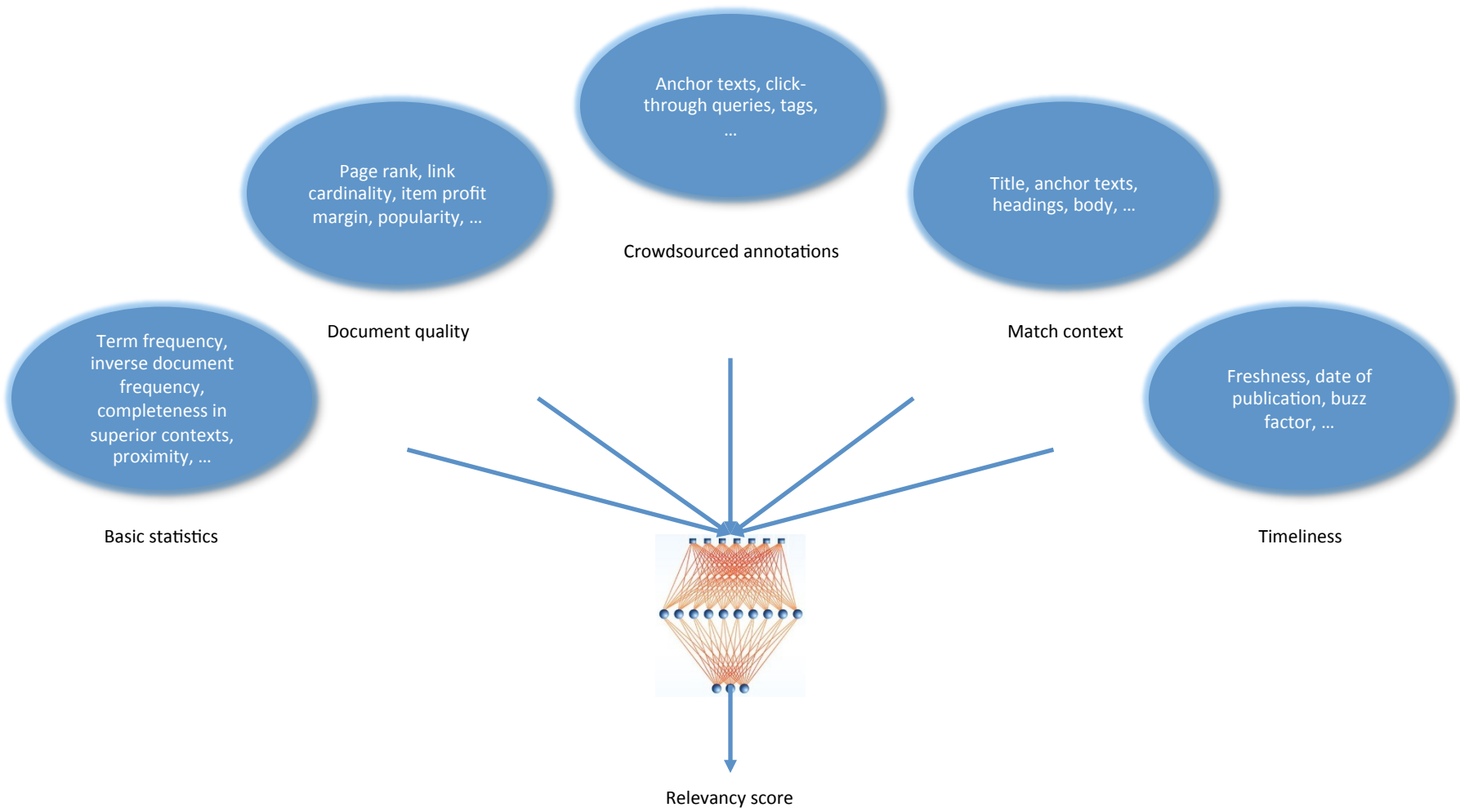
Searching The Content

Context Drilling



“If the result set is too large, only consider the superior contexts”

Relevancy



“Maximize the normalized discounted cumulative gain (NDCG)”

Processing The Results

- **Faceted browsing**
 - What are the distributions of data across the various document fields?
 - “Local” versus “global” meta data
- **Result arbitration**
 - Which results from which sources should be displayed in a federation setting?
 - How should the SERP layout be rendered?
- **Unsupervised clustering**
 - Can we automatically organize the results set by grouping similar items together?
- **Last-minute security trimming**
 - Does the user still have access to each result?

The screenshot displays a search engine interface with several key components:

- Refine Results Panel:** A top panel with filters for Source Title, Author Name, Year, Document Type, and Index Terms. It includes a 'limit to' button and an 'exclude' button.
- Refine your search Panel:** A middle panel with a list of search terms (e.g., 'albert einstein', 'theory of relativity') and a 'Web' tab.
- Main Search Results:** A central area showing search results for 'deer', including a line graph, a table of statistics (e.g., '9.08', '202,27 K'), and various links to related content like 'Deer Consumer Products Inc (US:DEER)' and 'White-tailed deer - Wikipedia, the free encyclopedia'.
- Clustering Panel:** A right-side panel titled 'clusters' showing a hierarchical list of categories such as 'All Results (259)', 'Deer Hunting (34)', 'Mule, Hunts (30)', 'Family (19)', 'Wildlife (18)', 'Fish and Wildlife (4)', 'White-tailed (3)', 'Other Topics (1)', 'Feeding, Regulations (2)', 'Department Of Wildlife (2)', 'Tennessee Wildlife Resources Agency (2)', 'Canada, UK Columbia (2)', 'Odocoileus virginianus (2)', 'Other Topics (4)', 'Management (23)', 'Park (20)', 'Animals (16)', 'Red Deer (13)', 'Control (10)', and 'Odocoileus virginianus (9)'. It includes a 'remix' button and a 'find in clusters' search box.

Data Mining

MapReduce: Simplified Data Processing on Large Clusters

Jeffrey Dean and Sanjay Ghemawat

jeff@google.com, sanjay@google.com

Google, Inc.

Abstract

MapReduce is a programming model and an associated implementation for processing and generating large data sets. Users specify a *map* function that processes a key/value pair to generate a set of intermediate key/value pairs, and a *reduce* function that merges all intermediate values associated with the same intermediate key. Many real world tasks are expressible in this model, as shown in the paper.

Programs written in this functional style are automatically parallelized and executed on a large cluster of commodity machines. The run-time system takes care of the details of partitioning the input data, scheduling the program's execution across a set of machines, handling machine failures, and managing the required inter-machine communication. This allows programmers without any experience with parallel and distributed systems to easily utilize the resources of a large distributed system.

Our implementation of MapReduce runs on a large cluster of commodity machines and is highly scalable: a typical MapReduce computation processes many terabytes of data on thousands of machines. Programmers find the system easy to use: hundreds of MapReduce programs have been implemented and upwards of one thousand MapReduce jobs are executed on Google's clusters every day.

1 Introduction

Over the past five years, the authors and many others at Google have implemented hundreds of special-purpose computations that process large amounts of raw data, such as crawled documents, web request logs, etc., to compute various kinds of derived data, such as inverted indices, various representations of the graph structure of web documents, summaries of the number of pages crawled per host, the set of most frequent queries in a

given day, etc. Most such computations are conceptually straightforward. However, the input data is usually large and the computations have to be distributed across hundreds or thousands of machines in order to finish in a reasonable amount of time. The issues of how to parallelize the computation, distribute the data, and handle failures conspire to obscure the original intent of the computation with large amounts of complexity in these issues.

As a reaction to this complexity, we developed an abstraction that allows us to express these computations in a way that allows us to parallelize them. The details of parallelization, scheduling, and load balancing are handled by the system, inspired by the *map* and *reduce* functions. The abstraction most of our computation is expressed in terms of a *map* operation to each logical record and a *reduce* operation to compute a set of intermediate values for each key. Applying a *reduce* operation to all records with the same key, in order to produce a single result, is done appropriately. Our use of the *map* and *reduce* functions is specified map and reduce functions. The *map* function is the primary mechanism for expressing computation as the primary mechanism for expressing computation.

The major contribution of this paper is a simple and powerful interface that allows programmers to express and distribution of their computations in a way that allows them to utilize the resources of a large distributed system with an implementation that is highly scalable and has high performance.

Section 2 describes the MapReduce programming model and gives several examples of computations that can be expressed in the MapReduce programming model. Section 3 describes our cluster-based computing environment. Section 4 describes several refinements of the programming model that we have found useful. Section 5 has performance measurements of our implementation for a variety of tasks. Section 6 explores the use of MapReduce within Google including our experiences in using it as the basis

SCOPE: Easy and Efficient Parallel Processing of Massive Data Sets

Ronnie Chaiken, Bob Jenkins, Per-Ake Larson, Bill Ramsey,

Darren Shakib, Simon Weaver, Jingren Zhou

Microsoft Corporation

{rchaiken, bobjen, palarson, brams, darrens, sweaver, jrzhou}@microsoft.com

ABSTRACT

Companies providing cloud-scale services have an increasing need to store and analyze massive data sets such as search logs, e-mails, etc. For cost and performance reasons, processing is done on large clusters of shared-nothing commodity machines. It is challenging to design a programming model that enables users to easily write programs that can efficiently and effectively utilize all resources in such a cluster and achieve maximum degree of parallelism.

The *Map-Reduce* programming model provides a good abstraction of group-by-aggregation operations over a cluster of machines. The programmer provides a *map* function that performs grouping and a *reduce* function that performs aggregation. The underlying run-time system achieves parallelism by partitioning the data and processing different partitions concurrently using multiple machines. However, this model has its own set of limitations: Users are forced to map their applications to the map-reduce model in order to achieve parallelism. For some applications this mapping is very unnatural. Users have to provide implementations for the *map* and *reduce* functions, even for simple operations like projection and selection. Such custom code is error-prone and hardly reusable. Moreover, for complex applications that require multiple stages of map-reduce, there are often many valid evaluation strategies and execution orders. Having users implement (potentially multiple) *map* and *reduce* functions is equivalent to asking users to specify physical execution plans directly in database systems. The user plans may be suboptimal and lead to performance degradation by orders of magnitude.

In this paper, we present a new scripting language, SCOPE (Structured Computations Optimized for Parallel Execution), targeted for large-scale data analysis that is under development at Microsoft. Many users are familiar with relational data and SQL. SCOPE intentionally builds on this knowledge but with simplifications suited for the new execution environment. Users familiar with SQL require little or no training to use SCOPE. Like SQL, data is modeled as sets of rows composed of typed columns. Every rowset has a well-defined schema. The SCOPE runtime provides implementations of many standard physical operators, saving users from implementing similar functionality repeatedly. SCOPE is being used daily for a variety of data analysis and data mining applications inside Microsoft.

SCOPE is a declarative language. It allows users to focus on the data transformation: required to solve the problem at hand and hides the complexity of the underlying platform and implementation details. The SCOPE compiler and optimizer are responsible for generating an efficient execution plan and the runtime for executing the plan with minimal overhead.

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the VLDB copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Very Large Database Endowment. To copy otherwise, or to republish, to post on servers or to redistribute to lists, requires a fee. VLDB '06, August 24-30, 2006, Auckland, New Zealand. Copyright 2006 VLDB Endowment. ACM 959-4-0000-000-0000-0.

To appear in OSDI 2004



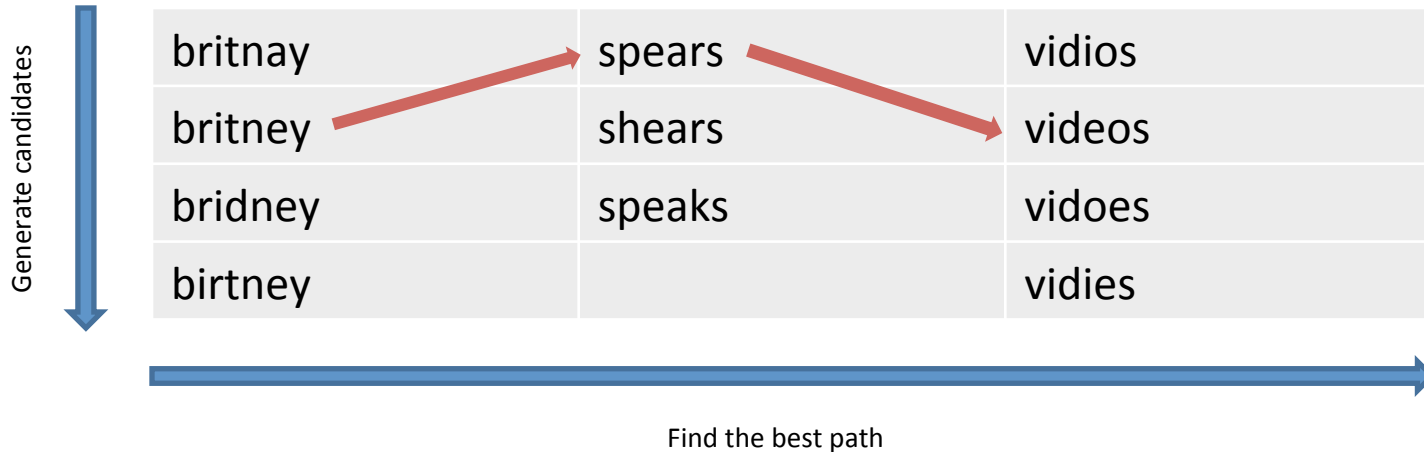
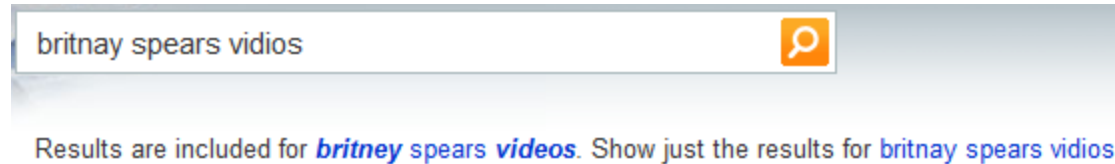
Applications

Spellchecking

488941	britney spears	29	britent spears	9	brinttany spears	5	brney spears	3	britiy spears	2	brirreny spears
40134	brittany spears	29	brittnany spears	9	britanay spears	5	broitney spears	3	britmeny spears	2	brittany spears
36315	brittney spears	29	britttany spears	9	britaniny spears	5	brotny spears	3	britneey spears	2	brirttany spears
24342	britany spears	29	btiney spears	9	britn spears	5	bruteny spears	3	britnehy spears	2	brirttney spears
7331	britny spears	26	birttney spears	9	britne spears	5	brtiyney spears	3	britnely spears	2	britain spears
6633	briteny spears	26	breitney spears	9	britneyn spears	5	brirttney spears	3	britnesy spears	2	britane spears
2696	britteny spears	26	brinity spears	9	britrney spears	5	gritney spears	3	britnetty spears	2	britaneny spears
1807	briney spears	26	britenay spears	9	brtiny spears	5	spritney spears	3	britnex spears	2	britania spears
1635	brittny spears	26	britneyt spears	9	britttney spears	4	bittny spears	3	britneyxxx spears	2	britann spears
1479	brintey spears	26	brittan spears	9	brtny spears	4	bnritney spears	3	britnity spears	2	britanna spears
1479	britanny spears	26	brittne spears	9	brytny spears	4	brandy spears	3	britntey spears	2	britannie spears
1338	britiny spears	26	brittany spears	9	rbittny spears	4	brbrittney spears	3	britnyey spears	2	britannt spears
1211	britnet spears	24	beitney spears	8	birtiny spears	4	breatiny spears	3	britterny spears	2	britannu spears
1096	britiney spears	24	birteny spears	8	bithney spears	4	breetney spears	3	brittneey spears	2	britanyl spears
991	britaney spears	24	brightney spears	8	brattney spears	4	breitney spears	3	britttney spears	2	britanyt spears
991	britnay spears	24	brintiny spears	8	breitny spears	4	brfitney spears	3	brittnyey spears	2	briteeny spears
811	brithney spears	24	britaney spears	8	breteny spears	4	briattany spears	3	brityen spears	2	britenay spears
811	briney spears	24	brittenny spears	8	brightny spears	4	bricity spears	3	briytney spears	2	britenet spears
664	birtney spears	24	brיתי spears	8	brintay spears	4	bricity spears	3	briltney spears	2	briteniy spears
664	brintny spears	24	brintwy spears	8	brinttey spears	4	brility spears	3	broteny spears	2	britenys spears
664	briteney spears	24	brittni spears	8	briotney spears	4	britttany spears	3	brtaney spears	2	britaney spears
601	bitney spears	24	brittnie spears	8	britanys spears	4	brinie spears	3	brtiiany spears	2	britin spears
601	brinty spears	21	birtitney spears	8	britley spears	4	brintenen spears	3	brtinay spears	2	britoryary spears
544	brittaney spears	21	birtany spears	8	brityneyb spears	4	brintne spears	3	brtinney spears	2	britymy spears
544	brittany spears	21	biteny spears	8	brityny spears	4	brityby spears	3	brtitany spears	2	brittaney spears
364	britey spears	21	bratney spears	8	brityty spears	4	brityay spears	3	brtiteny spears	2	brittnat spears
364	brittiny spears	21	brítani spears	8	brittner spears	4	brityaney spears	3	brtnet spears	2	brittnhey spears
329	brtney spears	21	britanie spears	8	brottany spears	4	britynie spears	3	brytyny spears	2	brintdy spears
269	brtney spears	21	briteany spears	7	brirtney spears	4	britynney spears	3	btney spears	2	brinteh spears
269	brintneys spears	21	brittay spears	7	birtney spears	4	britymney spears	3	drittney spears	2	brinteneny spears
244	britle spears	21	brittinay spears	7	biteney spears	4	britynar spears	3	pretney spears	2	brintney6 spears
244	brytney spears	21	brtany spears	7	bitiny spears	4	britlet spears	3	zbrittney spears	2	brintneye spears
220	breatney spears	21	brtiany spears	7	breateny spears	4	brityneuy spears	2	barittany spears	2	brintneyh spears
220	britiany spears	19	briney spears	7	brianty spears	4	brityney spears	2	bbbrittney spears	2	brintneym spears
199	brittney spears	19	brirtney spears	7	brintye spears	4	britymney spears	2	bbittney spears	2	brintneyyy spears
163	britrny spears	19	britrnaey spears	7	britianny spears	4	brittaby spears	2	bbritny spears	2	brintney spears
147	breatny spears	19	britlee spears	7	britley spears	4	brittery spears	2	bbrittany spears	2	brintneyj spears
147	brittiney spears	19	brityny spears	7	britlej spears	4	britthey spears	2	beitany spears	2	brintne spears
147	britty spears	19	brittanty spears	7	britleyuy spears	4	brittneay spears	2	beitny spears	2	brintnu spears
147	broutney spears	19	britttney spears	7	britletney spears	4	brittnat spears	2	beritney spears	2	brintony spears
147	brutney spears	17	birtny spears	7	brittnny spears	4	britttney spears	2	beritny spears	2	britrany spears
133	britteneny spears	17	brieny spears	7	britttian spears	4	brittney spears	2	betney spears	2	britrereny spears
133	briyney spears	17	brintty spears	7	brilyny spears	4	britttney spears	2	betny spears	2	brity spears
121	bittany spears	17	brithy spears	7	brrittany spears	4	brutney spears	2	bhriney spears	2	brittany spears
121	bridney spears	17	brittania spears	7	brityney spears	4	briyeny spears	2	biney spears	2	brittanyy spears
121	britalny spears	15	brinny spears	7	btitney spears	4	brinity spears	2	binety spears	2	brittang spears
121	britley spears	15	briten spears	7	britttany spears	4	brutteny spears	2	biretny spears	2	brittans spears
109	brietney spears	15	briterney spears	6	brityny spears	4	bruttiany spears	2	briritany spears	2	brittanyh spears
109	brithny spears	15	britheny spears	6	bhrittney spears	4	bryney spears	2	bririttany spears	2	brittanyyn spears
109	britrni spears	15	brithney spears	6	brithney spears	4	brythney spears	2	brirttany spears	2	brittany's spears
109	brittant spears	15	brittamy spears	6	breathney spears	4	brytne spears	2	birnty spears	2	brittanyt spears
98	brittney spears	15	brittmey spears	6	breaty spears	4	brytni spears	2	birtey spears	2	brittanyyy spears
98	brithy spears	15	britlet spears	6	bratny spears	4	brytnie spears	2	birtheny spears	2	brittary spears
98	brittiany spears	15	britrney spears	6	briatany spears	4	brvritney spears	2	birthieny spears	2	brittenie spears
98	britley spears	15	brintey spears	6	brint spears	4	dritley spears	2	birtnay spears	2	brittenty spears
89	brietny spears	14	brinet spears	6	britteney spears	4	priteny spears	2	birtnet spears	2	brittitney spears
89	brinetny spears	14	britley spears	6	britan spears	3	beittany spears	2	bitnet spears	2	brittley spears
89	brintny spears	14	britten spears	6	britynty spears	3	bichney spears	2	bittany spears	2	brittn spears
89	britrnie spears	12	beritney spears	6	brittitney spears	3	birityny spears	2	bnrittany spears	2	britttney spears
89	britley spears	12	brityny spears	6	brityny spears	3	birntey spears	2	bnrtney spears	2	britttney spears

<http://www.google.com/jobs/briney.html>

Spellchecking



1. Generate a set of candidates per query term using approximate matching techniques. Score each candidate according to, e.g., "distance" from the query term and usage frequency.
2. Find the best path in the lattice using the Viterbi algorithm. Use, e.g., candidate scores and bigram statistics to guide the search.

Entity Extraction

Levels of abstraction ↑

...
MAN				FOOD
N/proper	V/past/eat	DET	ADJ	N/singular
Richard	ate	some	bad	curry

1. Logically annotate the text with zero or more computed layers of meta data. The original surface form of the text can be viewed as trivial meta data.
2. Apply a pattern matcher or grammar over selected layers. Use, e.g., handcrafted rules or machine-trained models. Extract the surface forms that correspond to the matching patterns.

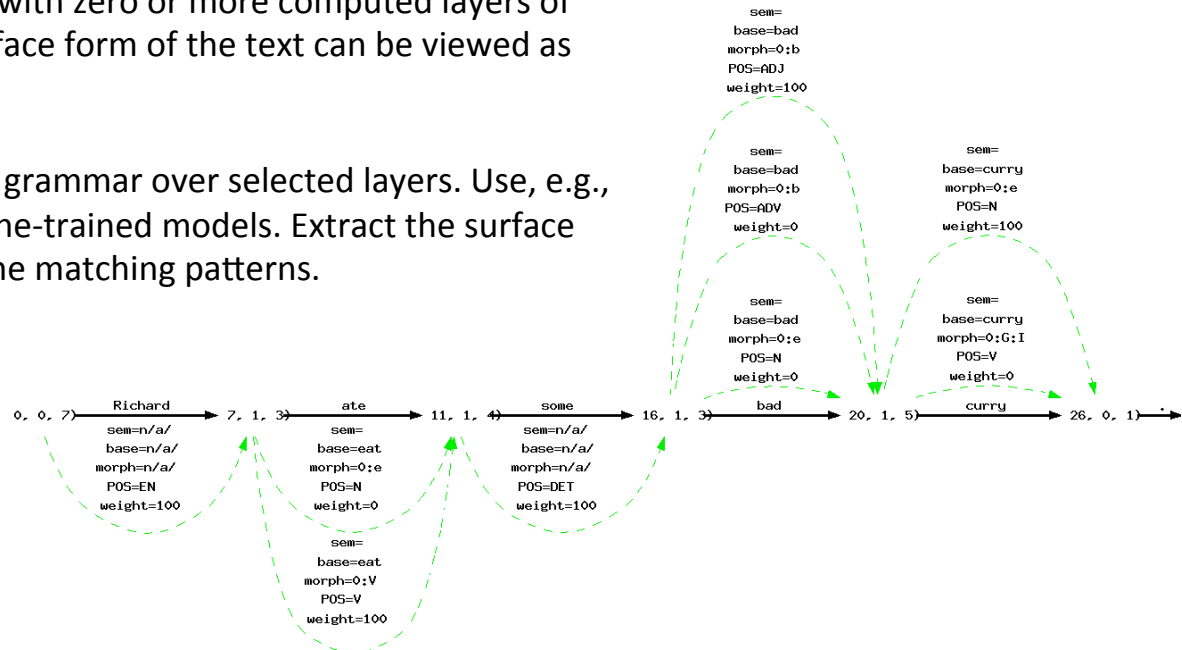
People

- Roger Federer (58)
- Andy Roddick (51)
- Lindsay Davenport (5)
- Andre Agassi (48)
- Maria Sharapova (45)
- Serena Williams (45)
- Alicia Molik (36)
- Marat Safin (34)
- Nikolay Davydenko (2)
- Joachim Johansson (2)
- Svetlana Kuznetsova

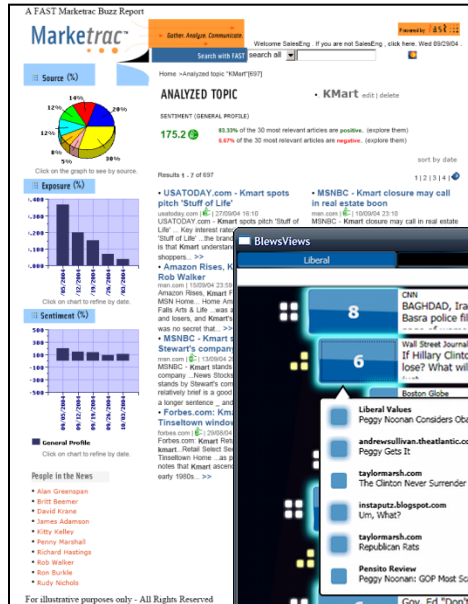
Refine your search

- albert einstein
- theory of relativity
- general theory of relativity
- bose einstein condensation
- physicists
- photoelectric effect
- special theory of relativity
- condensation
- speed of light
- bose-einstein condensation

[more >](#)



Sentiment Analysis



Jurys Boston Hotel
350 Stuart Street, Boston, 02116, United States
From £126 - £190*
Average visitor rating: 8.0 from 34 reviews
Hotel info, Reader reviews, Photos, Prices & availability
We have searched the web and found 34 reviews for this hotel
brand new and trying to be the best from A TripAdvisor Member, Chicago, IL, TripAdvisor.com
We spent 3 nights at Jurys' s beginning on the 4th of July, they had just opened on the 2nd....
Beautiful New Jurys Hotel from A TripAdvisor Member, Portsmouth, RI, TripAdvisor.com
Very impressed with this new hotel. Loved the decor - rooms were spacious and very clean. One of...
It would be a crime to stay anywhere else! from TripAdvisor Member, New York, NY, TripAdvisor.com
My company suggested I stay at Jurys on a red Boston. What a surprise. For an Irish...
Could not have been any better... from TripAdvisor Member, Montreal, Quebec, TripAdvisor.com
I went down to Boston from the 19th to the 25th and stayed at the wonderful...
Verdict is in from A TripAdvisor Member, TripAdvisor.com
I visited Boston July 31 to August 3 and had a great stay at Jurys. The room was beautifully...

“What is the current perception of my brand?”

“I want to stay at a hotel whose user reviews have a definite positive tone.”

“What are the most emotionally charged issues in American politics right now?”



1. To construct a sentiment vocabulary, start by defining a small seed set of known polar opposites.
2. Expand the vocabulary by, e.g., looking at the context around the seeds in a training corpus.
3. Use the expanded vocabulary to build a classifier. Apply special heuristics to take care of, e.g., negations and irony.

<http://research.microsoft.com/en-us/projects/blews/>

Contextual Search



“Sentences where someone says something positive about Adidas.”

`xml:sentence:(“adidas” and sentiment:@degree:>0)`

“Dates and locations related to D-Day.”

`xml:sentence:(“d-day” and (scope(date) or scope(location)))`

“Paragraphs that discuss a company merger or acquisition.”

`xml:paragraph:(string(“merger”, linguistics=“on”) and scope(company) and scope(price))`

“Paragraphs that contain quotations by Alan Greenspan, where he mentions a monetary amount.”

`xml:paragraph:quotation:(@speaker:“greenspan” and scope(price))`

“Sentences where the acronym MIT is defined.”

`xml:sentence:acronym:(@base:“mit” and scope(@definition))`

Persons that appear in **documents** that contain the word {soccer}



person@base
Jack Nicklaus (~10.0%)
Fred Davis (~10.0%)
Billie Jean King (~8.0%)
Richard Nixon (~8.0%)
John Wayne (~7.0%)
Margaret Smith (~7.0%)
Joe Frazier (~7.0%)
Irina Rodnina (~7.0%)
Mao Zedong (~6.0%)
Gordie Howe (~6.0%)
Richard M. Nixon (~6.0%)

[More...](#)

Example from Wikipedia

person@base
Diego Maradona (~4.0%)
David Beckham (~4.0%)
Alan Shearer (~3.0%)
Michelle Akers (~3.0%)
Mia Hamm (~3.0%)
Eric Wynalda (~3.0%)
Freddy Adu (~3.0%)
Michel Platini (~2.0%)
Stanley Matthews (~2.0%)
Oliver Neuville (~2.0%)
Bobby Moore (~2.0%)

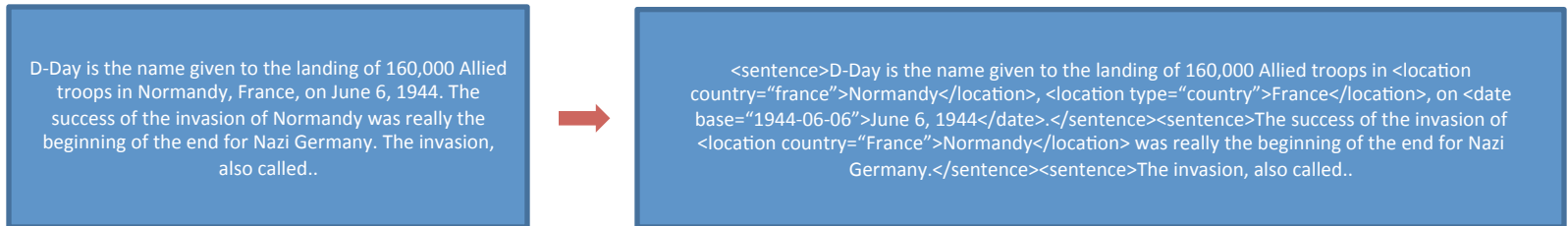
[More...](#)

Persons that appear in **paragraphs** that contain the word {soccer}

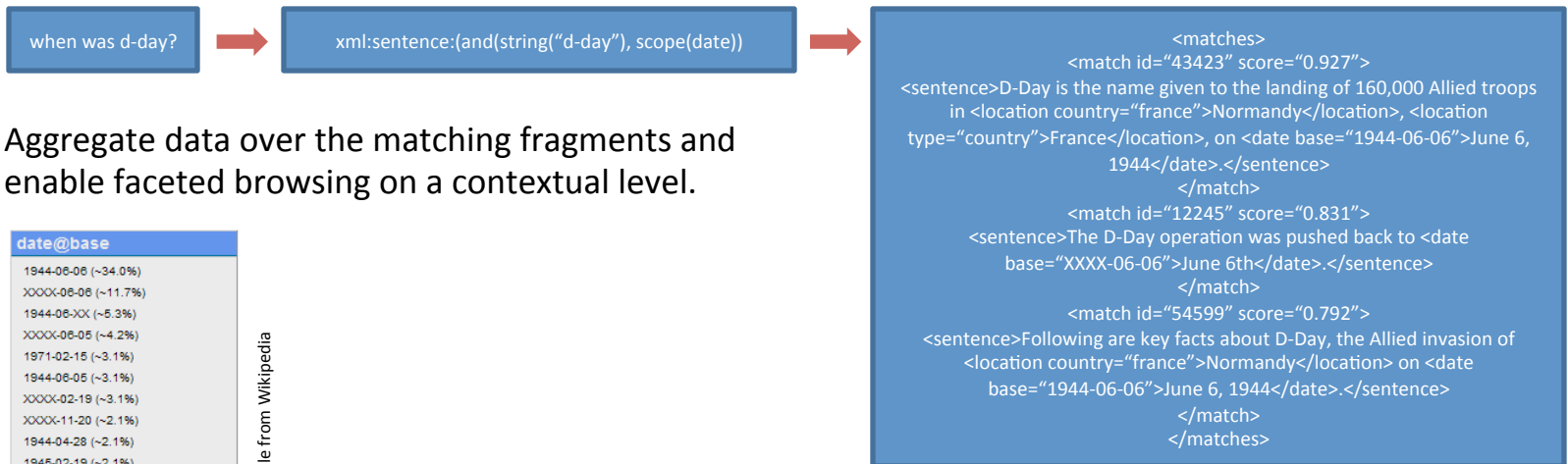


Contextual Search

1. During content processing, identify structural and semantic regions of interest. Mark them up in context, possibly decorated with meta data.



2. Make all the marked-up data fully searchable in a way that preserves context and where retrieval can be constrained on both structure and content. Possibly translate natural language queries into suitable system queries.



3. Aggregate data over the matching fragments and enable faceted browsing on a contextual level.

date@base
1944-06-06 (~34.0%)
XXXX-06-06 (~11.7%)
1944-06-XX (~6.3%)
XXXX-06-05 (~4.2%)
1971-02-15 (~3.1%)
1944-06-05 (~3.1%)
XXXX-02-19 (~3.1%)
XXXX-11-20 (~2.1%)
1944-04-28 (~2.1%)
1945-02-19 (~2.1%)
XXXX-09-15 (~2.1%)

More...

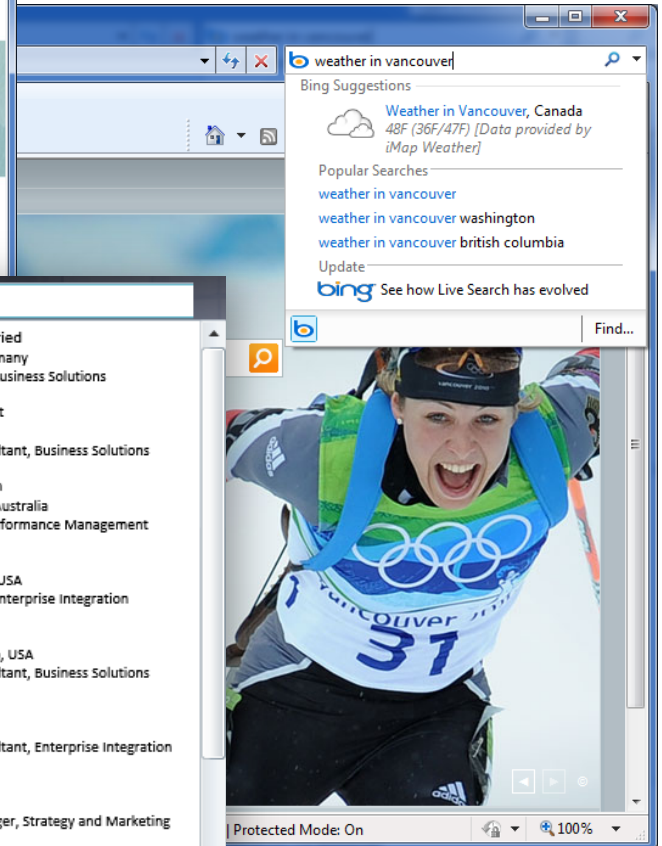
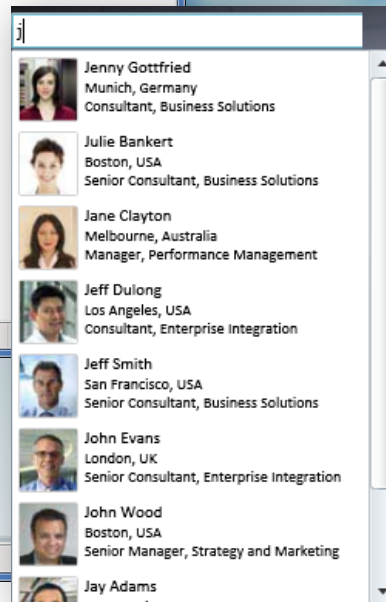
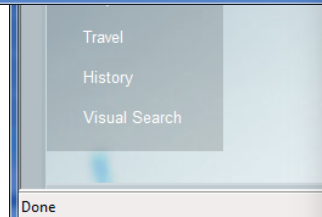
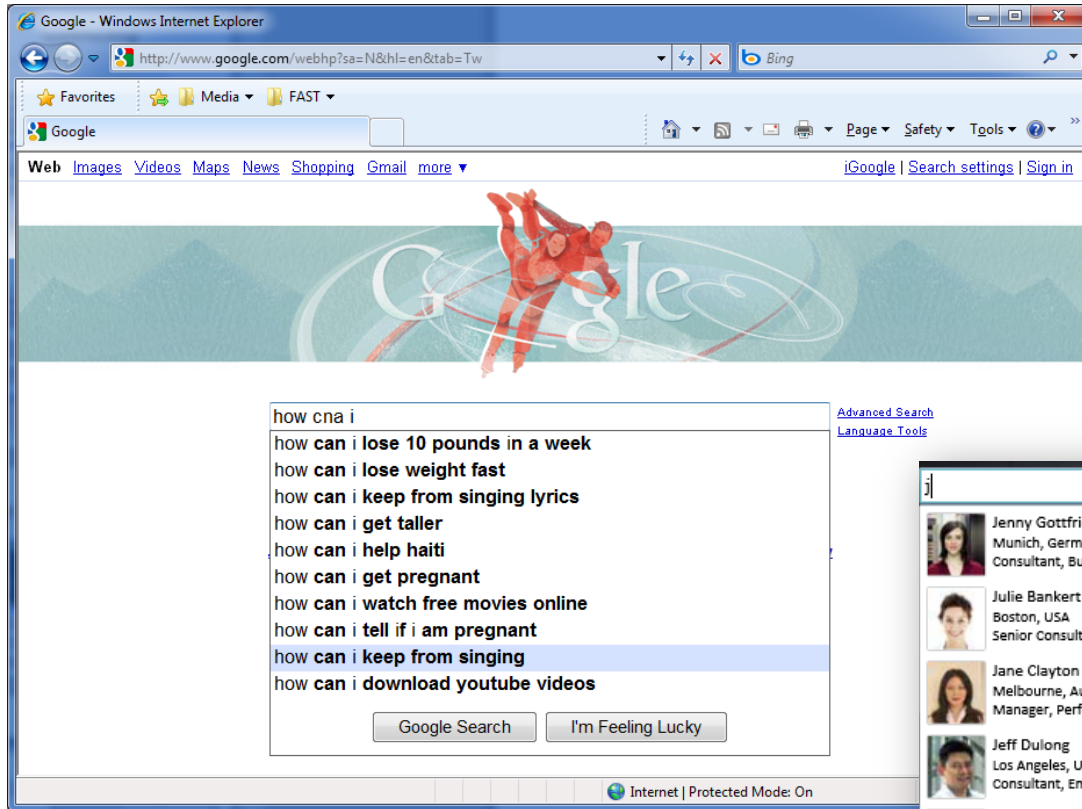
Example from Wikipedia

Machine Translation

The screenshot shows the Bing Translator interface. At the top, there are navigation links for Web, Images, News, More, MSN, and Hotmail, along with Sign in, Norway, and Preferences. The Bing logo is on the left, and the word "Translator" is in the center. Below the logo, there are links for Home, Tools, and Help. The main area features two language dropdown menus: "English" and "Danish", with a "Translate" button between them. Below the language selection, there is a text input field containing the English text: "The University of Oslo is Norway's largest and oldest institution of higher education. It was founded in 1811 when Norway was still under Danish rule. Today the University of Oslo has approx. 27 700 students and 5 900 employees. Four Nobel Prize winners indicates the quality of the research at the University." To the right of this text, the Danish translation is displayed: "Universitetet i Oslo er Norges største og ældste institution af de videregående uddannelser. Det blev grundlagt i 1811 hvor Norge var stadig under dansk regel. Universitetet i Oslo har i dag ca. 27 700 studerende og 5 900 medarbejdere. Fire Nobelprisen vindere angiver kvaliteten af forskningen på universitetet." At the bottom left, there is a "New" notice about a beta Haitian Creole translation engine.

The screenshot shows the Google Translate interface. At the top, there are navigation links for Web, Images, Videos, Maps, News, Shopping, Gmail, and more, along with a Help link. The Google Translate logo is prominently displayed. Below the logo, there is a "Translation" section with links for Translated Search, Translator Toolkit, and Tools and Resources. The main area features a "Translate text, webpages and documents" section with a text input field containing the same English text as the Bing screenshot. Below the text input, there are two dropdown menus: "Translate from: English" and "Translate into: Danish", with a "Translate" button to the right. Below the translation controls, the Danish translation is displayed: "Universitetet i Oslo er Norges største og ældste institution for videregående uddannelse. Det blev grundlagt i 1811, da Norge stadig var under dansk styre. I dag Universitetet i Oslo har ca. 27 700 studerende og 5 900 ansatte. Fire nobelpristagere angiver kvaliteten af forskningen på universitetet." At the bottom, there is a link to "Contribute a better translation".

Query Completion



Caption Generation

- **Intra-document search**
 - Locate and rank relevant document fragments
 - But do it fast!
- **Perceived relevancy**
 - First impressions count
 - Can make or break a service
- **Trends towards richer captions**
 - Format-specific interactivity
 - Actionable elements

Web Images Videos Maps News Shopping Gmail more ▾

Google why should i avoid sans serif fonts? Search Advanced Search

Web Show options... Results 1 - 10 of about 123,000 for why should i avoid sans serif fonts?. (0.07 seconds)

[How to Select Fonts for Your Website](#)
However, **sans serif fonts** can also be viewed as cold and impersonal. ... You can **avoid** this by choosing a font common to both operating systems. ... Your site's overall design **should** help you decide which style is best for your site.
[www.pallasweb.com/fonts.html](#) - [Cached](#) - [Similar](#)

Fonts
Glyphs in **sans-serif fonts**, as the term is used in CSS, have stroke endings a UA applying these guidelines **should** nevertheless **avoid** creating font-size ...
[www.w3.org/TR/CSS2/fonts.html](#) - [Cached](#) - [Similar](#)

[Credit Card Processing Experts - 5 Extra Credit Card Processing ...](#)
<p style="margin-bottom: 0in"><font To **avoid** paying these extra fees, you **should** know the terms of your credit card ...
[www.creditcardprocessingexperts.com/5_extra_credit_card_processing_charges_merchants_can_avoid.html](#) - [Cached](#)

Sans Serif Fonts
Avoid setting long passages of text in a light-weight **sans serif font**. Apart from lacking colour, continuous blocks of light text are hard to read. ...
[www.slideshare.net/mcmrbt/sans-serif-fonts](#) - [Cached](#) - [Similar](#)

[Web Design: TCR - Design Demos: Text: Serif vs. San-Serif](#)
A **sans-serif font**, such as Arial, lacks these tails. ... they **should** be used at a size large enough to **avoid** the problems shown below. ...
[www.webdesignref.com/examples/textex.htm](#) - [Cached](#) - [Similar](#)

Web Images Videos Shopping News Maps More | MSN | Hotmail

bing polyteknisk forening adresse

ALL RESULTS 1-10 of 997 results Advanced

Referat fra Oslo Talent Forum
...Netverk, Martine Bjørnstad: +47 991 61 084 Adresse: Innovativt Netverk Rosenkrantz gate 7, 0159 OSLO Innovativt Netverk er en undergruppe av **Polyteknisk Forening**
[www.innovativt.netverk.no/referat_politalentforum_020905.htm](#) - [Cached page](#)

Polyteknisk Forening, Foreninger og forbund Øvrige, Oslo - Gule Sider...
[www.gulesider.no/gsi/companyDetails.c?g=&lnid=2025M05](#) - [Cached page](#)

Innmeldingsskjema / Medlem / Hovedsiden - Polyteknisk Forening
Arbeidsgivers navn: Arbeidsgivers adresse: ... © **Polyteknisk Forening**
[www.polyteknisk.no/layout/set/print/Medlem/inmeldingsskjema](#) - [Cached page](#)

HABITAT Norge
HABITAT Norge [http://www.habitat-norge.org](#) Adresse: **Polyteknisk Forening**, Rosenkrantzgt. 7, 0159 OSLO Telefon: 47 22 42 68 70 Email: polyteknisk@polyteknisk.no Oslo, september 2006 ...
[www.habitat-norge.org/component?option=com_docman&task=doc_download&id=22&Itemid=27](#) - [Cached page](#) - PDF file

MORE ON THIS PAGE
Telefon til leder i Innovativt Netverk, Martine Bjørnstad: +47 991 61 084
Innovativt Netverk Rosenkrantz gate 7, 0159 OSLO
Innovativt Netverk er en undergruppe av **Polyteknisk Forening**
Tema: Oslo Talent Forum- a debate on: Globalization of Economies and Human Capital - Challenges ...
[Go to the page](#)

Office Business Applications: Unlocking the Business Value of IT Add to Wiki June 13, 2006

Summary Author People Comments Tags
...Business Value of IT Gurprit Singh Director, Emerging Technologies Microsoft Corporation
Sources of Business Performance Sources of Business Performance The Work of Business Create Lead Qualified? Retire Lead Create Oppt...

Sources of Business Performance The Work of Business "Real World" Information Work The Results Gap A New Breed of Application

Develop Customer Relationships Drive Innovation People Build Partner Relationships

The Future

