

INF3800/INF4800

Søketeknologi

2018.01.16

Foreleser



Aleksander Øhrn, Professor II
aleksaoh@ifi.uio.no

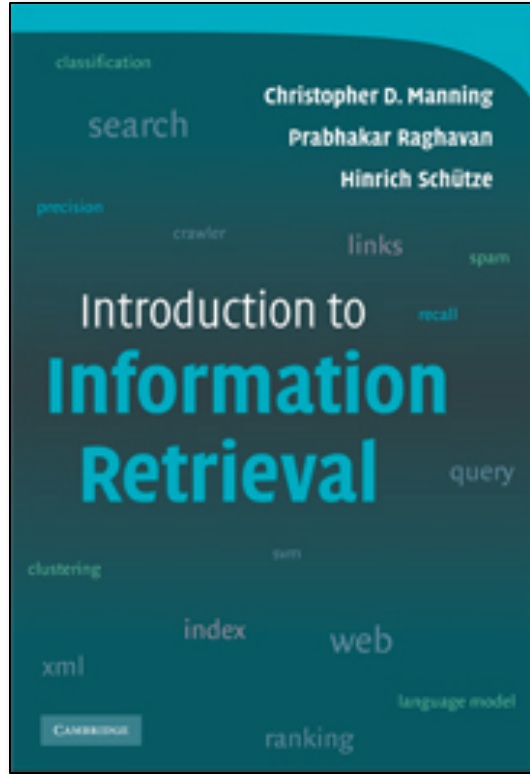
Gruppelærer



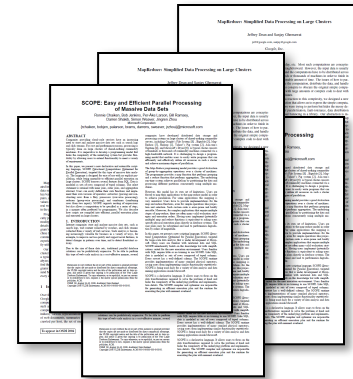
Ola Fosheim Grøstad
ola.fosheim.grostad@gmail.com

Pensum

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



+

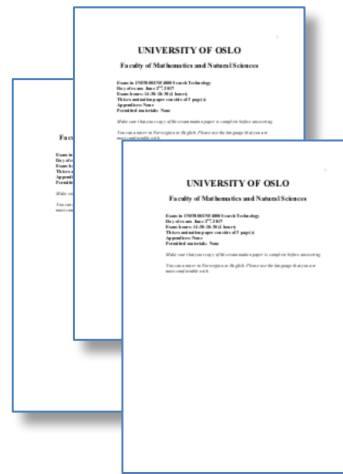


Øvinger

GitHub

<https://github.com/aohrn/inf3800-2018>

+



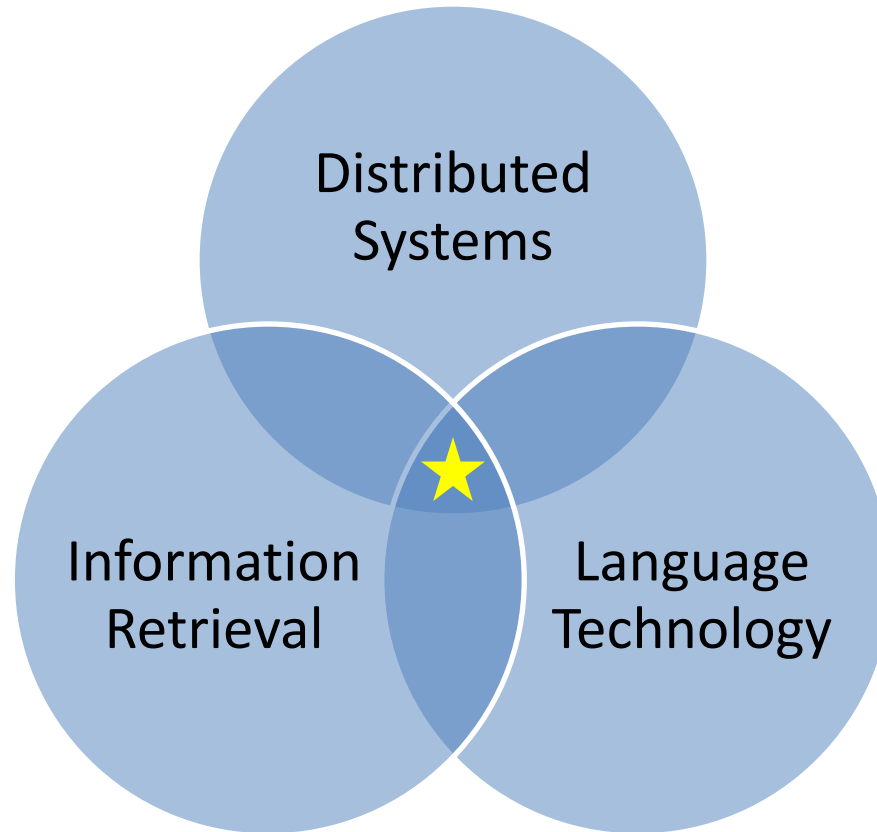
+

piazza

<https://piazza.com/uio.no/spring2018/inf38004800/home>

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+costello

elvis costello

Web **Images** **Videos** **Maps** **News** **Shopping** **More**

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_yjsf40geu4kz7FLyOQ4m5t3llyo4tpe/lost3f8f2-sb-top3f8fyfp-t-

lost

Web **Images** **Video** **Local** **Shopping** **More**

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: Pharma 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

Production | Cast and characters | Season synopsis | Matheson

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.

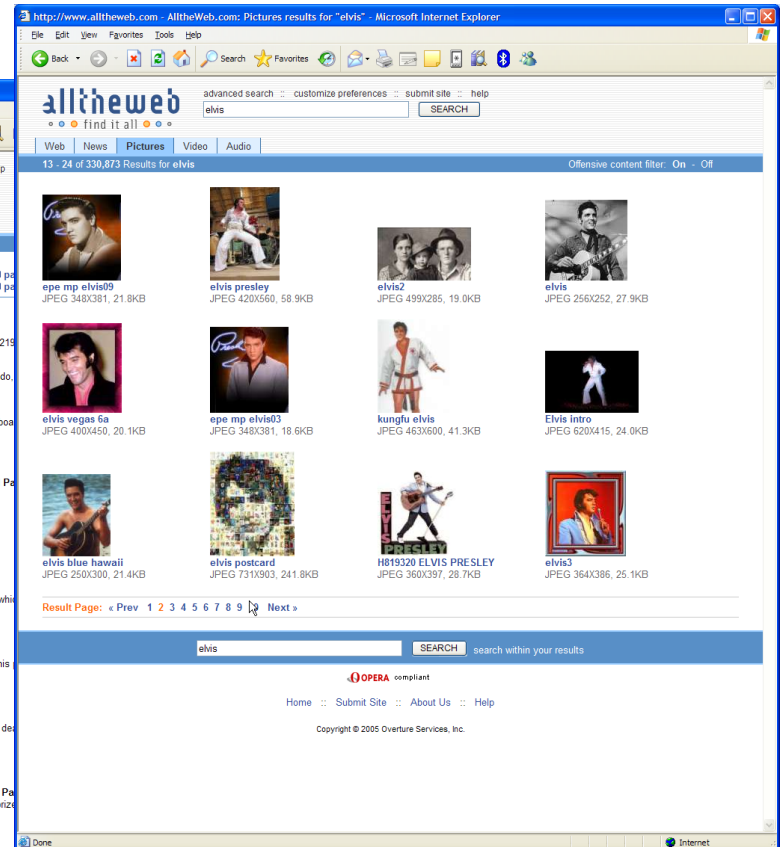
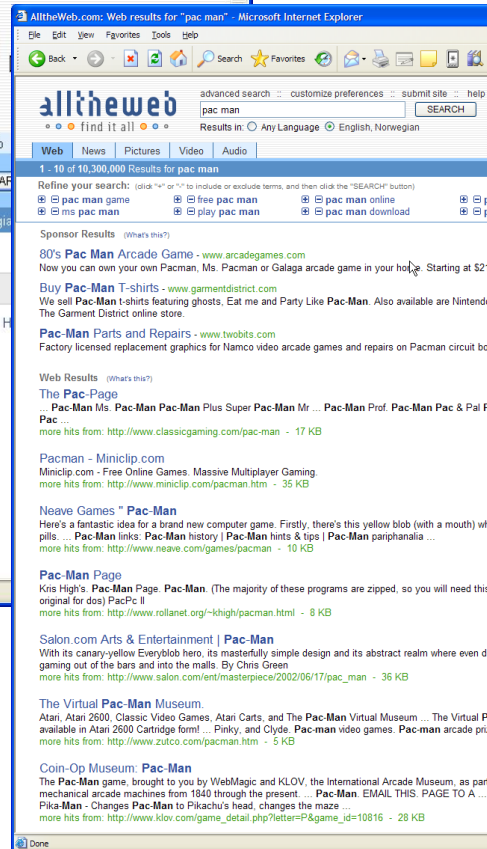
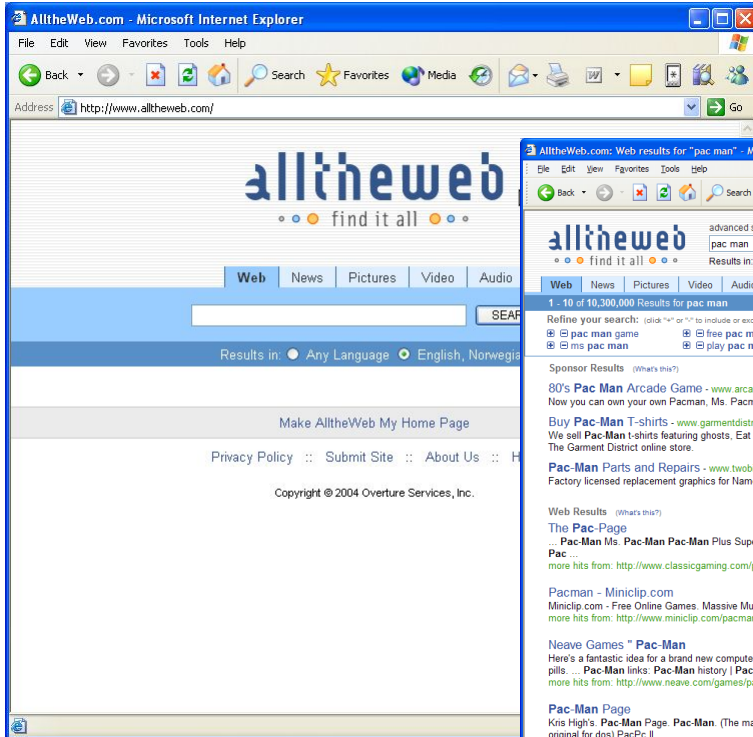
Lostpedia

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

FIRSTGOV.gov
The U.S. Government's Official Web Portal

Search: Enter Search Term(s) Go

in Federal Only Advanced

Home About Us Site Index Help Español Other Languages Welcome from President Bush

for Citizens for Businesses and Nonprofits for Federal Employees Government-to-Government

By Organization

- A-Z Agency Index
- Federal Executive
- Federal Legislative
- Federal Judicial
- Cross-Agency Portals
- State Government
- Local Government
- Tribal Government

Citizens: **Get It Done Online!**

- Find Government Benefits
- Apply for Government Grants
- Shop Government Auctions
- Apply for Government Jobs
- e-File Your Taxes
- Change Your Address
- Find Recreation Activities
- Renew Your Driver's License
- Get a Passport Application
- Apply for Social Security
- Check Immigration Case Status
- Contact Elected Officials
- Order Consumer Publications
- Weather Forecasts

Information by Topic

- Benefits and Grants
- Consumer Protection
- Defense and International
- Education and Jobs
- Environment, Energy and Agriculture

careerbuilder.com

Home Find Jobs Post Resumes My CareerBuilder Advice & Resources Career Fairs

WHO ARE THESE MONKEYS? >>>

Welcome new user... to the nation's largest employment network - find jobs locally or nationally!

- 1 Find Jobs: Search and apply instantly!
- 2 Set Up Job Alerts: Receive job matches via email
- 3 Post a Resume: List employers that you!

QUICK JOB SEARCH

Enter Keyword(s): (i.e. job title, company name)

Enter a City: City List

Select a State: All United States

Select a Category: All Job Categories

Search

JOB ALERTS

Find out when your favorite Saved Search has new results - via email.

You determine the frequency of emails: daily or weekly!

RESUMES

- Private or public
- Showcase multiple skills with up to 5 resumes
- Free and easy
- Hourly workers: Apply here

ADD A SEARCH JOB ALERT

POST A RESUME NOW

MOST POPULAR JOB CATEGORIES

- Accounting
- Contract/Lease
- Executive
- Information Technology
- Manufacturing
- Nonprofit
- Other
- Sales & Marketing

12.02.05 Ikke innlogget

Gule Sider® Telefonkatalogen™ Nummersøk Kart SMS Logg inn

Søk kun i firmanavn

Velg bransje | Velg sted | Detaljert søk

Kartsoek/Ruteplanlegger

Sammenlign og finn laveste pris

Canon Ixus 30 Sjekk priser

Apple iPod Photo 4GB Sjekk priser

Nettbutikker i Norge

Nyttige telefonnummere

Brann: 110

Politi: 112

Medisinsk nødhjelp: 113

Andre nyttige telefonnummere

SCIRUS
for scientific information only

About Us Newseum Advisory Board Submit Web Site Search Tips Contact Us

Basic Search Advanced Search Search Preferences

enstain big bang Search

Searched for: All of the words enstain AND big AND bang

Found: 17,025 total | 580 journal results | 16,445 Web results

Sort by: relevance | date

1. Big Bang: Yilmaz, Dec 2000

2. Beyond Einstein: Bang, BEYOND Einstein, from 12 Dec 2004

3. ENC Online: Curt Enstain, from 12 Dec 2004

IBM

Home Products Services & solutions Support & downloads My account

Seeing is believing

Big things are happening in the entry-level UNIX server world. The new IBM eServer p5 510 and 510 Express are causing quite a stir among small and medium-sized businesses. Learn more.

Learn about: Business consulting, IT services, Software products, Servers & storage

Shop for: Special offers, IT services, Software, Servers & storage, Notebooks

Get support: Downloads and drivers, Troubleshooting, Product publications, Training, Open a service request

Manage my account: Shopping cart, Orders and delivery, Invoices and payments, Contracts, Inventory

Resources for: Small & medium business, Large enterprise, Government, Education, IBM Business Partners, Journalists

楽天市場 ICHIBA

最新のモノ! 総合通販! 毎日! 楽天市場

楽天市場のサービス

楽天カード 使えば貯まる

楽天ポイント

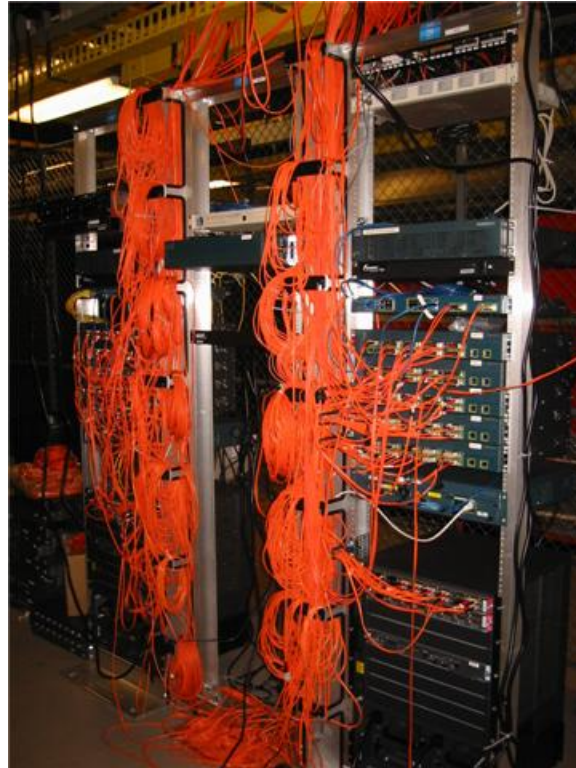
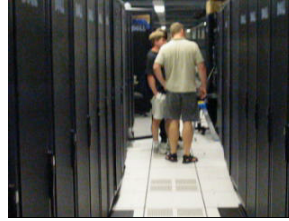
my Rakuten

Order your AIX upgrade online

THE HELP DESK IS OPEN

Data Centers

alltheweb.com 2000



Data Centers

Microsoft 2010



<http://www.youtube.com/watch?v=K3b5Ca6lZqE>

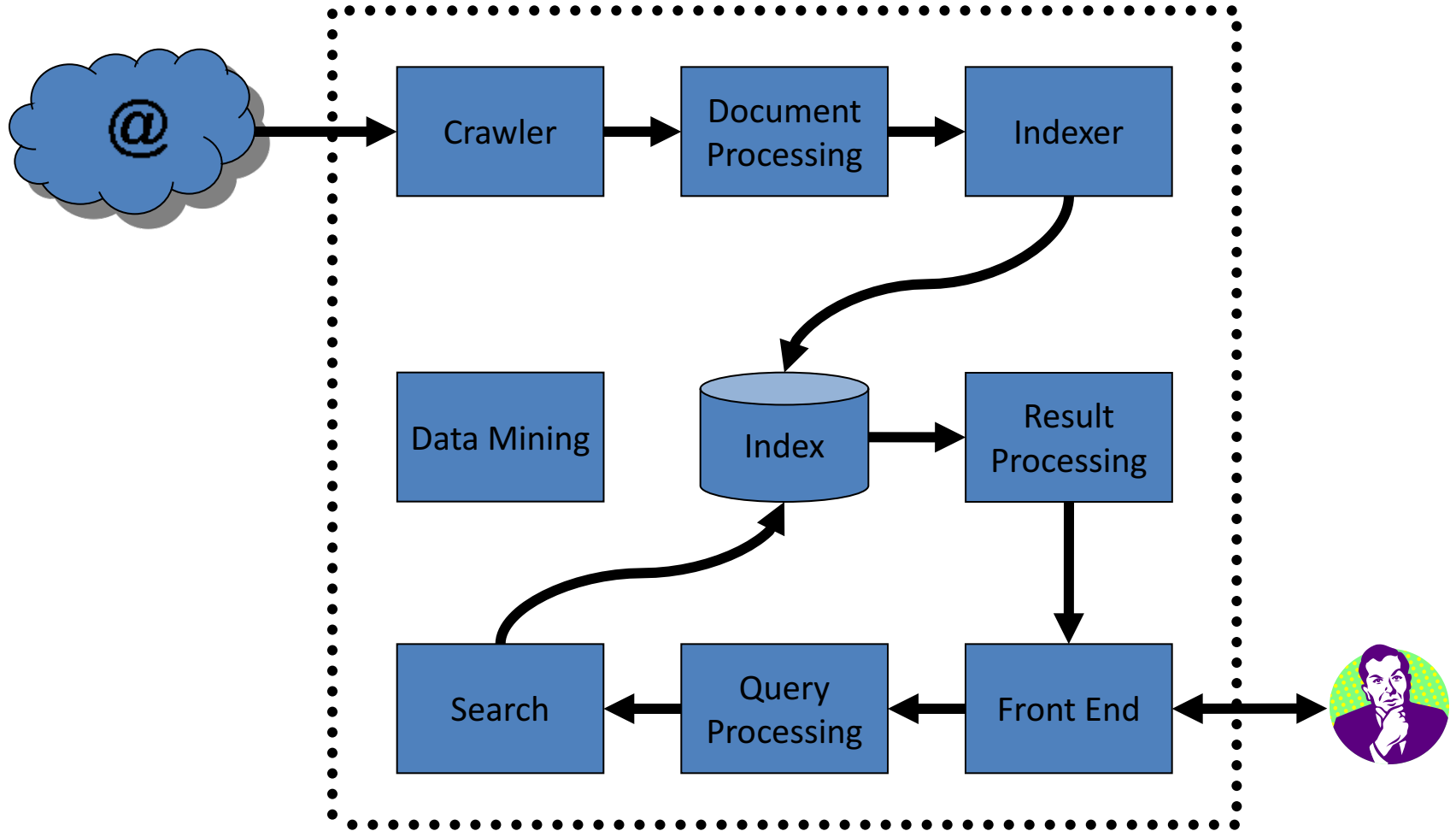
A screenshot of a YouTube search results page. The search bar contains the text "microsoft data centers". Below the search bar, there are navigation links for "Home", "Videos", and "Channels". The search results section shows two video thumbnails. The first video is titled "Microsoft OS Cloud Windows Azure Data Center - Google and Amazon" and has a duration of 2:53. The second video is titled "Microsoft Generation 4.0 Data Center Vision" and has a duration of 3:27. Both videos have a 5-star rating and show their respective view counts and upload dates.



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

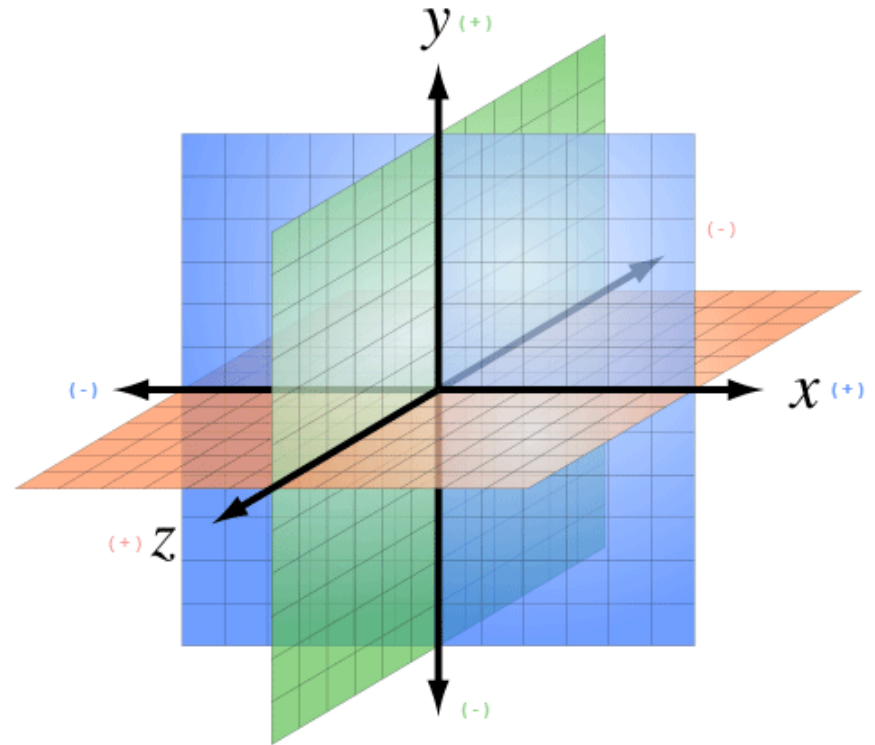
Search Platform Anatomy

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

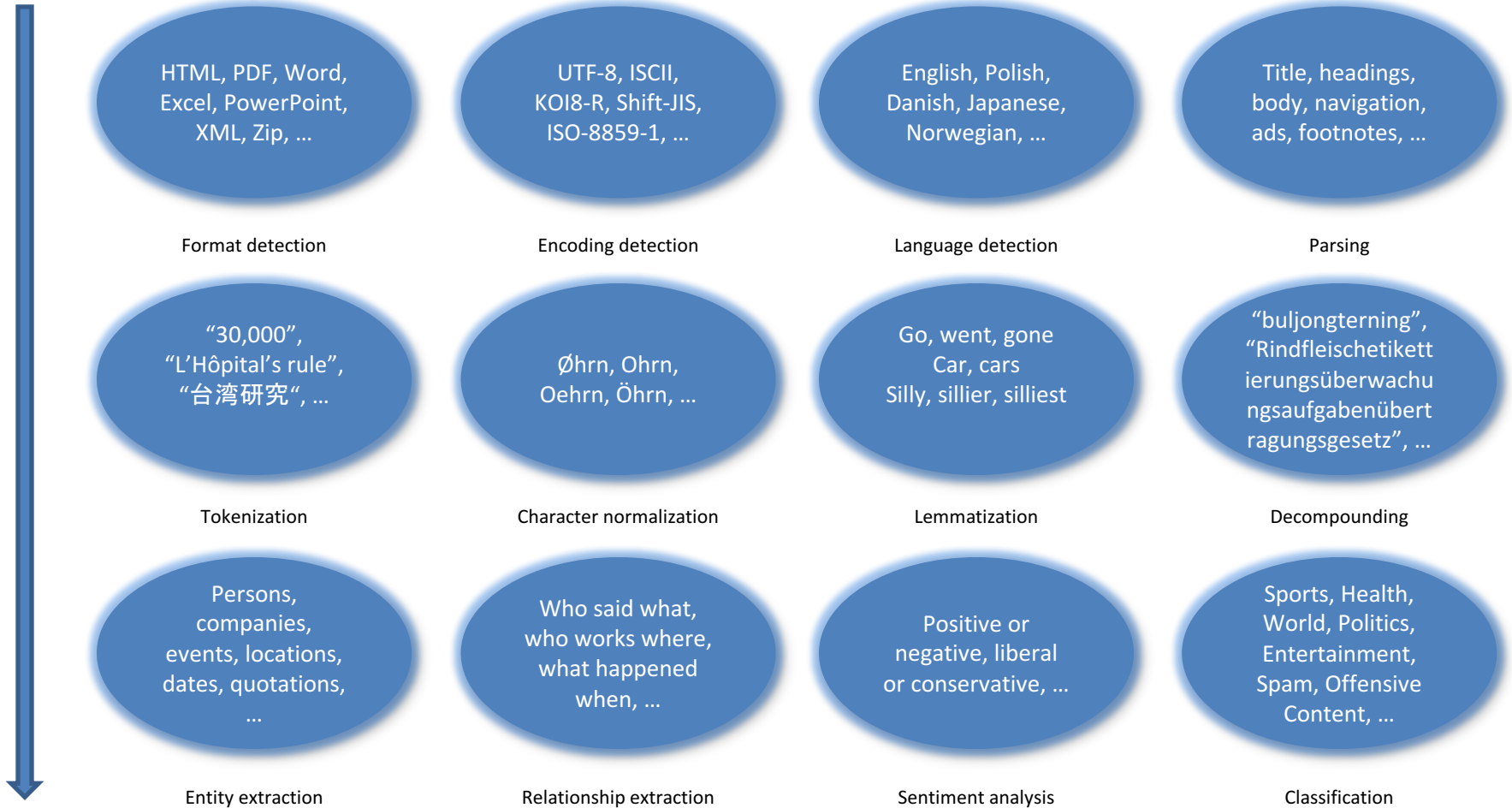
Scale through partitioning the data



Crawling The Web



Processing The Content

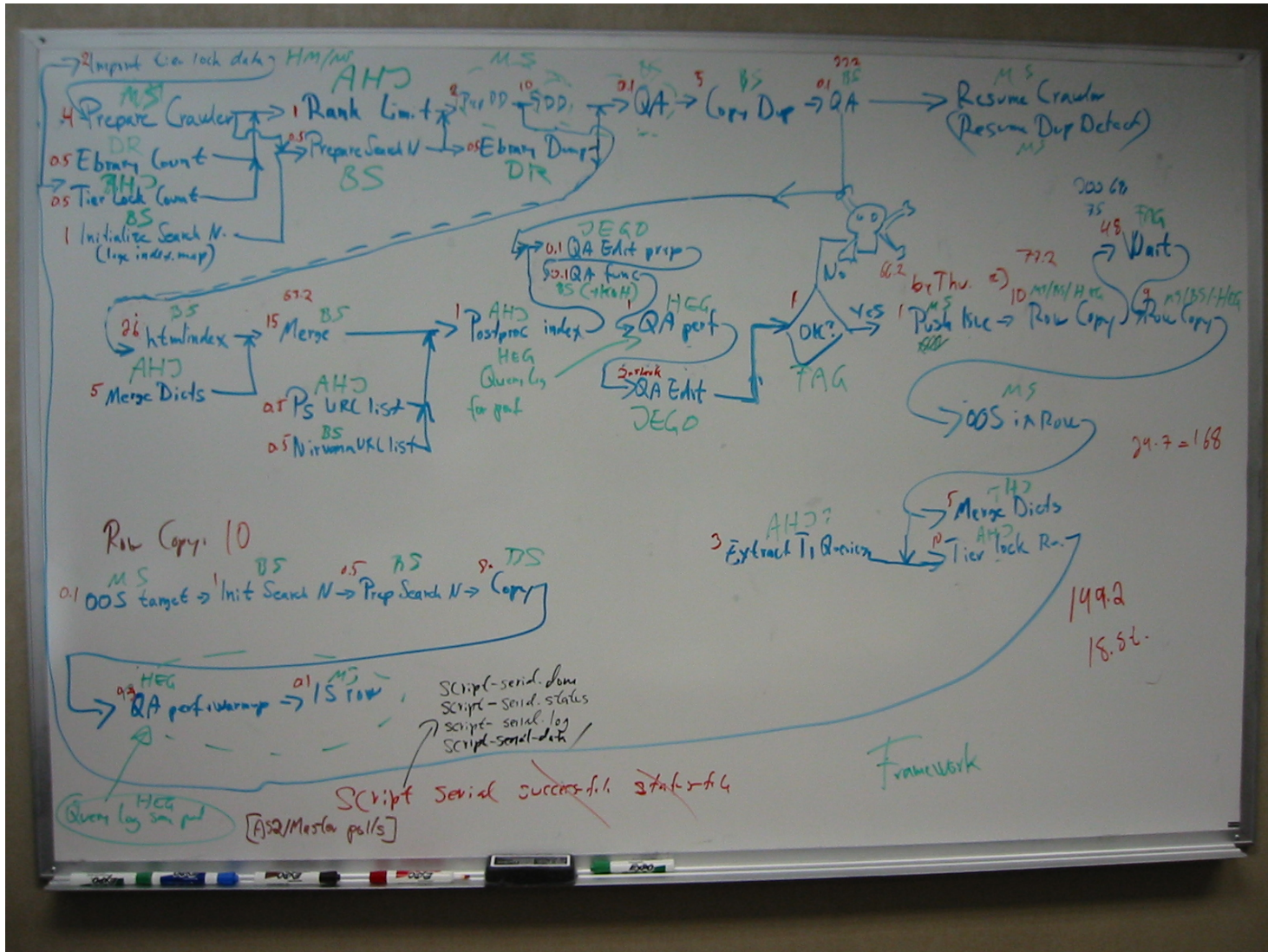


Creating The Index



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

Deploying The Index



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

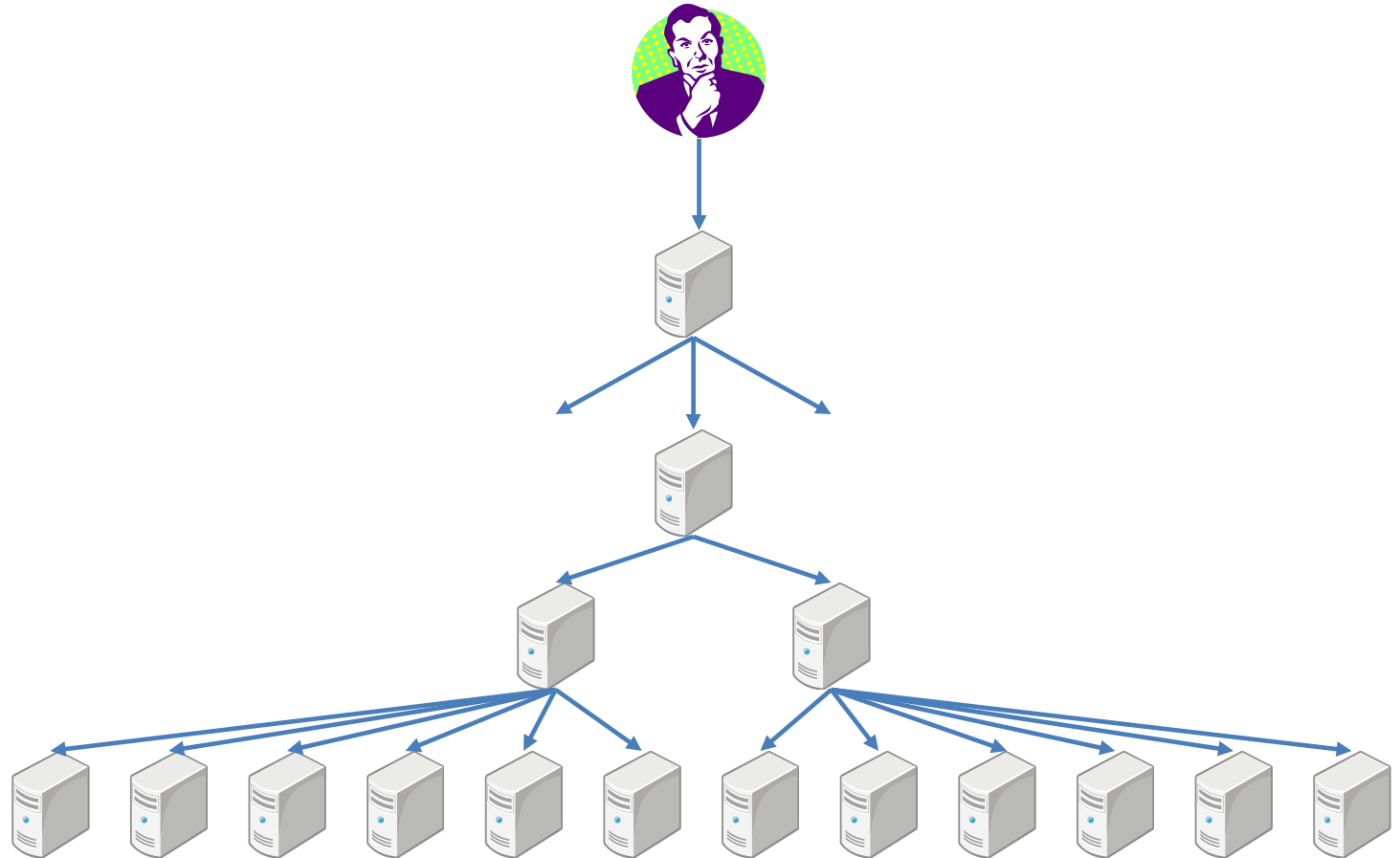
- Why?
- To skip postings that will not figure in the search results.
- How?
- Where do we place skip pointers?

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



Assess relevancy as we go along

Searching The Content



Federation
Query processing
Result processing

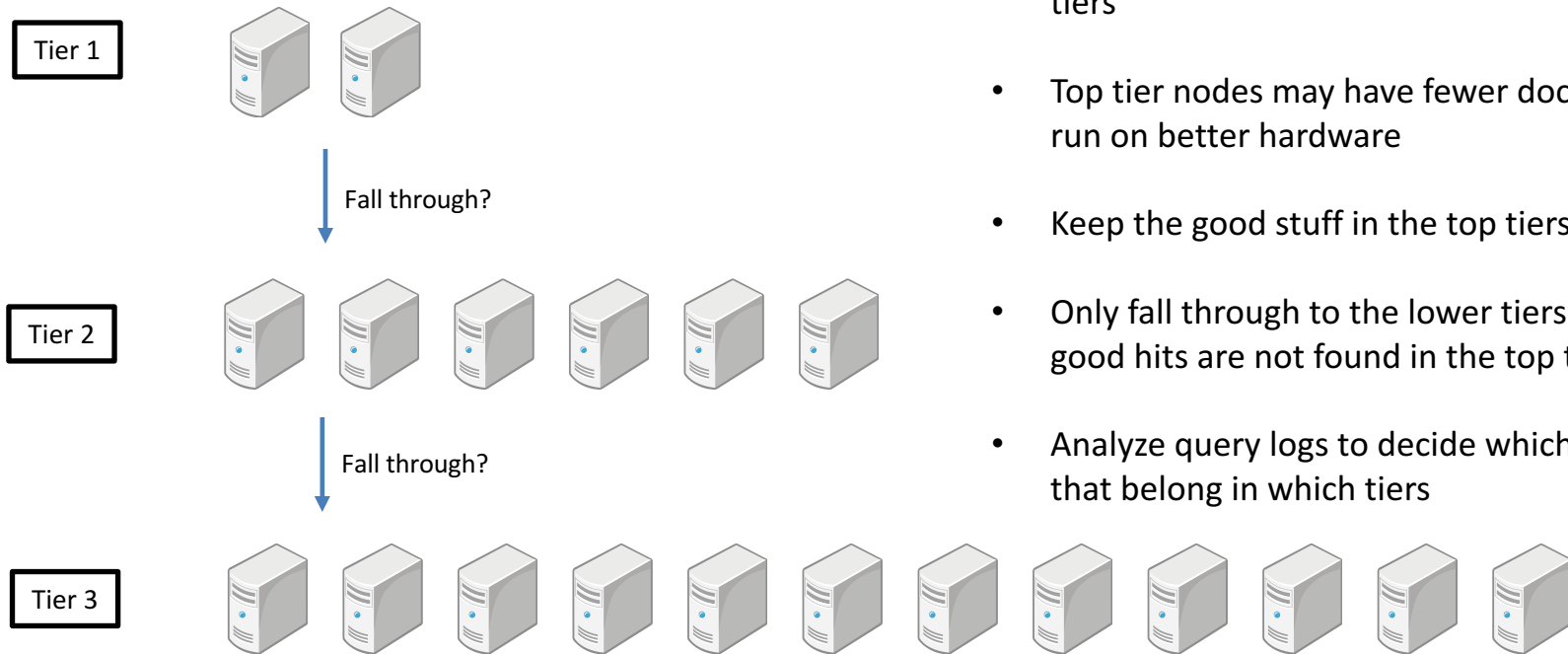
Dispatching
Merging

Searching
Caption generation

“Divide and conquer”

Searching The Content

Tiering

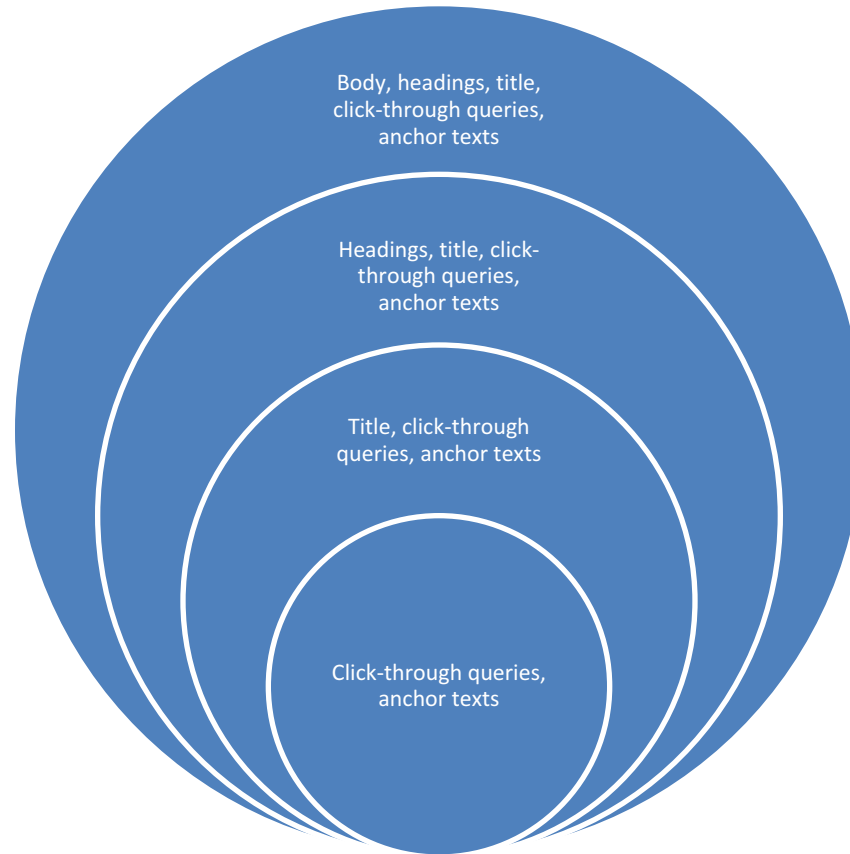


- Organize the search nodes in a row into multiple tiers
- Top tier nodes may have fewer documents and run on better hardware
- Keep the good stuff in the top tiers
- Only fall through to the lower tiers if not enough good hits are not found in the top tiers
- Analyze query logs to decide which documents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

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

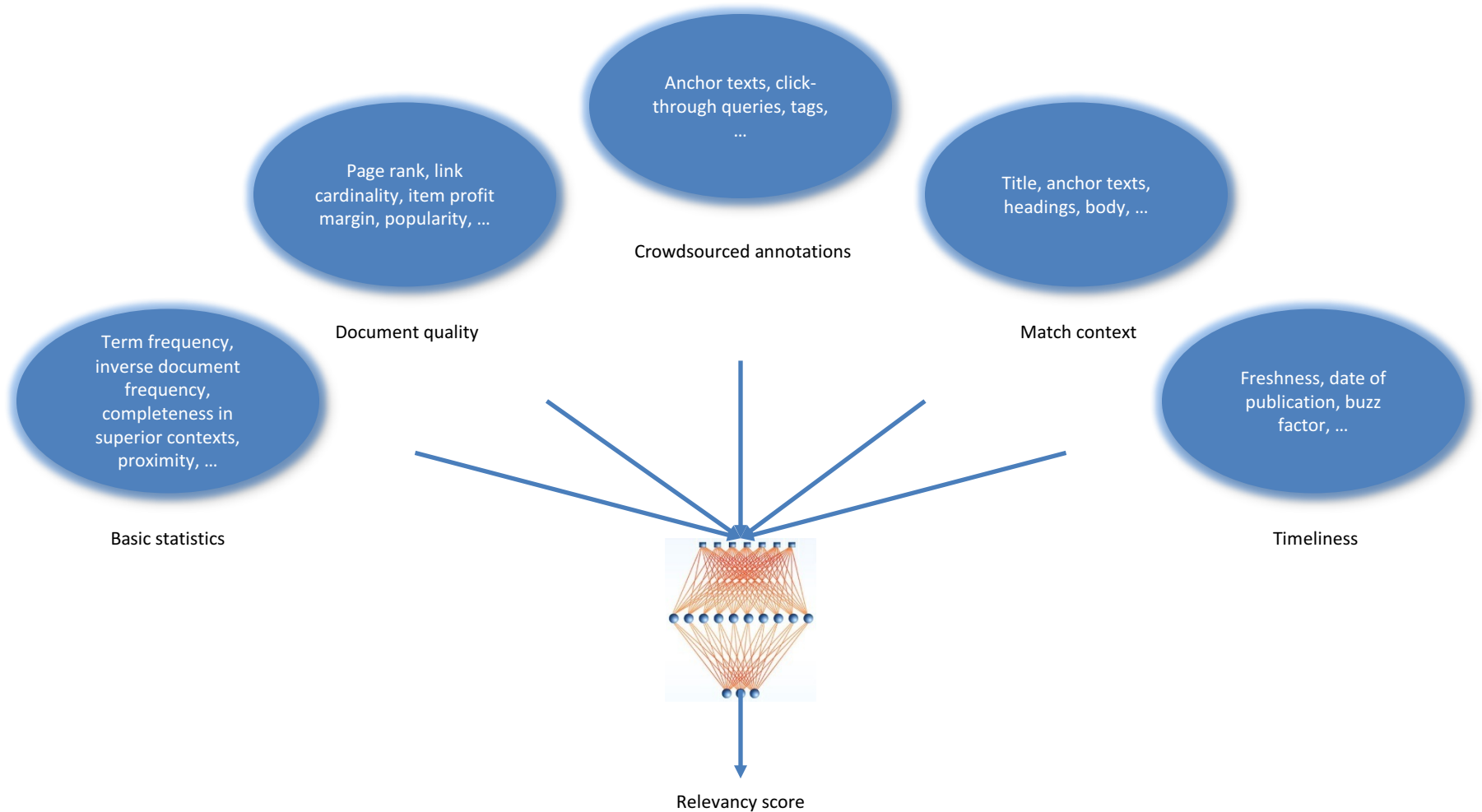
Searching The Content

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 table with columns for Source Title, Author Name, Year, Document Type, and Index Terms. It includes filters for various sources like 'Journal of Health Psychology' and 'American Journal of Cardiology', and index terms like 'heart infarction' and 'risk assessment'.
- Refine your search Panel:** A list of search terms such as 'albert einstein', 'theory of relativity', and 'special theory of relativity'.
- Main Search Results:** A Bing search result for 'deer' showing a price chart for 'Deer Consumer Products Inc (US:DEER)' with a price of 9.08 and a volume of 202.27 K.
- Clustering Panel:** A sidebar on the right titled 'clusters' showing a hierarchical list of categories like 'All Results (259)', 'Deer Hunting (34)', 'Mule, Hunts (30)', 'Family (19)', 'Wildlife (18)', and 'Fish and Wildlife (4)'. It includes a 'remix' button and a search box for clusters.

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 how failures conspire to obscure the original computation with large amounts of commodity machines are these issues.

As a reaction to this complexity, we developed an abstraction that allows us to express computations we were trying to perform in a simpler, more declarative style. This abstraction is inspired by the *map* and *reduce* operations and many other functional programming constructs. Most of our computation is expressed as a series of operations to each logical machine: compute a set of intermediate key/value pairs for a given key, in order to apply a *reduce* operation to all the intermediate values with the same key, in order to produce a final result. Our use of *map* and *reduce* is specified map and reduce functions, and a combiner as the primary mechanism for expressing the 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 declarative style with an implementation that achieves high performance on large clusters.

Section 2 describes the MapReduce programming model. Section 3 gives several examples of computations implemented in the MapReduce model. Section 4 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 and e-mails. For cost and performance reasons, processing is done on large clusters of shared-nothing commodity machines. It is challenging to develop 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 and/or special permissions from the publisher, ACM.
VLDB '08, August 24-30, 2008, Auckland, New Zealand.
Copyright 2008 VLDB Endowment. ACM 959-4-0000-000-0000-0000.

To appear in OSDI 2004

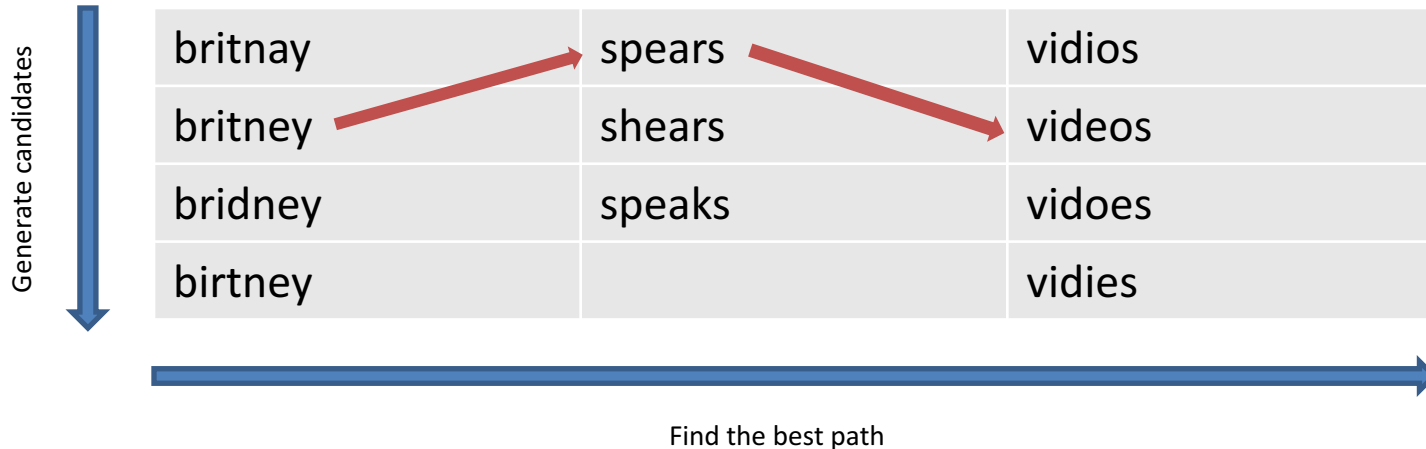
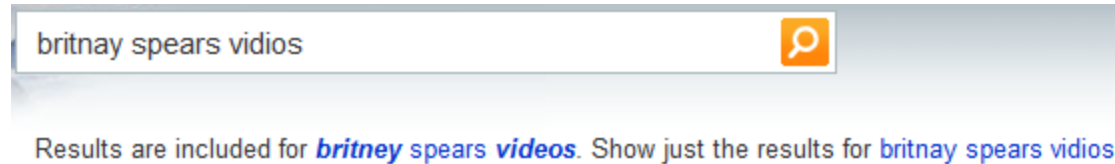


Applications

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	birtnney 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	britannia spears
1635	brittny spears	26	britneyt spears	9	britttney spears	4	bitny spears	3	britneyxxx spears	2	brittann spears
1479	brintey spears	26	brittan spears	9	brtny spears	4	brintney 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	rbitney spears	4	brbrintney spears	3	britnyey spears	2	britannt spears
1211	britnet spears	24	beitney spears	8	birtiny spears	4	breatingy 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	brattany spears	4	brtiney 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	brriatany spears	3	brityen spears	2	britenay spears
811	brtiney spears	24	brittenny spears	8	brightny spears	4	brityny spears	3	briytney spears	2	britenet spears
664	brirtney spears	24	briritni spears	8	brintay spears	4	brityey spears	3	briltney spears	2	briteniy spears
664	brintney spears	24	brirtny spears	8	brinttey spears	4	briltny spears	3	broteny spears	2	britenys spears
664	briteney spears	24	brittni spears	8	briotney spears	4	brriatany 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	birirtney spears	8	britley spears	4	brinteney spears	3	brtinay spears	2	britoryary spears
544	brittaney spears	21	brirtany spears	8	brirtneyb spears	4	brintne spears	3	brtinney spears	2	brityny spears
544	brittany spears	21	brteny spears	8	brirtney spears	4	brirtaby spears	3	brtitany spears	2	brittaney spears
364	britey spears	21	bratney spears	8	brirtny spears	4	briray spears	3	brtiteny spears	2	brittnat spears
364	brittiny spears	21	briritni spears	8	brirtner spears	4	brirainey spears	3	brtnet spears	2	brittnhey spears
329	brtney spears	21	briritane spears	8	brottany spears	4	briritnie spears	3	brytiny spears	2	brintdy spears
269	brtney spears	21	briteany spears	7	brirtney spears	4	brirtinney spears	3	btney spears	2	brirtneh spears
269	brirtneys spears	21	brittay spears	7	birrtney spears	4	brirtmney spears	3	drirtney spears	2	brirtneey spears
244	brirtne spears	21	brittinay spears	7	biteney spears	4	brirtnar spears	3	pretney spears	2	brirtney6 spears
244	brytney spears	21	brtany spears	7	bitiny spears	4	brirtnel spears	3	zbrirtney spears	2	brirtneye spears
220	breatney spears	21	brtiany spears	7	breateny spears	4	brirtneuy spears	2	barittany spears	2	brirtneyh spears
220	brtiany spears	19	brirney spears	7	brianty spears	4	brirtney spears	2	bbbrirtney spears	2	brirtneym spears
199	brirttney spears	19	brirtney spears	7	brintey spears	4	brirtmney spears	2	bbirtney spears	2	brirtneyyy spears
163	brirtny spears	19	brirtnaey spears	7	brirtianny spears	4	brirttaby spears	2	bbirtny spears	2	brirtney spears
147	breatny spears	19	brirtnee spears	7	brirtly spears	4	brirttery spears	2	bbrittany spears	2	brirtnejey spears
147	brirttiney spears	19	brirtony spears	7	brirtnej spears	4	brirtthey spears	2	beitany spears	2	brirtnee spears
147	brirtty spears	19	brirttanty spears	7	brirtneyu spears	4	brirttney spears	2	beitny spears	2	brirtnu spears
147	brirtney spears	19	brirtttney spears	7	brirttney spears	4	brirttnat spears	2	berrtney spears	2	brirtney spears
147	brirtney spears	17	brirtny spears	7	brirttnay spears	4	brirtttney spears	2	berrtny spears	2	brirttany spears
133	brirttney spears	17	brirtney spears	7	brirtttian spears	4	brirtttney spears	2	betney spears	2	brirttrey spears
133	briryney spears	17	brirtty spears	7	brirtyny spears	4	brirtttney spears	2	betny spears	2	brirtty spears
121	brittany spears	17	brirtthy spears	7	brirtttany spears	4	brirttney spears	2	bhriney spears	2	brirttany spears
121	bridtney spears	17	brirtttania spears	7	brirttney spears	4	brirttney spears	2	biney spears	2	brirtttany spears
121	brirtainy spears	15	brirttany spears	7	brirttney spears	4	brirttney spears	2	brintey spears	2	brirtttang spears
121	brirtney spears	15	brirtten spears	7	brirtttany spears	4	brirttney spears	2	brirttney spears	2	brirtttans spears
109	brirttney spears	15	brirtterney spears	6	brirttany spears	4	brirtttany spears	2	brirttany spears	2	brirtttangh spears
109	brirtthy spears	15	brirtthney spears	6	brirttney spears	4	brirttany spears	2	brirtttany spears	2	brirtttany spears
109	brirtni spears	15	brirttney spears	6	brirttney spears	4	brirttany spears	2	brirtttany spears	2	brirtttany's spears
109	brirttant spears	15	brirtttany spears	6	brirttney spears	4	brirttany spears	2	brirtttany spears	2	brirtttany spears
98	brirtney spears	15	brirtttney spears	6	brirttney spears	4	brirttany spears	2	brirtttany spears	2	brirtttany spears
98	brirtney spears	15	brirttney spears	6	brirttney spears	4	brirttany spears	2	brirtttany spears	2	brirtttany spears
98	brirttany spears	15	brirttney spears	6	brirttany spears	4	brirttany spears	2	brirtttany spears	2	brirtttany spears
98	brirttney spears	15	brirttney spears	6	brirttany spears	4	brirttany spears	2	brirtttany spears	2	brirtttany spears
89	brirtney spears	14	brirttney spears	6	brirttany spears	4	brirttany spears	2	brirtttany spears	2	brirtttany spears
89	brirtney spears	14	brirttney spears	6	brirttany spears	4	brirttany spears	2	brirtttany spears	2	brirtttany spears
89	brirttany spears	14	brirttney spears	6	brirttany spears	4	brirttany spears	2	brirtttany spears	2	brirtttany spears
89	brirttany spears	14	brirttney spears	6	brirttany spears	4	brirttany spears	2	brirtttany spears	2	brirtttany spears
89	brirttany spears	14	brirttney spears	6	brirttany spears	4	brirttany spears	2	brirtttany spears	2	brirtttany spears
89	brirttany spears	14	brirttney spears	6	brirttany spears	4	brirttany spears	2	brirtttany spears	2	brirtttany spears
89	brirttany spears	12	brirttney spears	6	brirttany spears	4	brirttany spears	2	brirtttany spears	2	brirtttany spears
89	brirttany spears	12	brirttney spears	6	brirttany spears	4	brirttany spears	2	brirtttany spears	2	brirtttany spears

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.

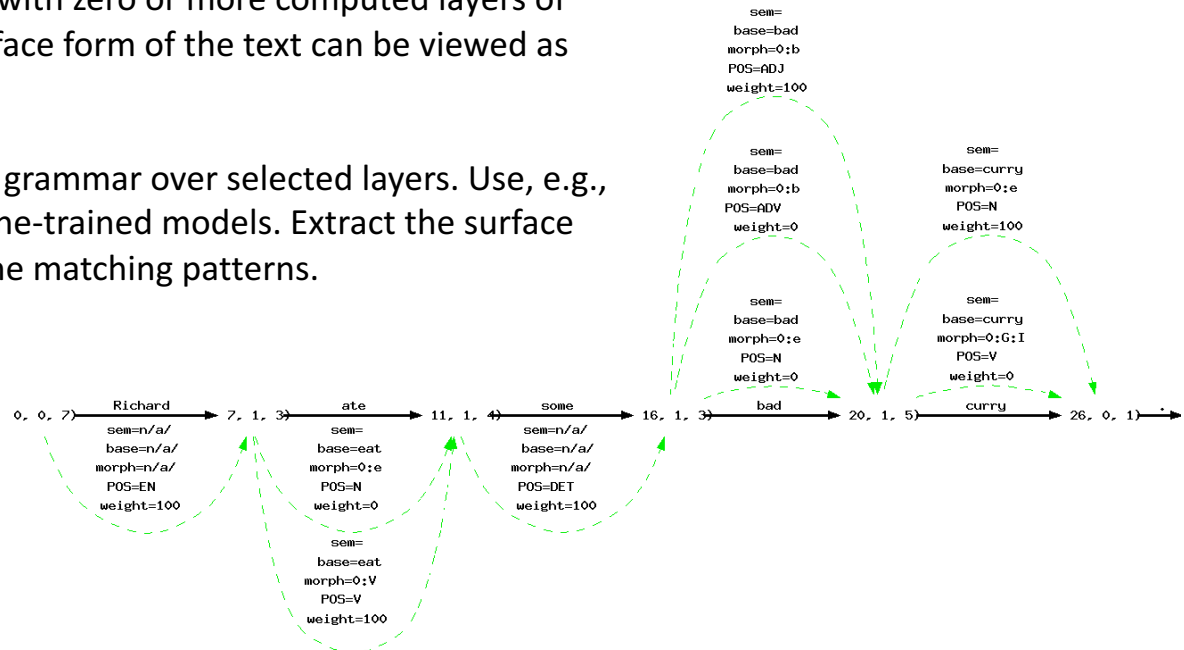
People

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

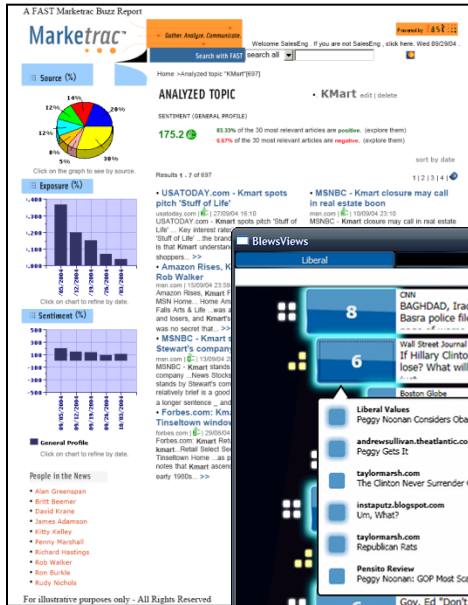
Refine your search

- 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



A screenshot of a TripAdvisor review for Jurys Boston Hotel. The review is from a member named 'TripAdvisor Member, Chicago, IL'. The review text reads: 'We spent 3 nights at Jurys' beginning on the 4th of July, they had just opened on the 2nd... Beautiful New Jurys Hotel from A TripAdvisor Member, Portsmouth, RI. 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. 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. I went down to Boston from the 19th to 5 nights and stayed at the wonderful... Verdict is in from A TripAdvisor Member. I visited Boston July 31 to August 3 and had stay at Jurys. The room was beautifully...'. The review has a 4-star rating and is from 34 reviews.

“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%)

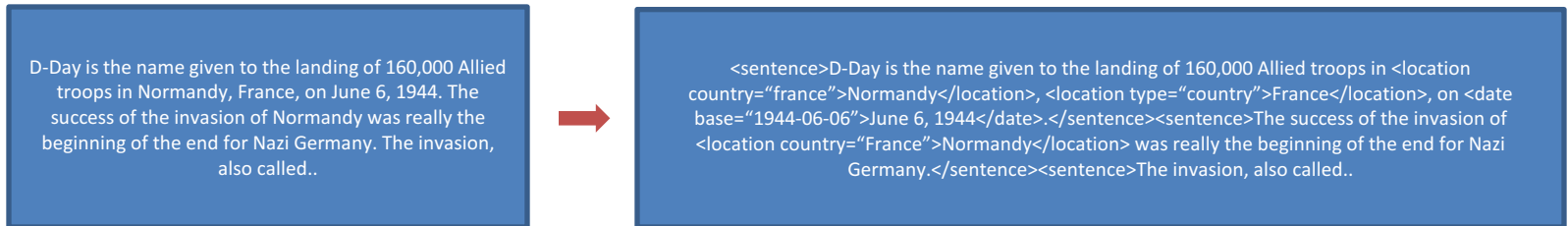
[More...](#)



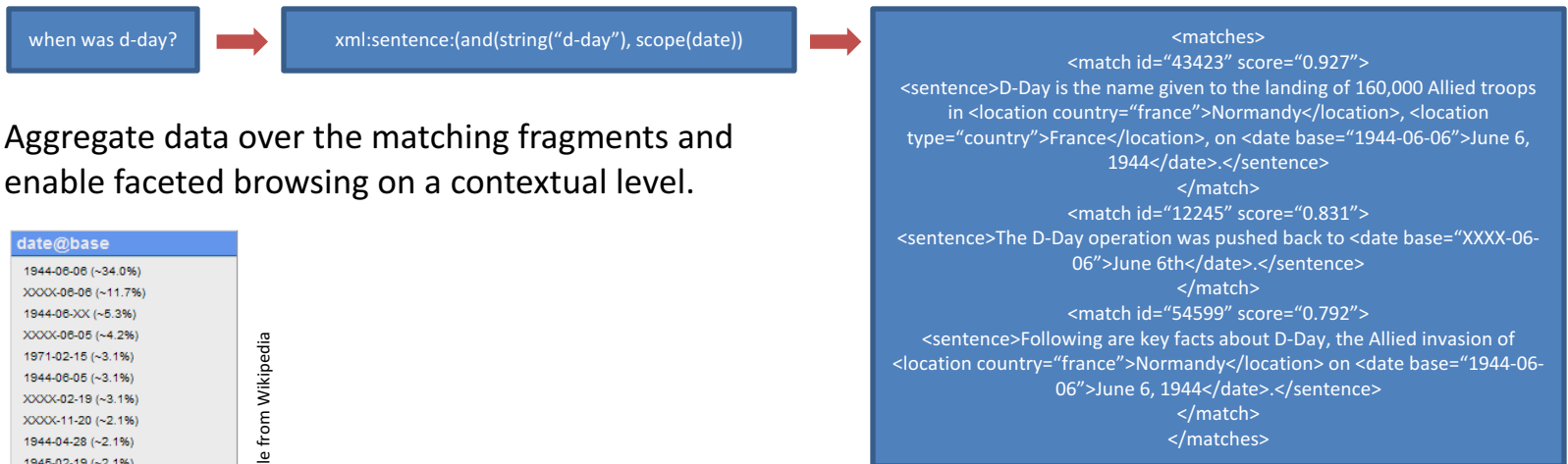
Persons that appear in **paragraphs** that contain the word {soccer}

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 website. 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 a language selection dropdown set to English, a bidirectional arrow, and another dropdown set to Danish. A "Translate" button is visible, along with "Clear All" and "Add to Favorites" options. Below this, there is a text input field with the URL "http://www.uio.no/english/about/". The translated text is displayed in a box on the right. At the bottom, there is a "New" announcement about a beta Haitian Creole translation engine.

Web Images News More MSN Hotmail Sign in Norway Preferences

bing Translator

Beta

Home | Tools | Help Free online translation service for a truly worldwide web

Languages English Danish Translate Clear All Add to Favorites

Enter text or webpage URL Report offensive translations

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.

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.

New: To help with Haitian relief efforts we have built a beta Haitian Creole translation engine. Simply select it in the Language drop-down box above.

The screenshot shows the Google Translate website. 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 prominent. Below the logo, there is a "Translation" section with links for Translated Search, Translator Toolkit, Tools and Resources, and a "Translate text, webpages and documents" section. The main area features a text input field with the same URL as the Bing Translator. Below this, there are dropdown menus for "Translate from" (English) and "Translate into" (Danish), along with a "Translate" button. The translated text is displayed in a box below. At the bottom, there is a "Contribute a better translation" link.

Web Images Videos Maps News Shopping Gmail more Help

Google translate

Translation

Translated Search

Translator Toolkit

Tools and Resources

Translate text, webpages and documents

Enter text or a webpage URL, or [upload a document](#).

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.

Translate from: English Translate

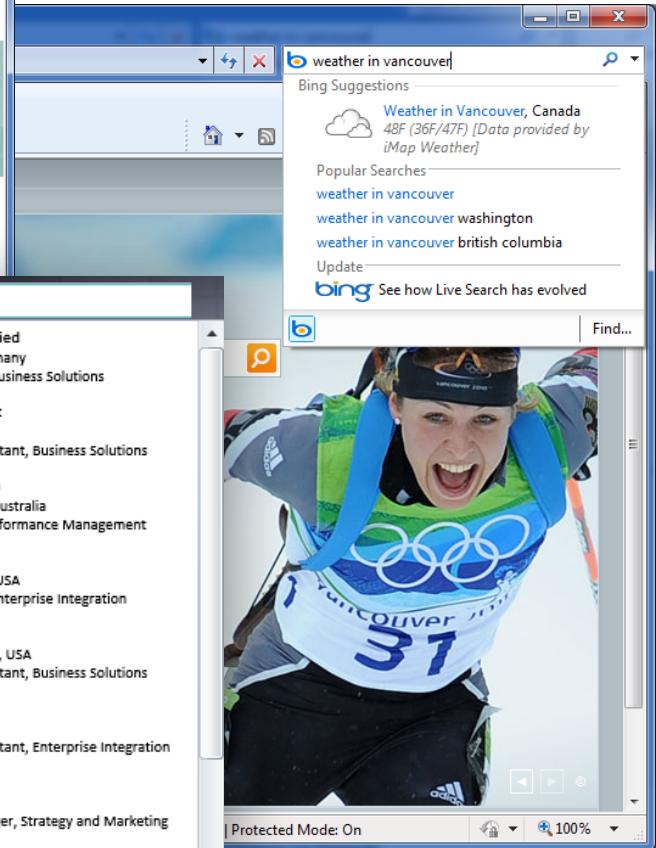
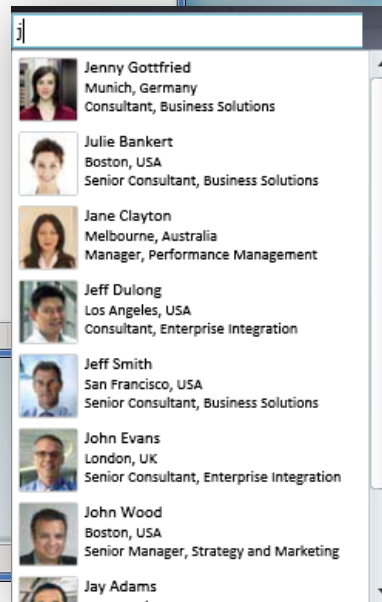
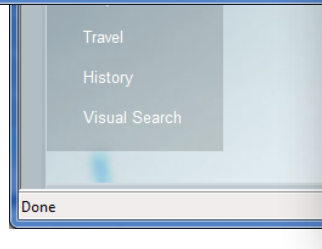
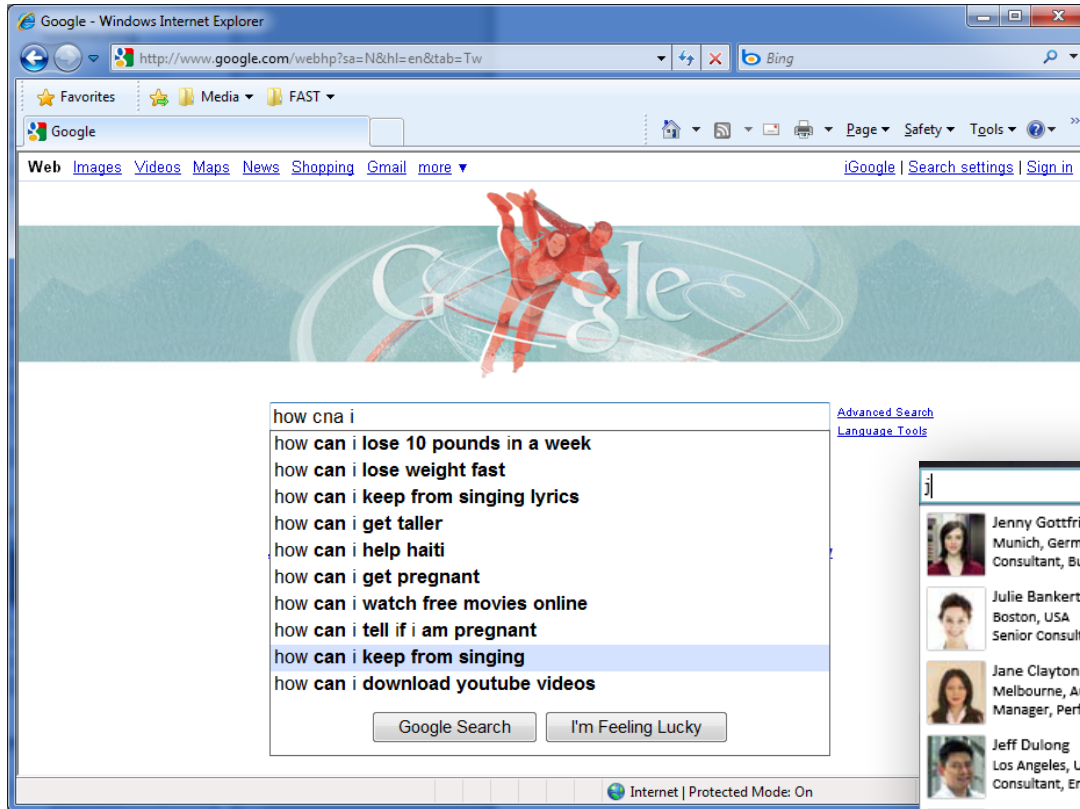
Translate into: Danish

English to Danish translation

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.

[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/gsc/companyDetails.c?g=&lineId=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

