



# Mastering Data-Intensive Collaboration and Decision Making





# Isabel Drost

## Nighttime:

Member of Apache Software Foundation.  
Organiser of Berlin Hadoop Get Together.  
Author/ Speaker on Apache Lucene, Hadoop, NoSQL.  
Reviewer OSS workshop WWW and JMLR OSS track\*.  
Co-Founder, committer and speaker Apache Mahout.

## Daytime:

Software developer.  
Guest lecturer at TU Berlin.  
Organiser Berlin Buzzwords 2010.



**Hello DICODE!**



# Hello DICODE!

Apache Hadoop?



# Hello DICODE!

Hudson?



# Hello DICODE!

Apache Lucene?



Apache HBase?

**Hello DICODE!**



# Hello DICODE!

```
patch -p0 < MAHOUT-89.diff
```



Apache CouchDB?

**Hello DICODE!**



# Hello DICODE!

Paxos?



# Hello DICODE!

Test driven development?



# Hello DICODE!

Committer?



# Hello DICODE!

BigTable?



# Hello DICODE!

NoSQL?



Maven?

**Hello DICODE!**



# Hello DICODE!

Map Reduce?



Continuous Integration?

**Hello DICODE!**



# Hello DICODE!

Apache Mahout?





Social Media

<http://vimeo.com/11677792>



# Information Management



structure  
Overload  
Data mining  
Real time updates  
Getting the word out Too much data at the same time  
Just in time information Data finds me  
collaboration Continuous updates  
organisation  
search

# Information Management



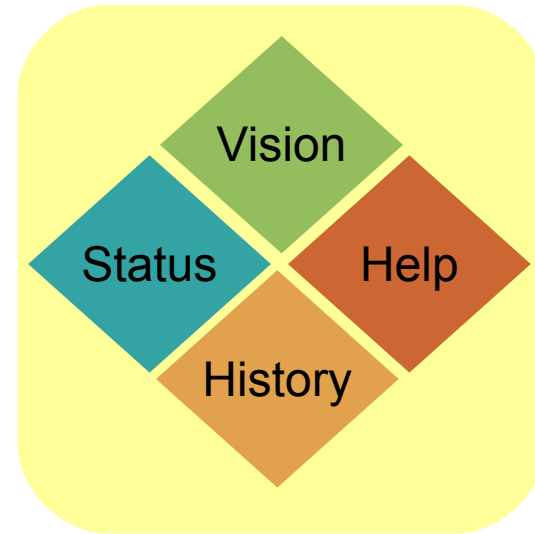
# Agenda

- Information overload: An example.
- Helping information workers.
- Whirl wind tour of IR FOSS Technologies.



# Information overload by example





Information management in a FOSS project.



Pretty new to free software development:

**Re: [Nutch-dev] Changing the score**

*Isabel Drost*

Thu, 16 Sep 2004 02:18:13 -0700

Shiraz Kanga wrote:

```
> So, to generalize what Greg said. The scoring algorithm could be based  
> upon the URL, any HTTP Header & the page content. So, it might suffice  
> to have a simple interface that looked like:  
>  
>     getScore (Url, Header, Content)
```

Scoring algorithms like pageRank not only use features, based on the page itself, but also use the link structure of the whole web(crawl) for defining the importance of a web page. Me, I would like to combine pure link based ranking with inner page features. Additionally I would like to consider the content of pages that surround (by means of links) the webpage to be scored.

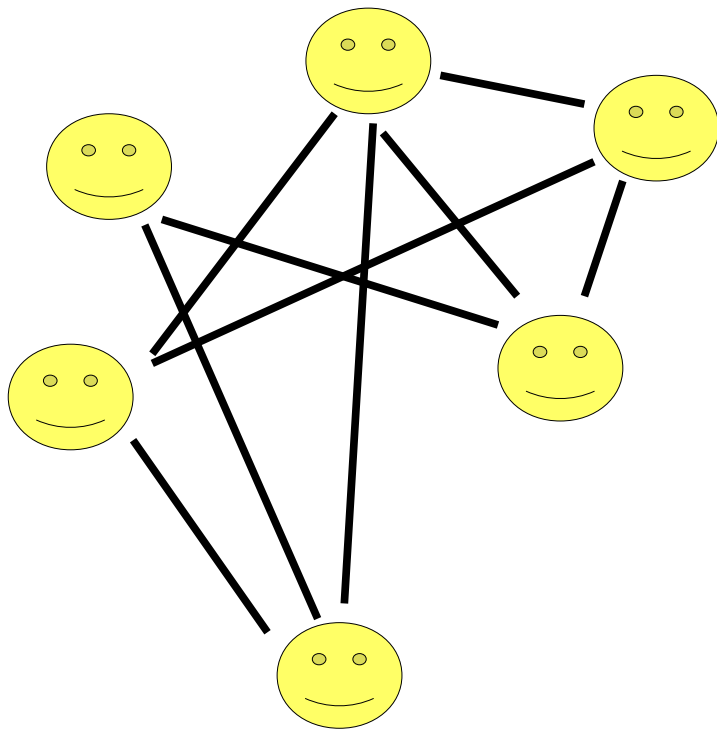
So to Doug's question "When are scores changed?": I would like to change the





# Sharing a vision

## Mess media



## Central hub

The screenshot shows the Apache Mahout website, which serves as a central hub for machine learning resources. The page features a navigation bar with links to the Dashboard, Apache Mahout, Mahout Wiki, and Quickstart. A search bar is located in the top right corner. The main content area is titled "Powered By Mahout" and includes a section for "Quickstart" with links to "Download and Installation" and "Running the Examples". Below this, there are sections for "Clustering", "Classification", and "Genetic Programming". A sidebar on the right contains a "MarkMail" search results section, displaying a list of messages related to Mahout, including topics like "Library for scalable logistic regression" and "Mahout 0.2 Released".



# Real time help requests

- Track feature discussion.
- Pick problems to work on.

Assigned: Unassigned

Reporter: [Robert Burrell Donkin](#)

Votes: 0

Watchers: 0

**Available Workflow Actions**

☐ [Submit Patch](#)

☐ [Resolve Issue](#)

**Operations**

☐ [Assign](#) this issue [\(to me\)](#)

☐ [Attach file](#) to this issue

	Assignee	Reporter	Pr	Status	Res
Example Tidy Up	Unassigned	<a href="#">Robert Burrell Donkin</a>	↑	Open	UNRES
iBase	Unassigned	<a href="#">Rabin Anil</a>	↑	Open	UNRES
Amazon Elastic MR, etc.	Unassigned	<a href="#">Grant Ingersoll</a>	↓	Open	UNRES
	Unassigned	<a href="#">Otis Gospodnetis</a>	↑	Open	UNRES
MAHOUT-118 Decode matrix methods	Unassigned	<a href="#">Daniel Nee</a>	↑	Open	UNRES
MAHOUT-60 Adding all scripts (for nightly build) to SVN repository.	Unassigned	<a href="#">Edward J. Yoon</a>	↑	Open	UNRES
MAHOUT-63 Mahout/Hama Integration	Unassigned	<a href="#">Edward J. Yoon</a>	↑	Open	UNRES

<https://issues.apache.org/jira/browse/MAHOUT-65>

**Mahout**

**Add Element Labels to Vectors and Matrices**

Created: 06/Jul/08 11:52 AM Updated: Wednesday 01:51 PM

Component/s: [Matrix](#)








Affects Version/s: 0.1

Fix Version/s: None

**Time Tracking:** Not Specified

**File Attachments:**

(Click on column header to sort)

	File Name	Da
	<a href="#">MAHOUT-65-name.patch</a>	20i
	<a href="#">MAHOUT-65-name.patch</a>	20i
	<a href="#">MAHOUT-65-name.patch</a>	20i
	<a href="#">MAHOUT-65d.patch</a>	20i
	<a href="#">MAHOUT-65c.patch</a>	20i
	<a href="#">MAHOUT-65b.patch</a>	20i
	<a href="#">MAHOUT-65.patch</a>	20i

**Issue Links:**

**Blocker**

This issue **blocks**:  
[MAHOUT-126](#) Prepare document vectors from the text

**Description**

Many applications can benefit by accessing elements in vectors and matrices using String labels in addition to

**All** **Comments** **Work Log** **Change History** **Subversion Commits** **FishEye**

[Jeff Eastman](#) added a comment - 06/Jul/08 12:09 PM

This patch introduces four new operations on Vectors:

1. `setLabelBindings(Map<String, Integer> bindings)` - sets a map of bindings between string labels and vector
2. `getLabelBindings()` - returns the map, creating an empty one if needed
3. `set(String label, double value)` throws `UnboundLabelException`, `IndexException` - sets the labeled element or
4. `get(String label)` throws `UnboundLabelException`, `IndexException` - returns the value indexed by the given label

This patch required reworking of the serializing/deserializing code so that the maps are included in the format run.

This is about as minimal a change as I could imagine, and I chose to just expose the map rather than try to do an kind of functionality that people will find useful?

Jeff

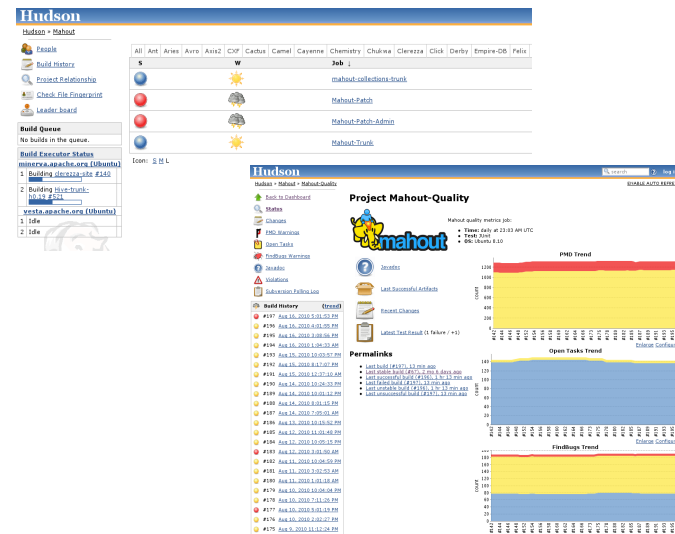


# What is the status - updates

No feedback.

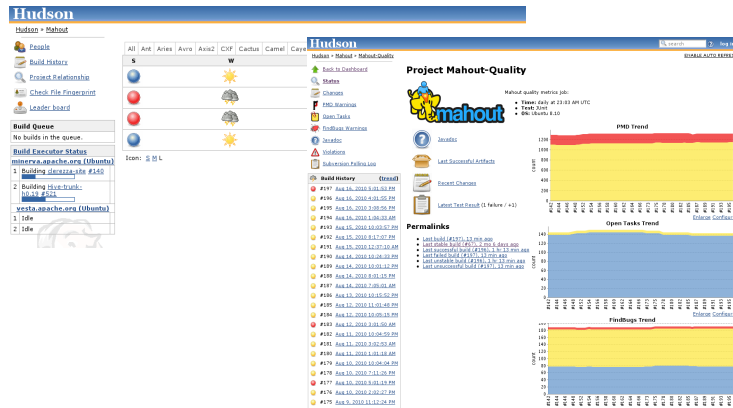


Fast feedback.

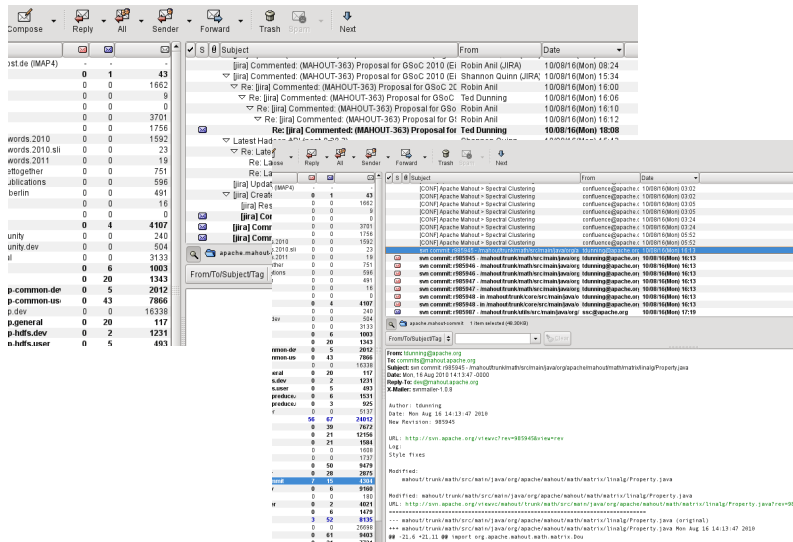




# Fast Feedback

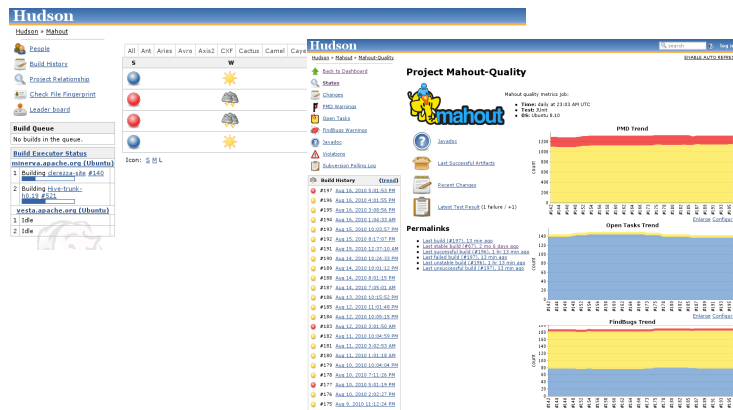


Mailing lists: Commits, Wiki changes, Issue tracker, user requests, organization

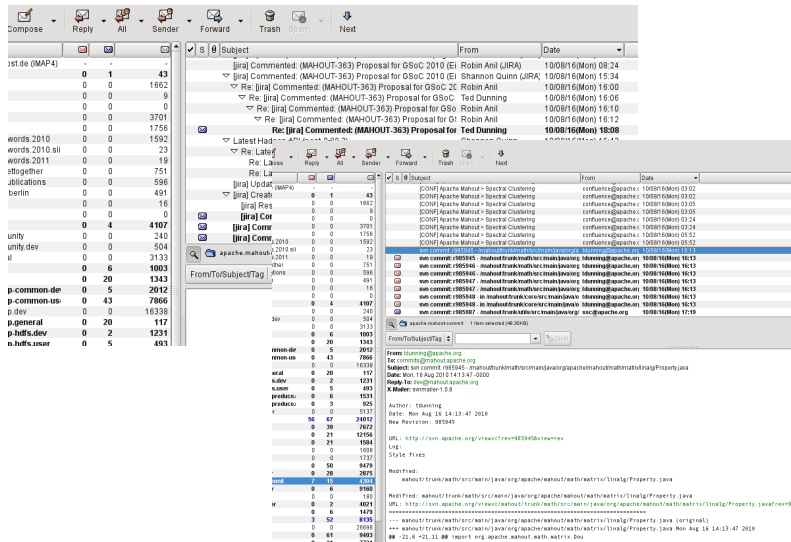




# Fast Feedback

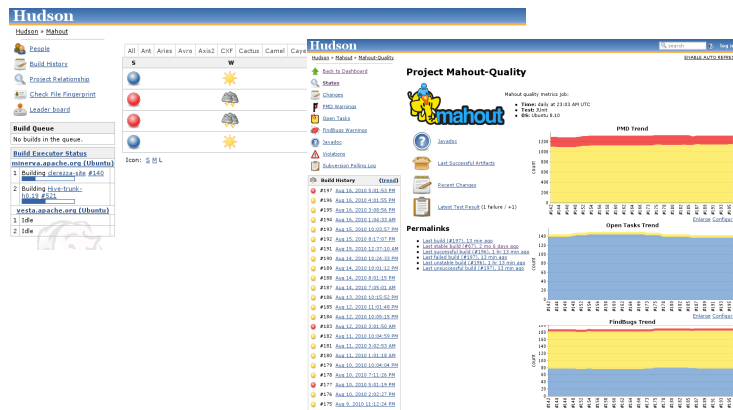


Mailing lists: Commits, Wiki changes,  
Issue tracker, user requests, organization






# Fast Feedback



# Updates from the web



Search terms:

Type:

Everything

▼

How often:

once a day

▼

Email length:

up to 20 results

▼

Your email:

Create Alert

Preview results

Monitor the Web for interesting new content

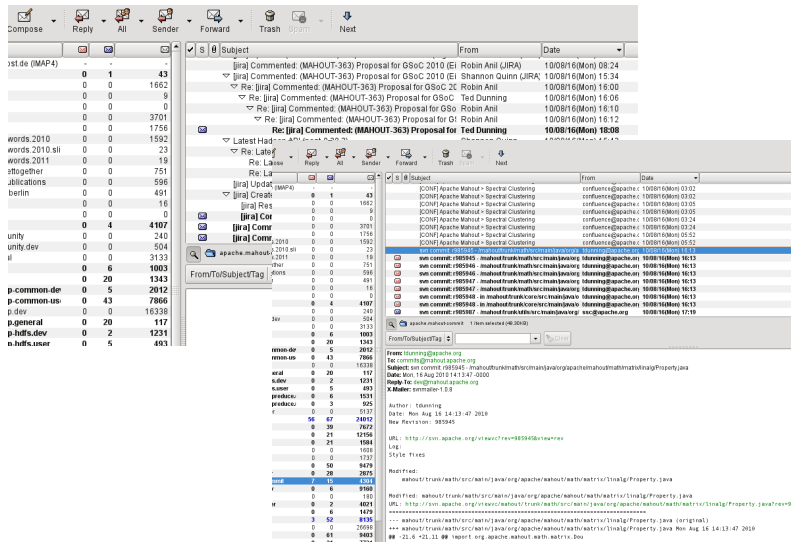
Google Alerts are email updates of the latest relevant your choice of query or topic.

Title	Author	Date
开源项目——apache.Mahout   写书：三学一动	LiAndy	Yesterday 17:26
run RowSimilarityJob job   新IT民工	admin	Friday 11:20
FeatherCast » Blog Archive » Episode 69 - Apache Mahout	Rich	Wednesday 13:28
Apache Tomcat - TR120 Hydraulic Drilling Rig - Diaphragm Wall ...	pretty	2010-08-08 19:55
Paste number 113299: org column view	unknown	2010-08-07 22:43
Authors' Podcast: Lucene in Action, Second Edition	grant_jingersoll	2010-08-03 01:13
Project Pitch: Quotation clustering with Mahout and Lucene	unknown	2010-08-02 23:21
Bayesian - Apache Mahout - Apache Software Foundation	unknown	2010-07-24 17:56
Mahout: Mammoth Scale Machine Learning	unknown	2010-07-23 20:21
RTI (KILL OR ATTACK) - Cluster search results? Google Predict ...	govindkanshi	2010-07-21 09:42
Twitter Weekly Updates for 2010-07-18 - nicolas leroy	nicolasleroy	2010-07-18 09:00
Apache Hadoop in Debian Squeeze   Debian-News.net - Your one stop ...	cj2003	2010-07-17 20:53

Episode 69 - **Apache Mahout**, August 11th, 2010. Better late than never, here's my conversation with Sean Owen about **Apache Mahout**, which graduate can listen to it [HERE](#), or via iTunes. ...

[Complete Story](#)

## Mailing lists: Commits, Wiki changes, Issue tracker, user requests, organization



## (Micro) blogs

[illegible]



# Archives

- Documentation: No time.
- Transparency and Traceability.
  - Code.
  - Discussions.
  - Decisions.







Code

Wiki pages

Project web page

SCM commit log

Developers' blogs

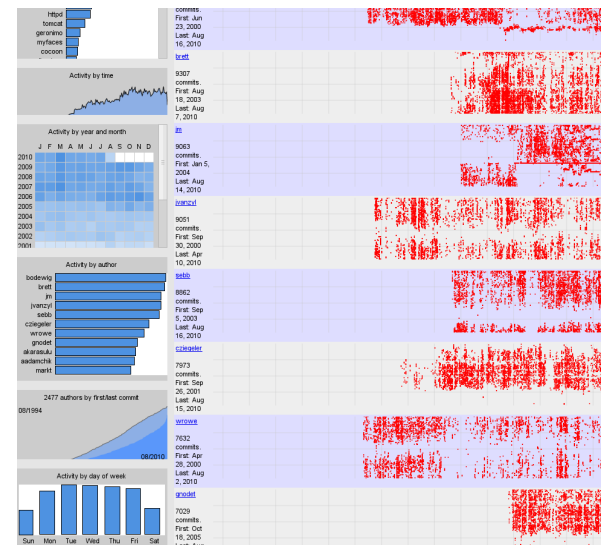
Developers' tweets

Mailing list archives

Issue tracker comments



# Tracing svn logs



<http://svnsearch.org/svnsearch/repos/ASF/search?view=plot&kind=day&kind=week&plotsort=commits>



# Searching code

The screenshot shows the Krugle search engine interface. At the top, there are navigation links: Home, Help, and Company. Below this is a search bar with the text 'mahout' entered. The search results are displayed in a table with columns for 'Code Search for mahout' and 'Related Results'. The results show various code snippets and links to related projects, including 'ClusteringComponent.java' and 'pom.xml'. A sidebar on the right contains a 'krugle BASIC' advertisement with the text 'Profit from your existing code', 'Manage code duplication', 'Fix bugs once across all projects', and a 'Download' button.

# Tracing svn logs



<http://svnsearch.org/svnsearch/repos/ASF/search?view=plot&kind=day&kind=week&plotsort=commits>



# Mailing lists

# Searching code

**krugle** 2.0.1.0.1 [Home](#) | [Help](#) | [Company](#)

[Open Source Code](#) | [Open Source Projects](#) | [SCM Comments](#)

[mahout](#) [Advanced Search](#) [Language](#) [Project](#) [Enter project name](#)

**Results** [Create Link](#)

Code Search for **mahout** 1 2 3 4 Next

Code Files 1-10 (out of about 32 matching files)

**ClusteringComponent.java** | [S](#) | [A](#) | [Apache-2.0](#)

```

37  /**
38   * Provide a plugin for clustering results. Can either be for search results (i.e. via Carcot2) or
39   * clustering documents (i.e. via Mahout)
40   */
41   * This engine is experimental. Output from this engine is subject to change in future releases.

```

**Related Results**

**krugle BASIC**  
Profit from your existing code  
- Manage code duplication  
- Fix bugs once across all projects  
Download your FREE copy of Krugle Basic

**pom.xml** | [U](#) | [UNICORE](#) | [Other](#)

```

23  <dependency>
24  <dependency>
25  <groupId>org.apache.mahout.hadoop</groupId>
26  <artifactId>mahout-core</artifactId>
27  <version>0.20.0</version>

```

**index.html** | [M](#) | [Apache-2.0](#)

... March 24th Introducing Apache Mahout - Grant Ingersoll, March 25 ...

**index.html** | [S](#) | [Apache-2.0](#)

... March 24th Introducing Apache Mahout - Grant Ingersoll, March 25 ...

**index.pdf** | [S](#) | [Apache-2.0](#)

... March 24th Introducing Apache Mahout - Grant Ingersoll, March 25 ...

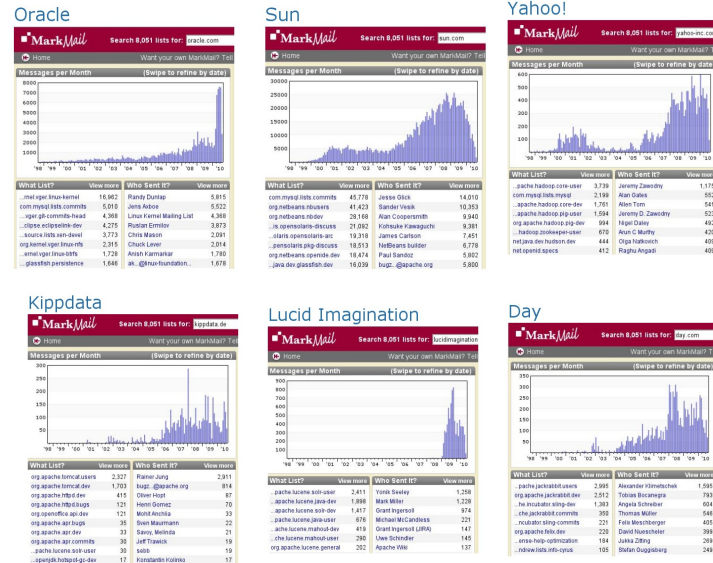
**index.xml** | [M](#) | [Apache-2.0](#)

```

51  <li><a href="http://eu.apachecon.com/c/aceu2009/sessions/136">Introducing Apache Mahout</a> - Grant Ingersoll, March 25th @ 10:38/11>
60  <li><a href="http://eu.apachecon.com/c/aceu2009/sessions/137">Lucene/Solr Case Studies</a> - Erik Batcher, March 25th @ 11:30/11>

```

WordsCommonKeysWithNCaps.txt | [H](#) | [HyperType \(berlios\)](#) | [GPL-2.0](#)



# Tracing svn logs



<http://svnsearch.org/svnsearch/repos/ASF/search?view=plot&kind=day&kind=week&plotsort=commits>



<http://svnsearch.org/svnsearch/repos/ASF/search?view=plot&kind=day&kind=week&plotsort=commits>



# Information management

- BI over collaboration data.



# Integrated Information Management



Search



Discovery

Search



Recommend  
Content

Location  
Aware

Discovery

Search



Recommend  
Content

Location  
Aware

Discovery

Organisation

Search



Recommend  
Content

Location  
Aware

User  
Feedback

Discovery

Organisation

Search



Adding recommendations



17 30  
20 45  
SO 14 30  
BRAD PITT DIANE KRUGER  
INGLOURIOUS BASTERDS.  
BY QUENTIN TARANTINO  
DOLBY STEREO SR-D

ODEON





# Show most relevant ads



### ALUMINIUM Baseballschläger 30" American Baseball

von [Outdoor 4 You - Shop](#)

★★★★☆ (4 Kundenrezensionen) [Mehr zu diesem Artikel](#)

---

Preis: **EUR 17,58**

**Auf Lager.**  
Verkauf und Versand durch [NORMANI](#).

Noch 5 Stück auf Lager.

**4 neu** ab EUR 17,58

Marken-Uhren mit Tiefpreis-Garantie finden Sie im [Uhren-Shop](#) bei Amazon.de/Uhren.

[Größeres Bild](#)

Für Kunden: Stellen Sie Ihre eigenen Bilder ein.

## Produktmerkmale

- Baseballschläger aus Aluminium
- mit rutschfestem Griff
- Absoluter Hammerpreis

## Wird oft zusammen gekauft

Kunden kaufen diesen Artikel zusammen mit [Baseball in Official Size & Weight](#) von IMPI Sports

 + 



**Preis für beide: EUR 22,57**

**Beides in den Einkaufswagen**

Diese Artikel werden von verschiedenen Verkäufern verkauft und versendet. [Details anzeigen](#)

## Kunden, die diesen Artikel gekauft haben, kauften auch

Seite 1 von 23



[Leder Quarzsandhandschuhe schwarz S-XXL](#)



[Balaclava 3-Loch](#)  
★★★★☆ (4) EUR 3,50



[Pfefferspray KO-FOG 40ML](#)  
★★★★☆ (9) EUR 5,95



[Baseballschläger Holz 32" American Baseball natur](#)



# Show most relevant ads

clearasil®

erwandte Suchbegriffe: [clearasil pickelstift](#), [world of warcraft](#).

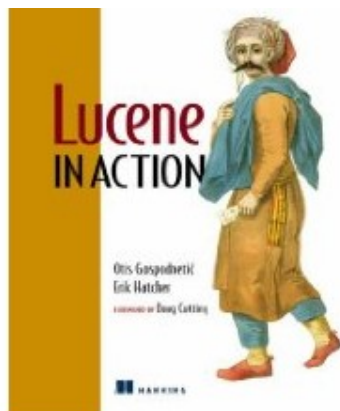
este Ergebnisse

« Zurück | Seit

1.  **Clearasil 44161 Tiefenreinigung Antibakterielle Reinigungspads, 60er Packung**  
Neu kaufen: **EUR 5,99** 2 Angebote ab **EUR 5,15**  
Lieferung bis **Samstag, 22. März**: Bestellen Sie innerhalb der nächsten **23 Stunden**  
**Kostenlose** Lieferung möglich.  
★★★★★ (1)  
**Drogerie & Bad:** Alle 13 Artikel ansehen
2.  **Clearasil Ultra Anti-Pickel Reinigungspads, 65 Stück von Clearasil (Ba**  
Neu kaufen: **EUR 7,99**  
Gewöhnlich versandfertig in 1 bis 3 Wochen.  
**Kostenlose** Lieferung möglich.  
**Drogerie & Bad:** Alle 13 Artikel ansehen
3.  **World of Warcraft: Wrath of the Lich King (Add-on)** von Vivendi Unive  
Vista / XP)  
Neu kaufen: **EUR 39,99**  
Vorbestellbar  
**Kostenlose** Lieferung möglich.  
**Games:** Alle Artikel ansehen



# Show most relevant ads



[See larger image](#)

[Share your own customer images](#)

[Publisher: learn how customers can search inside this book.](#)

~ [Otis Gospodnetic](#) (Author), [Erik Hatcher](#) (Author)  
★★★★☆ (20 customer reviews)

List Price: ~~\$44.95~~

Price: **\$29.67** & this item ships for **FREE with Super Saver Shipping**. [Details](#)  
You Save: **\$15.28 (34%)**

**In Stock.**

Ships from and sold by **Amazon.com**. Gift-wrap available.

[23 new](#) from \$19.99    [15 used](#) from \$17.11

	Also Available in:	List Price:	Our Price:	Other Offers:
Paperback (2)		<del>\$44.99</del>	<b>\$29.69</b>	

## Frequently Bought Together

Customers buy this book with [Building Search Applications: Lucene, LingPipe, and Gate](#) by Manu Konchady



+



**Price For Both: \$70.12**



Add both to Cart

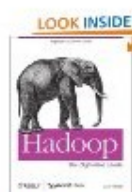
Add both to Wish List

[Show availability and shipping details](#)

## Customers Who Bought This Item Also Bought



[Solr 1.4 Enterprise Search Server](#) by David



[Hadoop: The Definitive Guide](#) by Tom White



[Hibernate Search in Action](#) by Emmanuel



[Introduction to Information Retrieval](#) by



[Collective Intelligence in Action](#) by Satnam Alag



# Recommending places

<http://www.flickr.com/photos/jfclere/4061801735>

<http://www.flickr.com/photos/25831000@N08/4156701164>

<http://www.flickr.com/photos/alainpicard/4175214747>

[http://www.flickr.com/photos/joachim\\_s\\_mueller/2417313476/](http://www.flickr.com/photos/joachim_s_mueller/2417313476/)

[http://www.flickr.com/photos/sebastian\\_bergmann/1244514498](http://www.flickr.com/photos/sebastian_bergmann/1244514498)

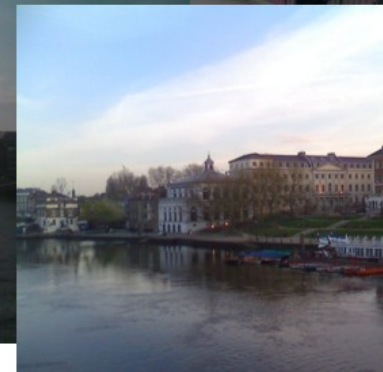
<http://www.flickr.com/photos/philfotos/4510197138/>

[http://www.flickr.com/photos/claudio\\_ar/2643165035/](http://www.flickr.com/photos/claudio_ar/2643165035/)

[http://www.flickr.com/photos/claudio\\_ar/2643180457/](http://www.flickr.com/photos/claudio_ar/2643180457/)



Thanks to Falko Menge for the pictures of Brussels.





# Recommending people

**People you may know**

**Shane Curcuru**  
Conference Lead at Apache  
Softw

**Ugo**  
Com  
Profe

**Arjé**  
CTO  
lead,  
Mem  
Foun

Invite | x

Invite | x

Invite | x





# Organisation of information







Any size

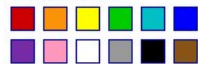
[Medium](#)  
[Large](#)  
[Icon](#)  
[Larger than...](#)  
[Exactly...](#)

Any type

[Face](#)  
[Photo](#)  
[Clip art](#)  
[Line drawing](#)

Any color

[Full color](#)  
[Black and white](#)  
[Specific color](#)



Related searches: [oakland raiders](#)



**Oakland Airport**  
625 x 471 - 103k - jpg  
[visitingdc.com](#)



**Oakland Ranks Fifth**  
538 x 359 - 44k - jpg  
[bayareahomegirl.com](#)



**Oakland**  
900 x 600 - 171k - jpg  
[globalsecurity.org](#)



**OAKLAND First**  
400 x 400 - 27k  
[bayassociation.org](#)



**Oakland Gaudy Lexus 2**  
500 x 340 - 73k - jpg  
[lexusenthusiast.com](#)



**The Oakland Baseball**  
450 x 305 - 39k - jpg  
[sportsbusinesssims.com](#)



**Oakland**  
600 x 320 - 48k - jpg  
[webpages.scu.edu](#)



**Learn more about**  
550 x 366 - 56k - jpg  
[tripadvisor.com](#)

Any size

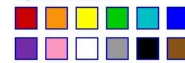
[Medium](#)  
[Large](#)  
[Icon](#)  
[Larger than...](#)  
[Exactly...](#)

Face

[Photo](#)  
[Clip art](#)  
[Line drawing](#)

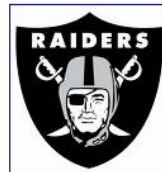
Any color

[Full color](#)  
[Black and white](#)  
[Specific color](#)

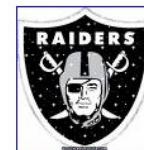


[Reset options](#)

Related searches: [oakland raiders](#)



**Oakland, CA 94621**  
513 x 545 - 13k - gif  
[nflfootballstadiums.com](#)



**All Graphics »**  
262 x 278 - 43k - gif  
[coolchaser.com](#)



**Oakland Sideshow and**  
720 x 480 - 44k  
[channels.com](#)



**Detroit Tigers v**  
594 x 396 - 51k  
[zimbio.com](#)



**Oakland, by Davey D**  
359 x 512 - 26k - jpg  
[sfbayview.com](#)



**Detroit Tigers v**  
443 x 594 - 74k  
[zimbio.com](#)



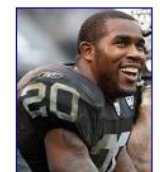
**Dallas Cowboys v**  
594 x 404 - 60k  
[zimbio.com](#)



**Distributed by Tubemogul. The**  
720 x 480 - 18k  
[channels.com](#)



**oakland@coe.ufl.edu**  
379 x 471 - 110k - jpg  
[coe.ufl.edu](#)



**#20 of the Oakland**  
467 x 594 - 86k  
[zimbio.com](#)



Discovering new information



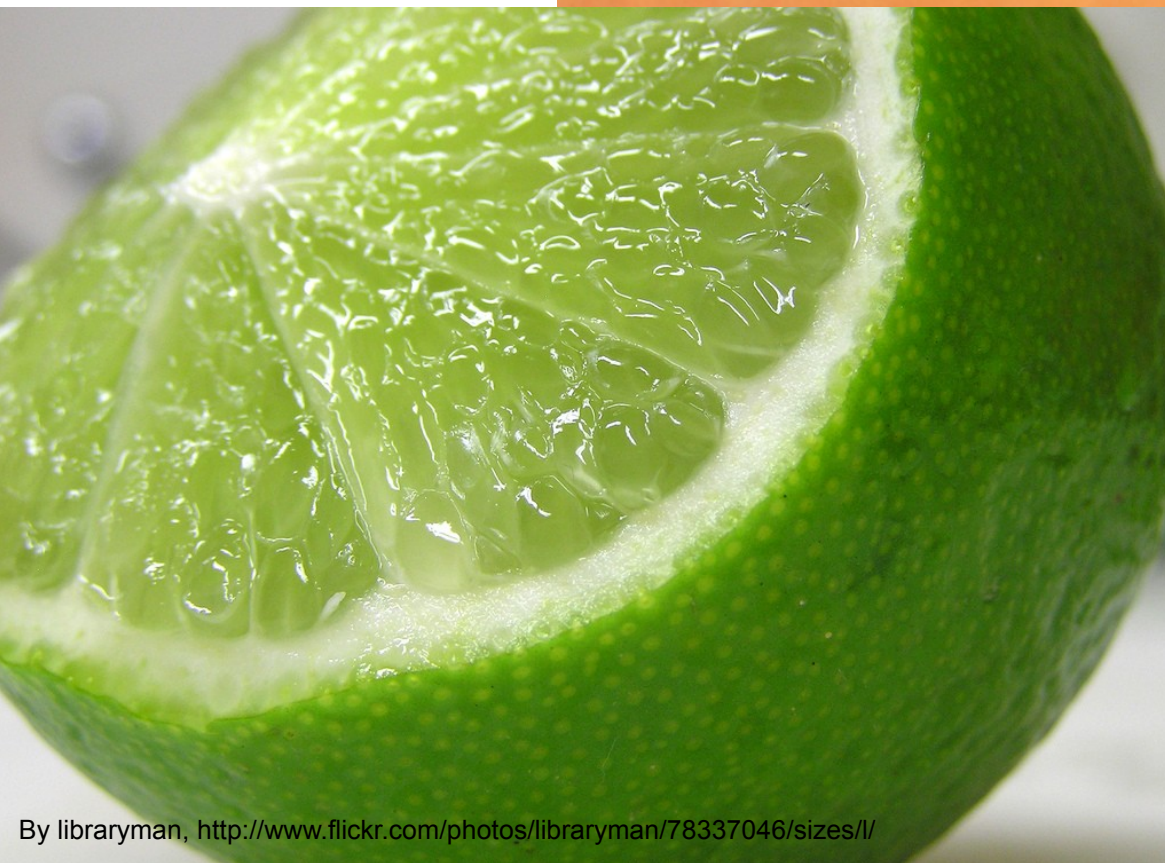


January 3, 2006 by Matt Callow  
<http://www.flickr.com/photos/blackcustard/81680010>









By crypto, <http://www.flickr.com/photos/crypto/3201254932/sizes//>

By libraryman, <http://www.flickr.com/photos/libraryman/78337046/sizes//>





By quinnanya, <http://www.flickr.com/photos/quinnanya/2806883231/>



By libraryman, <http://www.flickr.com/photos/libraryman/78337046/sizes//>



By crypto, <http://www.flickr.com/photos/crypto/3201254932/sizes//>



Visit our premier sponsor stores on CNET

Powermat

Samsung Phones

Blackberry

Sprint Phones



## Get CNET on your mobile device

Learn about CNET's mobile Web site, which features all the reviews, news, and video you'll find on CNET--optimized for your mobile device.

Related searches: [mobile software](#), [india](#), [virgin-mobile](#), [t-mobile](#), [virgin mobile](#)

6,255 PRODUCT REVIEW RESULTS FOR "mobile"

0 COMPARE SELECTED

### NARROW YOUR RESULTS

#### PRODUCT REVIEWS (6,255)

[Cell Phones \(1,782\)](#)

[Cell Phones and Accessories \(2,432\)](#)

[Mobile Files \(6\)](#)

[Personal Communication and Navigation \(2,515\)](#)

[See all matching products +](#)

#### NEWS, PHOTOS & VIDEOS (22,179)

[Today \(10\)](#)

[Past 7 days \(102\)](#)

[Past 30 days \(392\)](#)

[Past year \(4,048\)](#)



### Sony Ericsson Xperia X10 - sensuous black (unlocked)

Reviewed on 03/23/2010

Editors' rating:



User rating:



**CNET's bottom line:** The Sony Ericsson Xperia X10 offers a slick, user-friendly design with a solid feature set and good call quality. Data performance was uneven, however, and the Timescape application and keyboard miss some details. [Read CNET's review](#)

As shown: \$999.00 See manufacturer site for availability

[+ Save to list](#)

[Compare](#)



### Samsung Vibrant (T-Mobile)

Reviewed on 07/14/2010

Editors' rating:



User rating:



**CNET's bottom line:** Sleek and fast, the Samsung Vibrant is one of T-Mobile's top smartphones for entertainment, but its feature set takes a toll on the battery life. [Read CNET's review](#)

**\$559 to \$589**  
at 4 stores

[+ Save to list](#)

[Compare](#)



### Nokia 5230 Nuron (T-Mobile)

Reviewed on 03/22/2010

**\$299**  
at 1 store



Options

[clear projects](#)

isabell drost



BETA


[Start new search](#)

## Search Results for isabell drost

Found 890 results in 0.108 seconds. Displaying page 1 of 89, sorted by relevancy ▼

Did you mean: [isabel drost](#)





About | **New features!** | Search feeds | Search plugins | Download | Can't search | ...

Web MSN Yahoo Wiki Images News Jo

tiger

Tree Visualization

- All Topics (100)
  - Tiger Woods (14)
  - Panthera Tigris (8)
  - News (5)
  - Tiger Woods PGA TOUR (5)
  - Wikipedia (5)
  - Music (4)
  - Resources on Tigers (4)
  - Siberian Tiger (4)
  - Team (4)
  - Company (3)
  - [more](#) | [show all](#)

Top 100 results of about **15000000** for

- 1 [TigerDirect.com - Computers, Com](#)  
The Best Deals on Computers, Con  
**TigerDirect.com.**  
<http://www.tigerdirect.com/> [Cuil, E
- 2 [Tiger - Wikipedia, the free encyclo](#)  
The **tiger** (*Panthera tigris*), a memb  
Native to much of eastern and ...  
<http://en.wikipedia.org/wiki/Tiger> [/
- 3 [TigerWoods.com: The Official Site](#)  
**Tiger's** official site, in English and J  
<http://www.tiaerwoods.com/> [Ask. ]



Location aware information





looking for .. (eg Nike)

in or close to .. (eg The Hague)

[or choose a category](#)
[Select a district](#)

[Results](#)
[Map](#)

[Visit the Amsterdam homepage](#) | 
 [Google](#) | 
 [Facebook](#) | 
 [Twitter](#) | 
 [RSS](#) | 
 [Go to TomTom.com](#)

34,802 results found for Amsterdam

Home (Amsterdam)

Location: Ontario > Amsterdam

1 to 10 of 84,802 | next


Refine Results

- Categories
- Consultancies (7472)
  - Studios and prod .. (4275)
  - Advertising (3324)
  - Clothing (3197)
  - Artists (2995)

- Distance
- From central London:
  - within 0.5 km (4652)
  - within 1 km (14 391)
  - within 2 km (32 745)
  - within 3 km (51 561)

- District
- Downtown (19 863)
  - Oud-Zuid (12 501)
  - Zuideramstel (5903)
  - North (5881)
  - East and Watergra .. (5833)

Sponsored Results



The Bicycle Friendly

Veerstraat 27/2HG, 1075 SM Amsterdam

Category: Courier

Getting letters and parcels in and around Amsterdam by bike 8.30-18.30 hours ..

[more info](#) | 
 [e-mail](#) | 
 [website](#) | 
 [route](#) | 
 [add to](#)

3.3 km

write review

de fietsvriende



A la Bonne Chaise

From Hogendorpstraat 107-2, 1051

Category: Furniture Upholstery

For professional reupholstery and trusted furniture ..

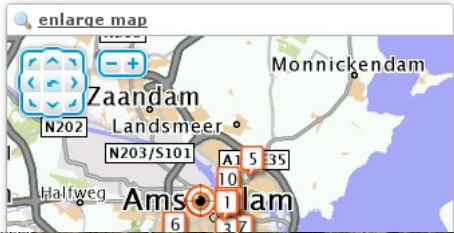
[more info](#) | 
 [e-mail](#) | 
 [website](#) | 
 [route](#)

1.4 km

write review

All results sorted by: relevance | distance

- 1 Im Amsterdam
- Herengracht 418-1hg, 1017 BZ Amsterdam
- Category: consultancy
- [more info](#) | 
 [phone](#) | 
 [website](#) | 
 [route](#)
- 2 Asito Amsterdam Centre
- Keienbergweg 50/C-50/D, 1101 GO Amsterdam
- Category: Janitor





# Whirl wind tour of FOSS IR



# Once upon a time



Lucene: A java library, indexing made trivial.



# Once upon a time



A library is not enough!



# Once upon a time



Scaling to the world wide web (of 2003/4).



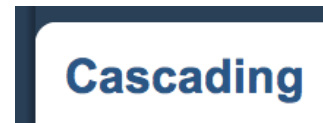
# Once upon a time



Adding more functionality



# Once upon a time



**Sqoop**  
SQL-to-Hadoop database



... and even more functionality.



**Cassandra**

A highly scalable, eventually consistent, distributed, structured key-value store.

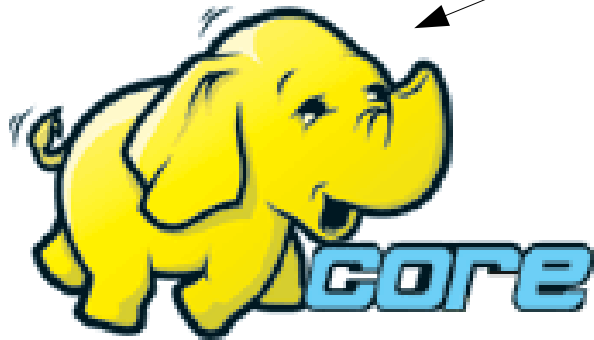


More than Log Analysis

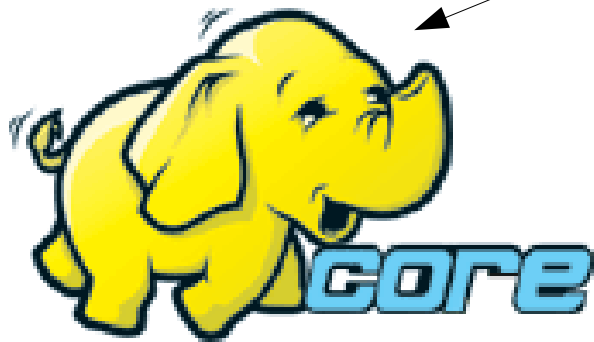
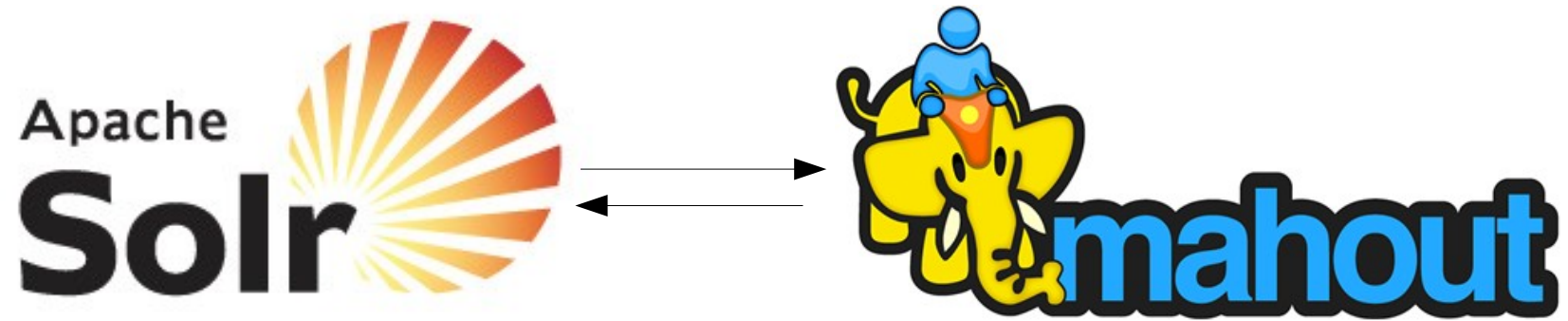




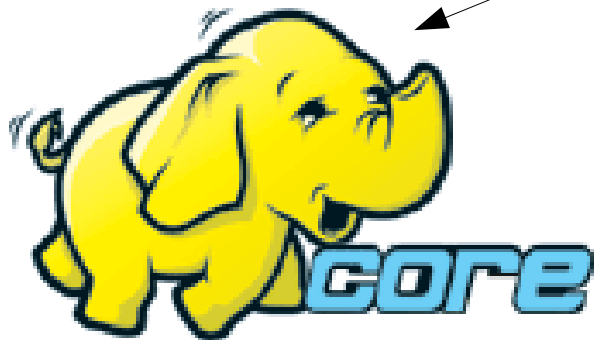
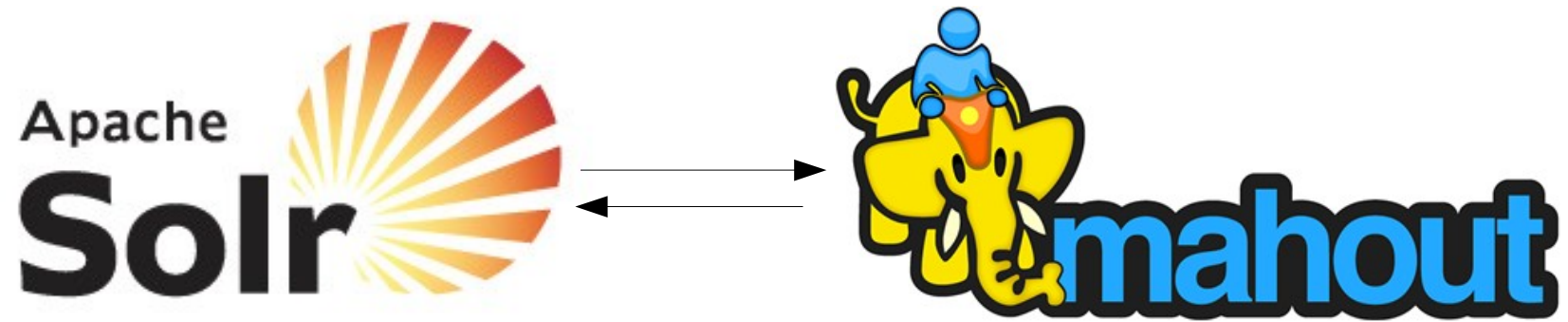




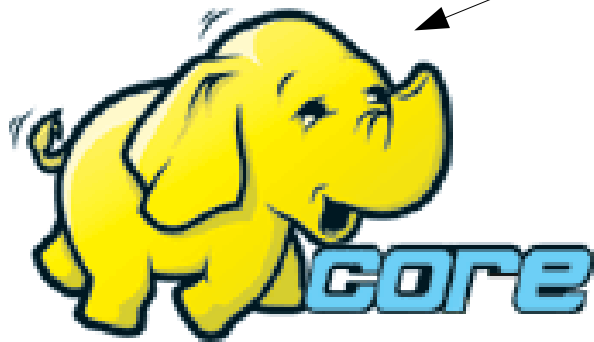
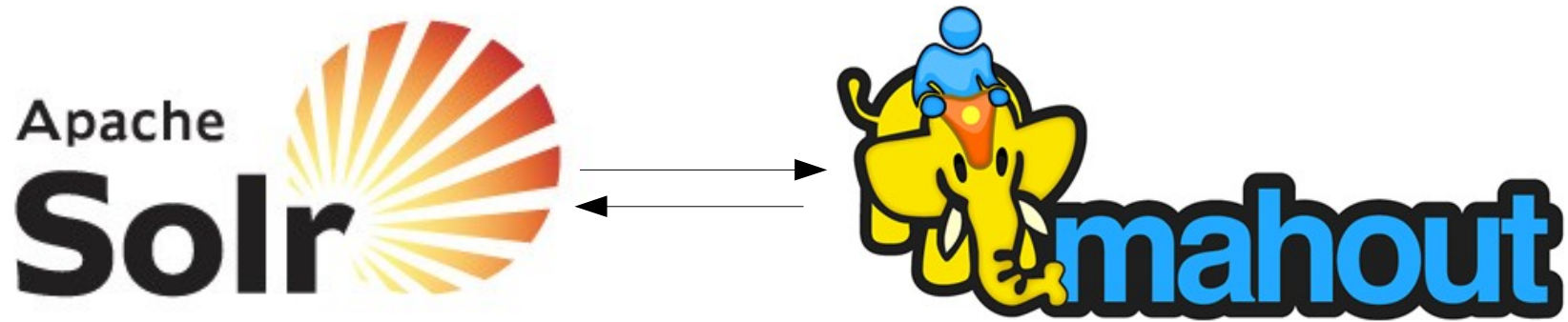












?



### The Colorful History of the Internet and Its Astonishingly Problematic Future

The history of the Internet is a colorful and complex one, and it is a history that is still being written. The Internet is a global network of computers and servers that are connected by a series of cables and wireless signals. It is a network that has revolutionized the way we communicate, work, and play. However, the Internet is also a source of many problems, including privacy concerns, cybercrime, and the digital divide. In this article, we will explore the colorful history of the Internet and its astonishingly problematic future.

**The 1940s, 1950s, and 1960s.** In the late 1940s and 1950s, when computers were still bulky and expensive, they were used for simple tasks like calculating and bookkeeping. These early computers were often found in government offices and universities. The first computer network, called ARPANET, was created in 1969. It was a network of four computers at four different universities in the United States. The network was used for research and communication between the universities. In the 1960s, the network grew to include more than 100 computers at more than 100 universities. The network was used for a variety of purposes, including research, education, and communication. The network was also used for military purposes. The network was used to communicate between the Pentagon and the universities. The network was used to share information about the Soviet Union. The network was used to share information about the Vietnam War. The network was used to share information about the Cuban Missile Crisis. The network was used to share information about the Kennedy assassination. The network was used to share information about the moon landing. The network was used to share information about the Vietnam War. The network was used to share information about the Cuban Missile Crisis. The network was used to share information about the Kennedy assassination. The network was used to share information about the moon landing.

It is important to note that since the 1960s, the number of transistors per unit area has been doubling every one-and-a-half years—thus increasing computing power. This amazing progression of circuit fabrication is called Moore's Law, named after Gordon Moore, a pioneer in the integrated circuit field and founder of the Intel Corporation (ibid.).

What are the future of the Internet? The future of the Internet is uncertain, but it is likely to be a bright one. The Internet is a global network of computers and servers that are connected by a series of cables and wireless signals. It is a network that has revolutionized the way we communicate, work, and play. However, the Internet is also a source of many problems, including privacy concerns, cybercrime, and the digital divide. In this article, we will explore the colorful history of the Internet and its astonishingly problematic future.



# Why?

Why should I waste my time with  
doing stuff for free?



Work on what you want...

when you want.

<http://www.flickr.com/photos/abnelgonzalez/2058764760/>





Share and discuss.

November 16, 2005 [phil h]  
<http://www.flickr.com/photos/hi-phi/64055296>



Learn from the best.



Soft Skills.

<http://www.flickr.com/photos/ajawin/3587215356/>








Make work visible and re-usable.





## Powered By Mahout

Are you using Mahout to do Machine Learning? Care to share?



-  Booz Allen Hamilton uses Mahout's clustering algorithms. See <http://www.slideshare.net/ydn/3-biometric-hadoopsummit2010>
-  Buzzlogic uses Mahout's clustering algorithms to improve ad targeting
-  Kauli, one of Japanese Adnetwork, uses Mahout's clustering to handle clickstream data for predicting audience's interests and intents.
-  Mippin uses Mahout's collaborative filtering engine to recommend news feeds
-  Yahoo! Mail uses Mahout's Frequent Pattern Set Mining. See <http://www.slideshare.net/hadoopusergroup/mail-antispam>

**NOTE: Please add links in alphabetical order. Links here do NOT imply endorsement by Mahout, its committers or the Apache Software Foundation and are for informational purposes only.**



## Powered By Mahout

Are you using Mahout to do Machine Learning? Care to share?

- Booz Allen Hamilton uses Mahout's clustering algorithms. See <http://www.slideshare.net/ydn/3-biometric-hadoopsummit2010>
- Buzzlogic uses Mahout's clustering algorithms to improve ad targeting
-  Kauli, one of Japanese Adnetwork, uses Mahout's clustering to handle clickstream data for predicting audience's interests and intents.
-  Mippin uses Mahout's collaborative filtering engine to recommend news feeds
- Yahoo! Mail uses Mahout's Frequent Pattern Set Mining. See <http://www.slideshare.net/hadoopusergroup/mail-antispam>

**NOTE: Please add links in alphabetical order. Links here do NOT imply endorsement by Mahout, its committers or the Apache Software Foundation and are for informational purposes only.**



**mahout**



as a Service





## Powered By Mahout

Are you using Mahout to do Machine Learning? Care to share?

- Booz Allen Hamilton uses Mahout's clustering algorithms. See <http://www.slideshare.net/ydn/3-biometric-hadoopsummit2010>
- Buzzlogic uses Mahout's clustering algorithms to improve ad targeting
-  Kauli, one of Japanese Adnetwork, uses Mahout's clustering to handle clickstream data for predicting audience's interests and intents.
-  Mippin uses Mahout's collaborative filtering engine to recommend news feeds
- Yahoo! Mail uses Mahout's Frequent Pattern Set Mining. See <http://www.slideshare.net/hadoopusergroup/mail-antispam>

**NOTE: Please add links in alphabetical order. Links here do NOT imply endorsement by Mahout, its committers or the Apache Software Foundation and are for informational purposes only.**





Jumpstart your project with proven code.



January 8, 2008 by dreizehn28  
<http://www.flickr.com/photos/1328/2176949559>



A black and white photograph of two men in a discussion. The man on the left is gesturing with his hands while speaking. The man on the right is listening intently, resting his chin on his hand. The text "Discuss ideas and problems online." is overlaid in white.

Discuss ideas and problems online.

November 16, 2005 [phil h]  
<http://www.flickr.com/photos/hi-phi/64055296>

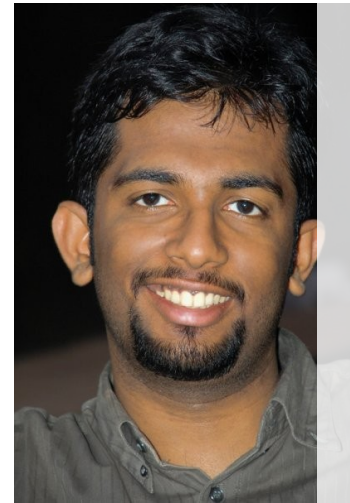




Become a committer.







Sebastian Schelter  
Jake Mannix  
Benson Margulies  
Robin Anil  
David Hall  
AbdelHakim Deneche  
Karl Wettin  
Sean Owen  
Grant Ingersoll  
Otis Gospodnetic  
Drew Farris  
Jeff Eastman  
Ted Dunning  
Isabel Drost



# Become a committer: Of Apache Mahout



*Emeritus:*  
Niranjan Balasubramanian  
Erik Hatcher  
Ozgur Yilmazel  
Dawid Weiss



\*-user@\*.apache.org

\*-dev@\*.apache.org



Image by: Patrick McEvoy

Interest in solving hard problems.

Being part of lively community.

Engineering best practices.

Bug reports, patches, features.

Documentation, code, examples.





Thanks to Tim Lossen et. al for taking amazing pictures of the conf.





# Berlin Buzzwords 2011

Search/ Store/ Scale

May/ June 2011

Thanks to Tim Lossen et. al for taking amazing pictures of the conf.



# October 7th 2010: Hadoop\* Get Together Berlin

- Max Heimeel – Mahout @ TU Berlin.
- Sebastian Schelter – Mahout recommender systems.

**Apache Hadoop Get Together**

**newthinking store**

Tucholskystr. 48

Berlin, Bundesland Berlin [Get Directions](#)

\* UIMA, Hbase, Lucene, Solr, katta, Mahout, CouchDB, pig, Hive, Cassandra, Cascading, JAQL, ... talks welcome as well.



\*-user@\*.apache.org

\*-dev@\*.apache.org



Image by: Patrick McEvoy

Interest in solving hard problems.

Being part of lively community.

Engineering best practices.

Bug reports, patches, features.

Documentation, code, examples.



# Final note

These slides are available for download under the

<http://creativecommons.org/licenses/by-sa/2.0/>

license at

<http://isabel-drost.de/hadoop/slides/dicode.pdf>



Personal Information management.



140 RSS feeds from blogs

Publications at min 6 conferences

NIPS, ICML, SigIR, KDD, ECML, COLT

Put papers to citeulike

4 Google alerts

5 Hashtags on Twitter

50 Mailing lists

Follow maybe 10 really closely.

Get in touch face-to-face

Apache Hadoop Get Together, Apache Dinner,  
Scrumtisch, FOSDEM, FrOSCon, Apache Con  
JUGBB, Devvxx, c-base, Mahout Hackathon,  
Berlin Buzzwords

Following 150 ppl on Twitter





# Information management

- Personal information management.