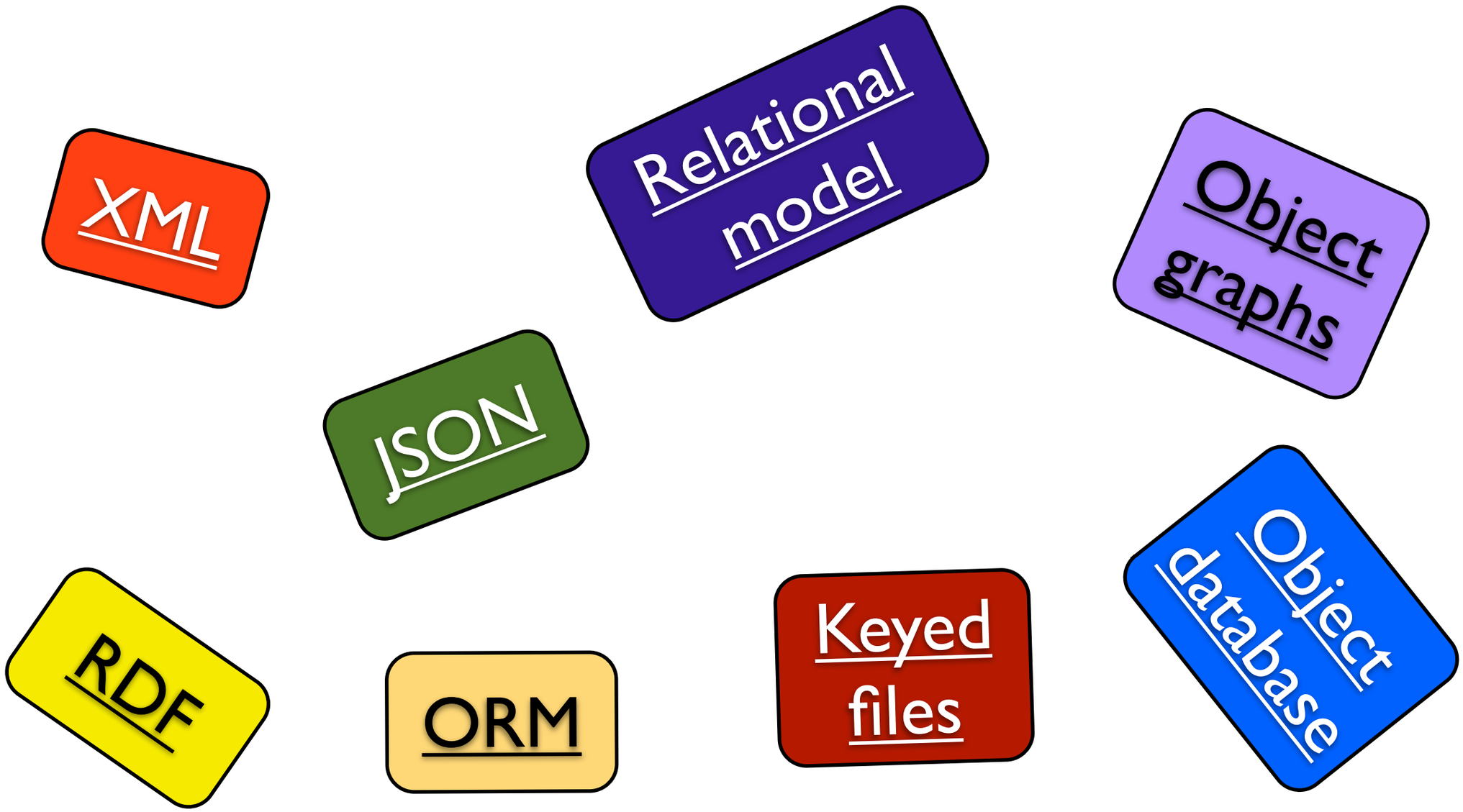
Modern Data Technologies

Software Languages Team
University of Koblenz-Landau
Ralf Lämmel and Andrei Varanovich

General models of data



Modern Applications

Amazon sold a record 158 items every second on Nov. 29

By Douglas Stanglin, USA TODAY

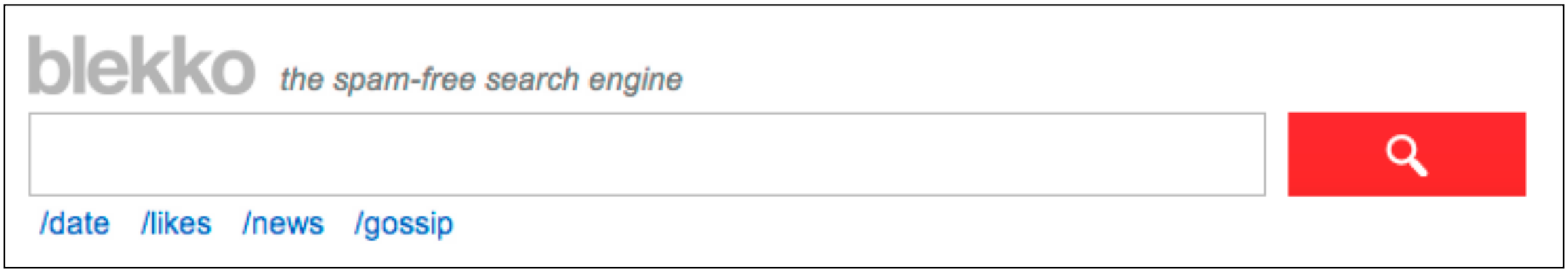
Updated 2010-12-28 3:11 PM

CAPTION By Matt Cardy, Getty Images

Amazon.com has announced that on Nov. 29, its peak holiday shopping day, customers were buying online at a record-breaking 158 items per second, with the newest Kindle its best-selling product ever.

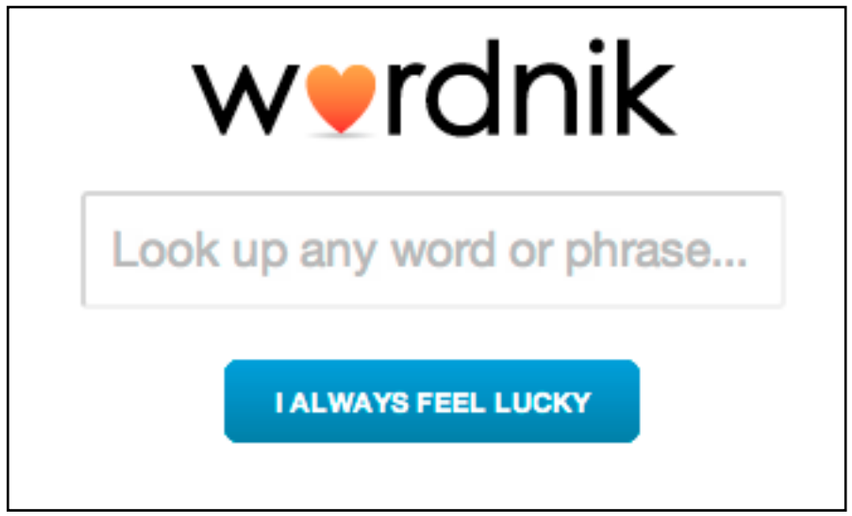
http://content.usatoday.com/communities/ondeadline/post/2010/12/amazon-says-kindle-was-its-best-selling-product-ever-this-year-/1#.UFADLaQtgsd

Google Search I'm Feeling Lucky



blekko's mission: blekko is a consumer facing search engine focused on delivering high quality, relevant, spam-free search results. We believe search should be open, transparent and collaborative. For this reason we combine traditional algorithmic search with the expertise of our users and partners to eliminate spam and deliver results from only the most reputable, best quality sites. This combination creates a highly differentiated editorial search experience that is fundamentally changing search and content discovery online.

http://blekko.com/



Wordnik is a new way to discover meaning. Wordnik shows definitions from multiple sources, so you can see as many different takes on a word's meaning as possible. For more information about the sources of our dictionary definitions, please see the Colophon page.

http://www.wordnik.com/

Ralf Lämmel @reallynotabba 8 Sep

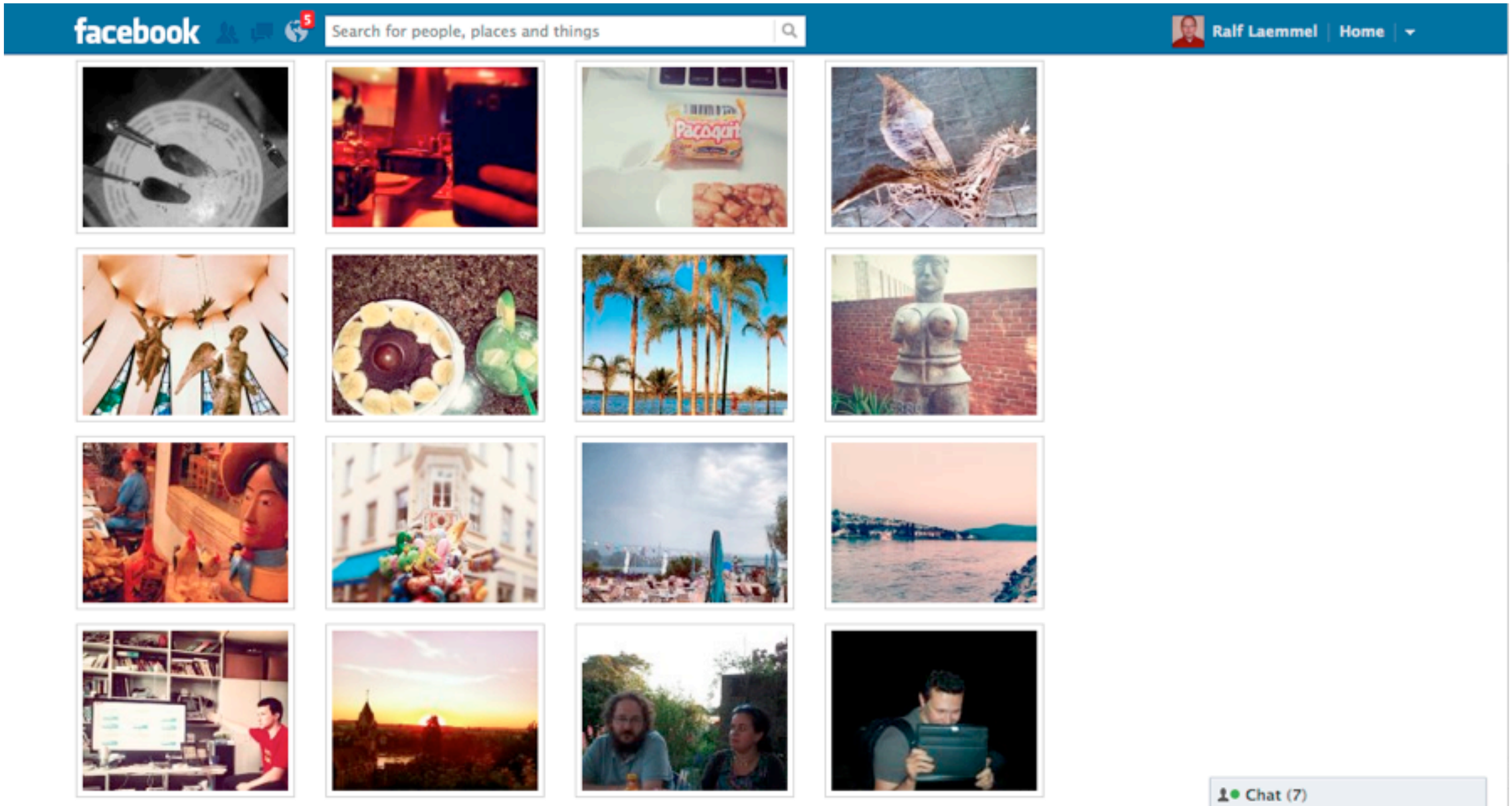
Andrei (@DrGigaBit) & me delivered a "modern web programming" course for #Debeka yesterday; all slideware is here: github.com/avaranovich/we…

Collapse Reply Delete Favorite

2 RETWEETS	1 FAVORITE	

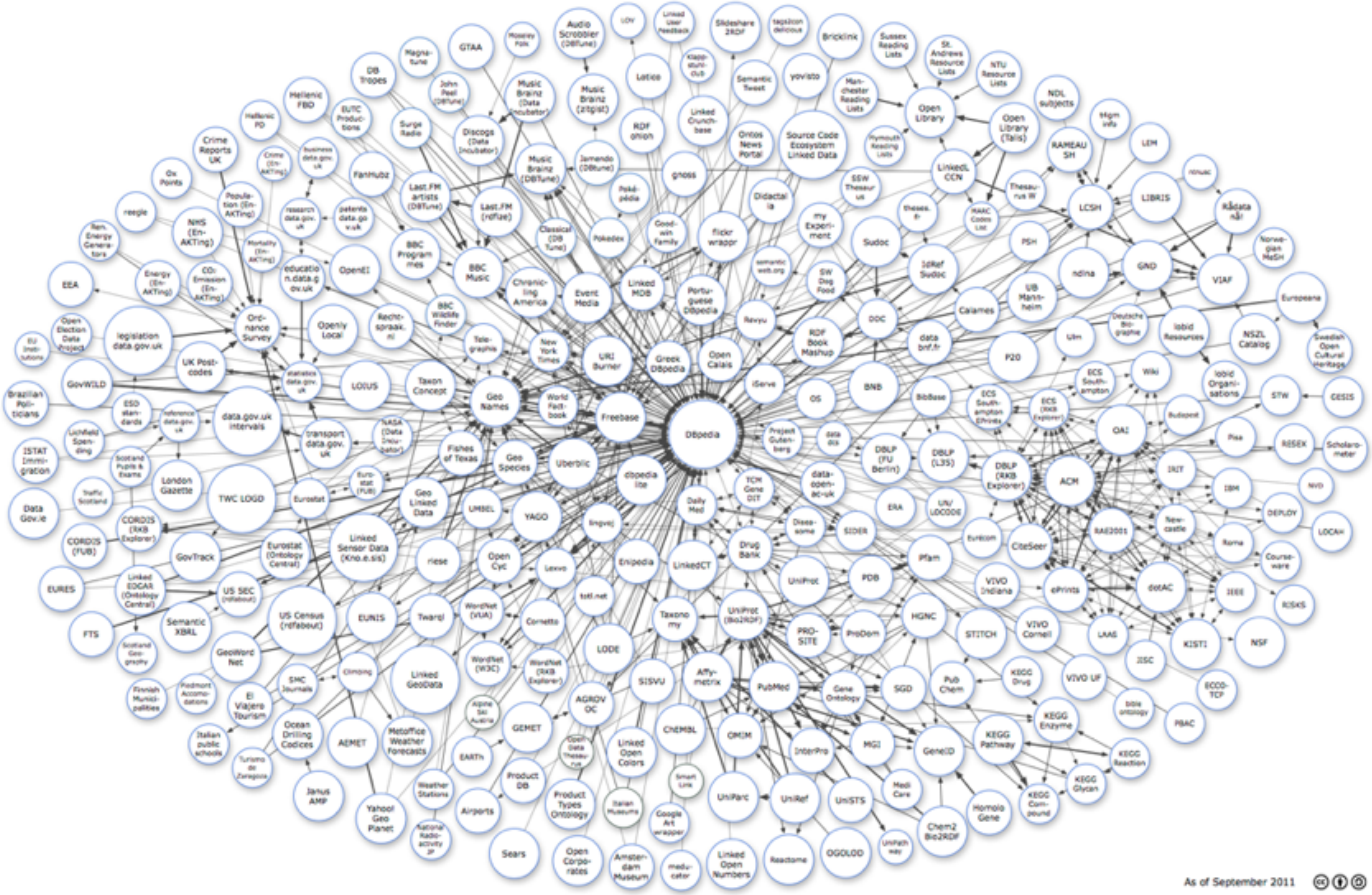
10:36 AM - 8 Sep 12 · Details

https://twitter.com/reallynotabba/status/244353505060614144



https://www.mafiawars.zynga.com/fbconnect?

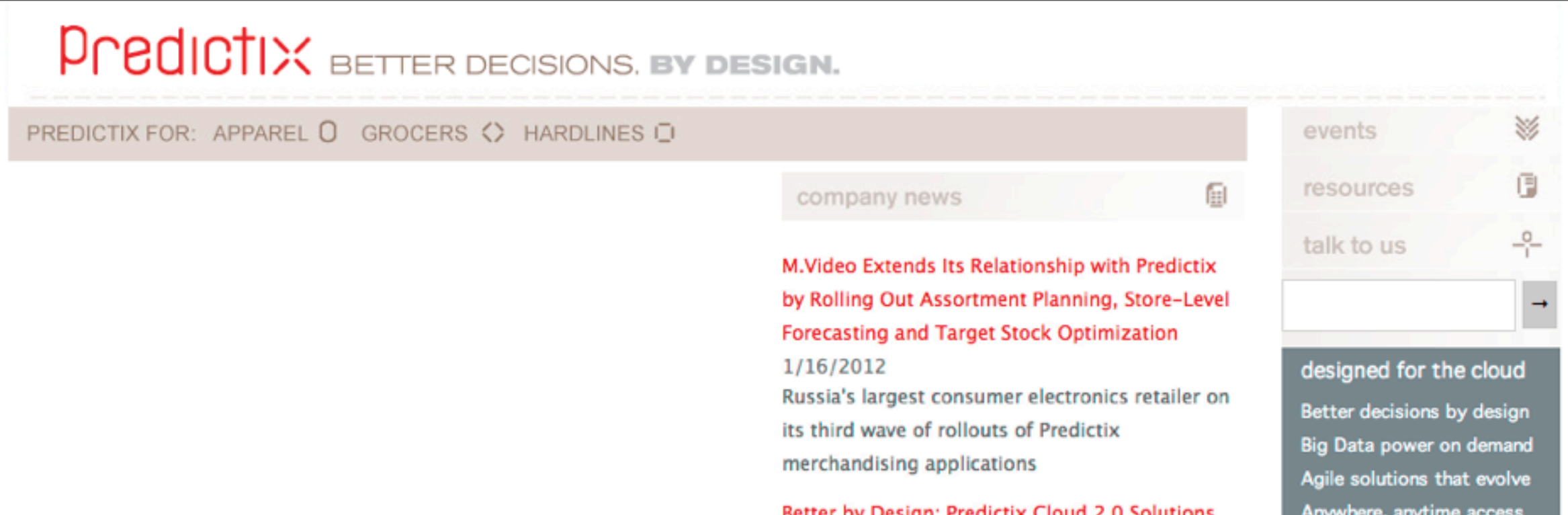
Linked Data



<http://richard.cyganiak.de/2007/10/lod/>

Pentaho Big Data Analytics Center

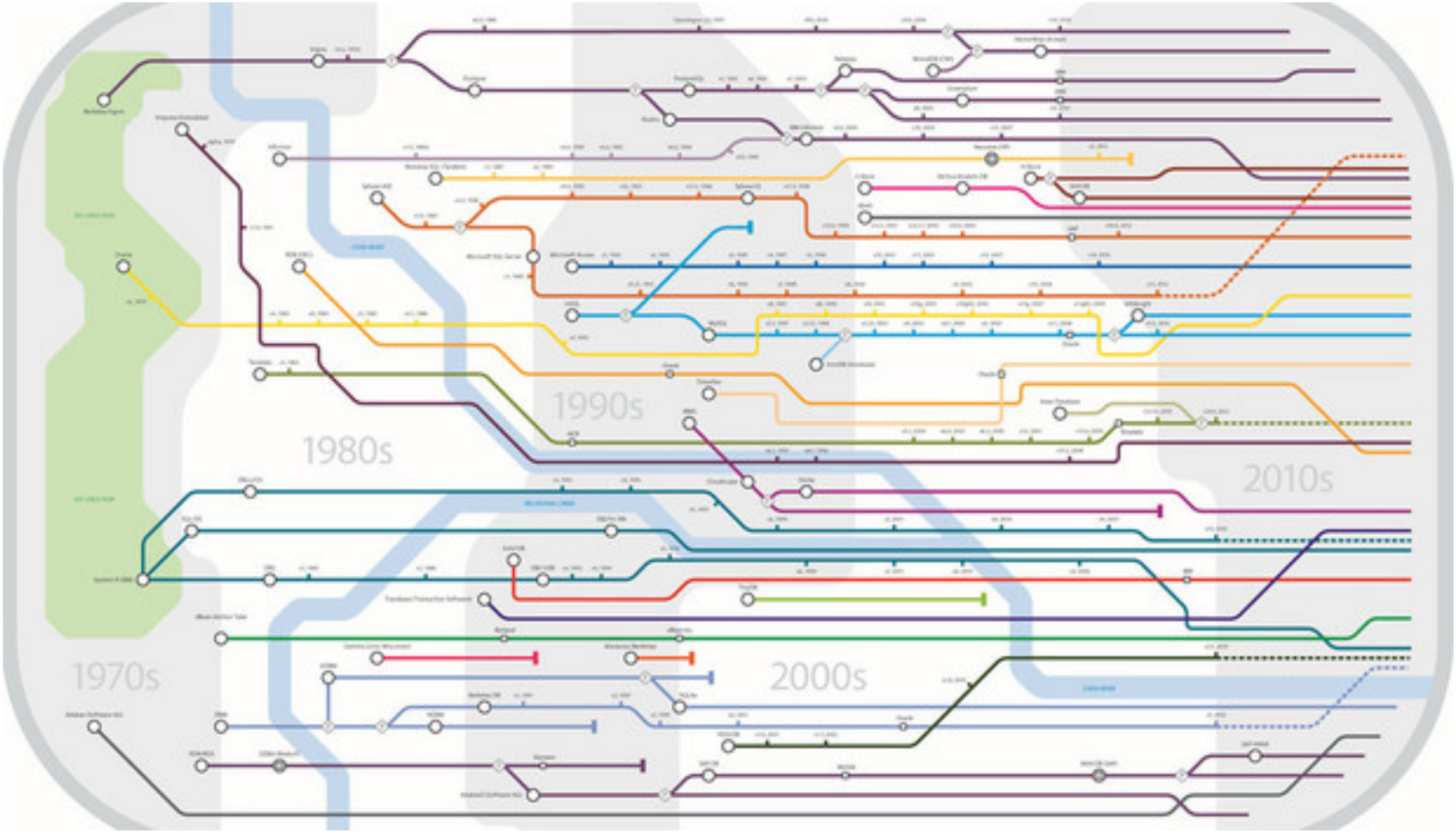
http://www.pentaho.com/big-data/



http://www.predictix.com/

Too many technologies.
Too little time.

Genealogy of RDBMSs



[http://www.searchdatacenter.de/themenbereiche/business-applications/ressourcen/articles/367517/]
Schaubild: Die Genealogie relationaler Datenbank Management Systeme
vom Hasso Plattner Institut, Potsdam.

A

- Apache Accumulo
- Apache Derby

B

- BaseX
- Berkeley DB
- BlackRay

C

- C-Store
- Calpont
- Apache Cassandra
- Cdb (software)
- Comparison of database access
- CouchDB
- CSQL
- CUBRID

D

- Database Deployment Manager
- Database Management Library
- Dataphor
- Db4o
- Dbm
- User:Dodilp/sandbox
- User:Dodilp/sandbox2
- Drizzle (database server)
- Druid (database designer)

E

- Eloquera
- Eloquera database
- EnterpriseDB
- EXist

F

- Firebird (database server)
- FleetDB
- FlockDB
- Fyracle

G

- Gadfly (database)
- GCfilms
- GCstar
- Gizzard (Scala framework)
- GNOME-DB
- GT.M

H

- H-Store
- H2 (DBMS)
- Habanero.NET
- **HBase**
- Hibari (database)
- HSQLDB
- Hypertable

I

- *InfiniDB*
- Infobright
- Ingres (database)

J

- JGnash

K

- Kexi

L

- LucidDB

M

- *Maatkit*
- MariaDB
- Membase
- MemcacheDB
- Metakit
- Mnesia
- MonetDB
- **MongoDB**
- Mozilla Raindrop
- MSQL-JDBC
- MySQL

N

- Neo (object-relational toolset)
- **Neo4j**
- NeoDatis ODB

O

- OpenOffice Base
- OpenQM
- OrientDB

P

- Perst
- PostgreSQL

R

- Redis
- Rel (DBMS)
- RRDtool

S

- Sedna (database)
- Shapado
- SmallSQL
- Sones GraphDB
- SQLite
- Strozzi NoSQL (RDBMS)

T

- Tarantool
- Terrastore
- TreapDB
- TxtSQL

V

- Virtuoso Universal Server
- VoltDB

W

- Wakanda (software)

Z

- Zope Object Database

[http://en.wikipedia.org/wiki/Category:**Free**_database_management_systems] 12 Sep 2012

A

- Amazon DynamoDB

B

- Basho Technologies

C

- Apache Cassandra
- Clusterpoint
- Column (data store)
- Column family
- CouchDB

D

- Db4o
- User:Dodilp/sandbox
- User:Dodilp/sandbox2
- Dynamo (storage system)

E

- Eloquera
- Eloquera database

F

- FleetDB
- Fluidinfo

H

- Hazelcast
- Hibari (database)

I

- IBM Lotus Notes
- IBM WebSphere eXtreme Scale
- InfiniteGraph

K

- Keyspace (distributed data store)

L

- LevelDB

M

- MarkLogic
- Membase
- Meronymy SPARQL Database Server
- **MongoDB**
- MongoHQ
- MultiValue

O

- ObjectDB
- OrientDB

R

- Rasdaman
- Redis
- **Riak**

S

- Standard column family
- Super column family

T

- Tarantool
- Terrastore

V

- Virtuoso Universal Server

W

- Wakanda (software)

X

- Xeround

[http://en.wikipedia.org/wiki/Category:**NoSQL**] 12 Sep 2012

6

- 6fusion

A

- Abiquo Enterprise Edition
- Amazon DynamoDB
- Amazon Elastic Block Store
- Amazon Elastic Compute Cloud
- Amazon Machine Image
- Amazon Route 53
- Amazon Simple Email Service
- Apache Drill
- **Apache Hadoop**
- Apache Hama

B

- Basho Technologies
- BigCouch
- Bitnami
- Wikipedia talk:Articles for creation/On demand business process management

C

- C12G
- Calpont
- Cascading
- Cloud business process management
- Cloud.bg
- Cloud.com
- Cloudant
- Cloudera
- Cloudkick
- CloudSigma
- Couchbase

D

- User:Dodilp/sandbox

E

- ElasticHosts
- Enlight cloud
- Enomaly Inc
- EnStratus
- Eucalyptus (computing)

F

- FlexiScale
- Fujitsu Global Cloud Platform

G

- GoGrid
- GreenButton
- GreenQloud

H

- HP Cloud Services

I

- IBM BlueWorks Live
- IBM cloud computing
- Iland
- *InfiniDB*

K

- Kaavo

L

- LingCloud
- Linode

M

- MapR
- Media Temple
- Music Kickup

N

- Nimbula
- Nimbus (cloud computing)

O

- OnApp
- OpenNebula
- OpenQRM
- OpenStack
- OpSource

R

- Rackspace Cloud
- **Riak**
- RightScale

S

- Scalr
- Sector/Sphere
- Skytap
- Sqrrl
- Sun Cloud

T

- User:TechForGood/sandbox
- TurnKey Linux Virtual Appliance Library

V

- Virtual Internet

W

- Wakame-vdc

[http://en.wikipedia.org/wiki/Category:**Cloud_infrastructure**] 12 Sep 2012

Guiding questions

- How to do a course on data technologies?
- What data technologies to select?
- What characteristics to exercise?

Background check regarding audience

- Are you a (near-to) ‘developer’?
- Are you a (near-to) ‘DBA’?

- Are you accustomed reading ***code*** (in Java, ...)?
- Are you accustomed reading ***schemas & queries***?
- Do you want to try out things during the course?
- Any expectations that you want make heard now?

So then, how to do a course on data technologies?

Focus on '***modern***' data technologies!

- **LINQ**: Language INtegrated Queries
- **NoSQL**: One representative per NoSQL DB type
- **MapReduce**-style data processing
- Distributed file systems

In computing, **NoSQL** (mostly interpreted as "not only SQL"[...]) is a broad class of database management systems identified by its non-adherence to the widely used relational database management system model; that is, NoSQL databases are not primarily built on tables, and as a result, generally do not use SQL for data manipulation.

[http://en.wikipedia.org/wiki/NoSQL] 12 Sep 2012

Characteristics of a NoSQL database

- It does not use SQL as its query language.
- It may not give full ACID guarantees.
- It has a distributed, fault-tolerant architecture.
- It is highly optimized and specialized.
- It is highly scalable (horizontally).

[http://en.wikipedia.org/wiki/NoSQL] 12 Sep 2012

Course picks 1 representative per NoSQL DB type

- Works with the typical scenarios
- Best-in class community buzz / user adoption
- Strong community ecosystem
- Free
- List of systems
  - Key-value databases (Riak)
  - Document-oriented databases (MongoDB)
  - Graph databases (Neo4J)
  - Column-oriented databases (HBase)

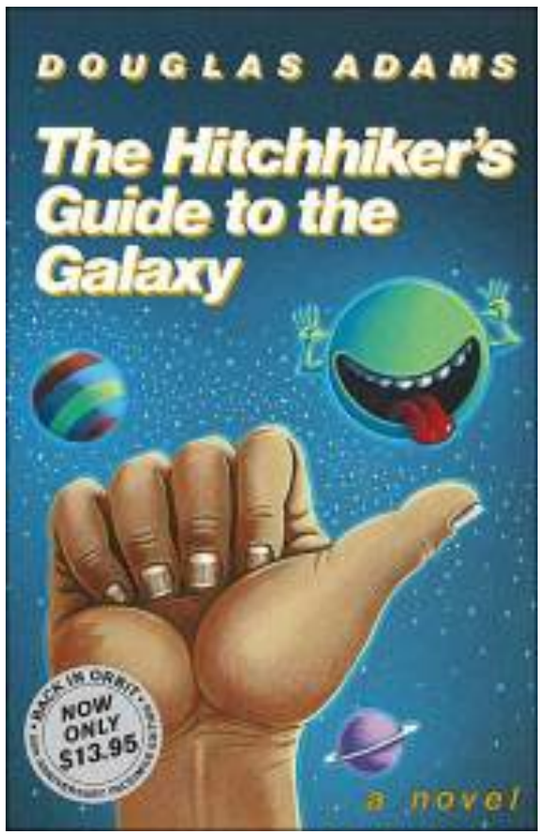
MapReduce is a programming model for processing large data sets, and the name of an implementation of the model by Google. MapReduce is typically used to do distributed computing on clusters of computers. [...] The model is inspired by the map and reduce functions commonly used in functional programming. [...]. MapReduce libraries have been written in many programming languages. A popular free implementation is Apache Hadoop.

[http://en.wikipedia.org/wiki/MapReduce] 12 Sep 2012

Use ***101companies*** throughout the course!

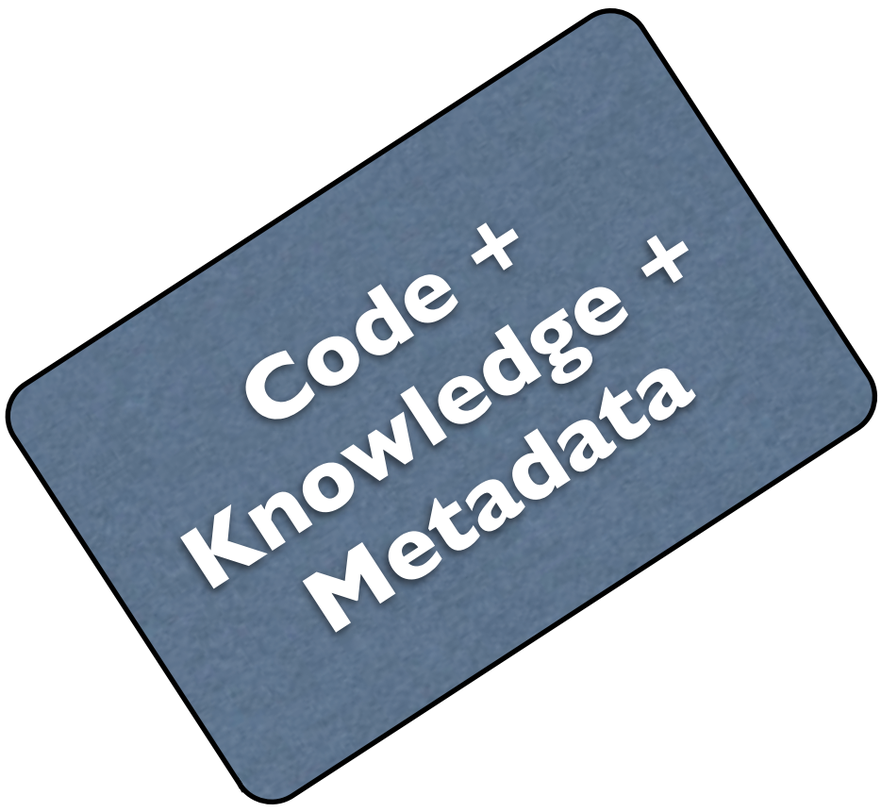
http://101companies.org

The Hitchhiker's Guide to the ***Software*** Galaxy



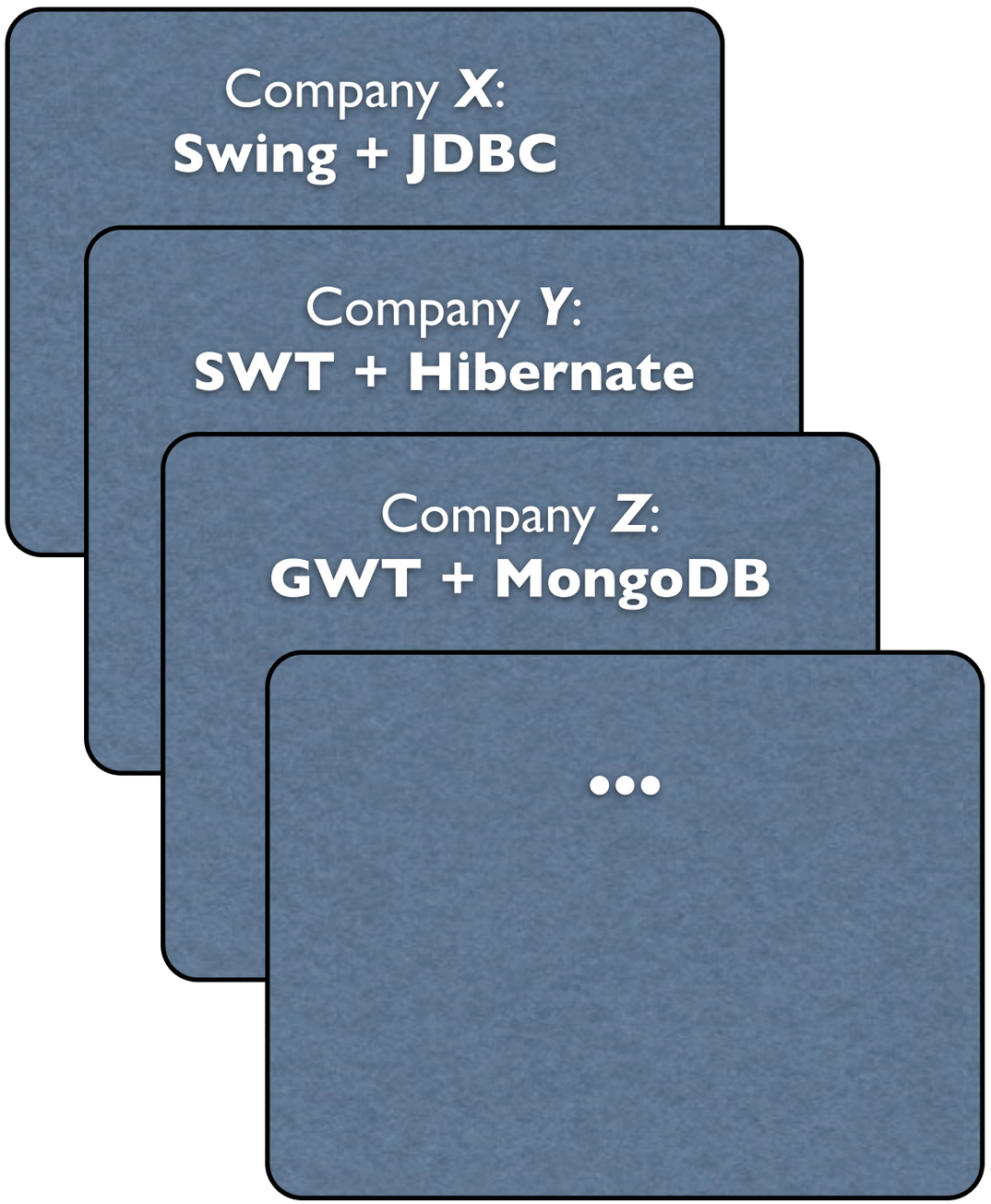
Wannabe Wikipedia for Software Developers. :-)

What's 101companies?



A **community project** aiming at a **knowledge base** about software **technologies and languages** based on implementations of a human-resources management system.

Why is it called “101companies”?

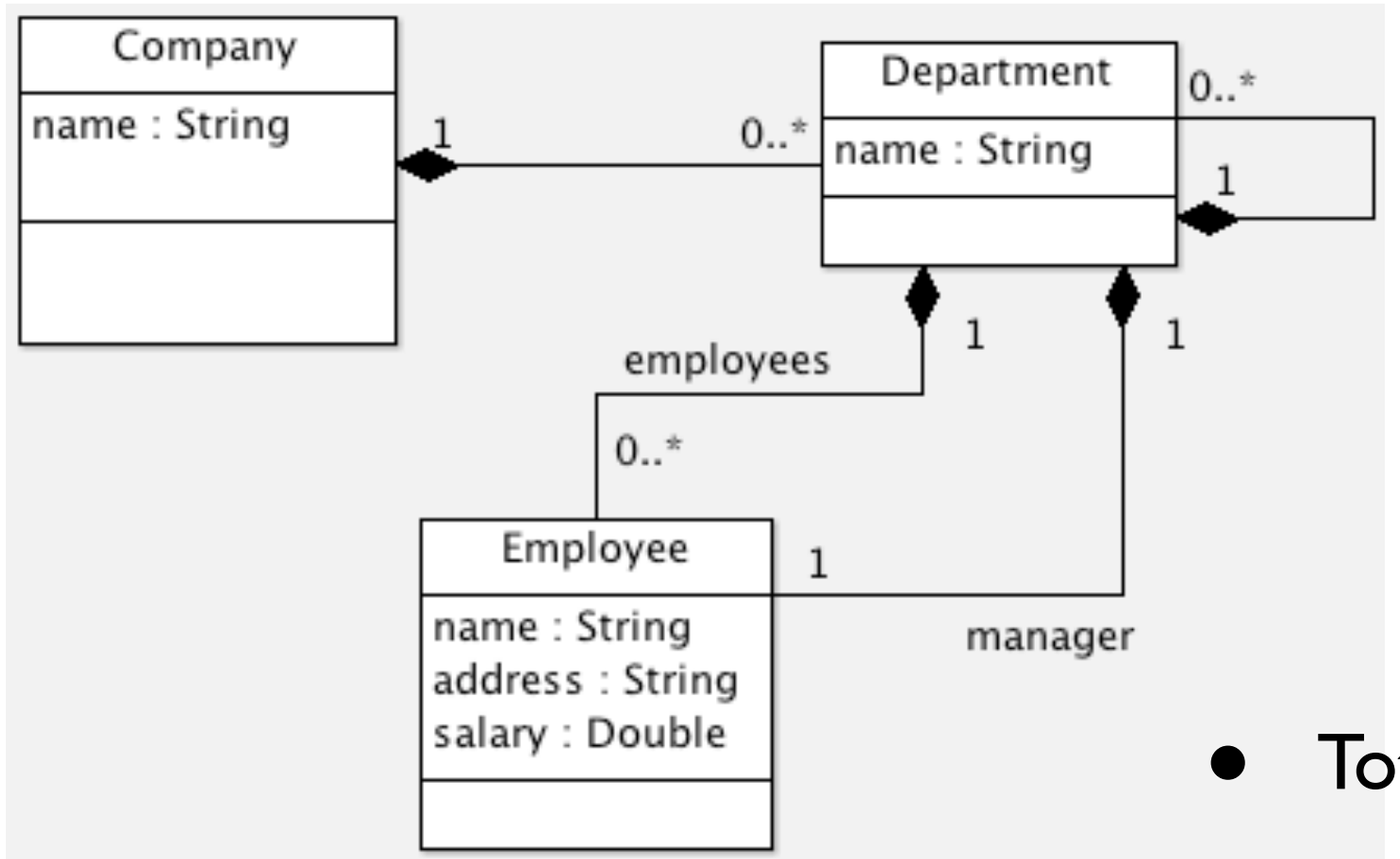


“101 ways of building a HRMS.”

or

“Building a HRMS for 101 companies.”

A Human Resources Management System



- Total salaries
- ~~Increase salaries~~
- Cut salaries
- Edit employee data
- Import / export company data

Demo

A *101companies* implementation using ***HTML5*** and local storage

company.html

file://localhost/Users/laemmel/projects/101repo/contributions/html5local/company.html

101Companies HTML5 implementation with local storage

Company

Name: Meganalysis save

Departments:

Research

Development

select

Total: 7226.765625 cut

Load company from local storage

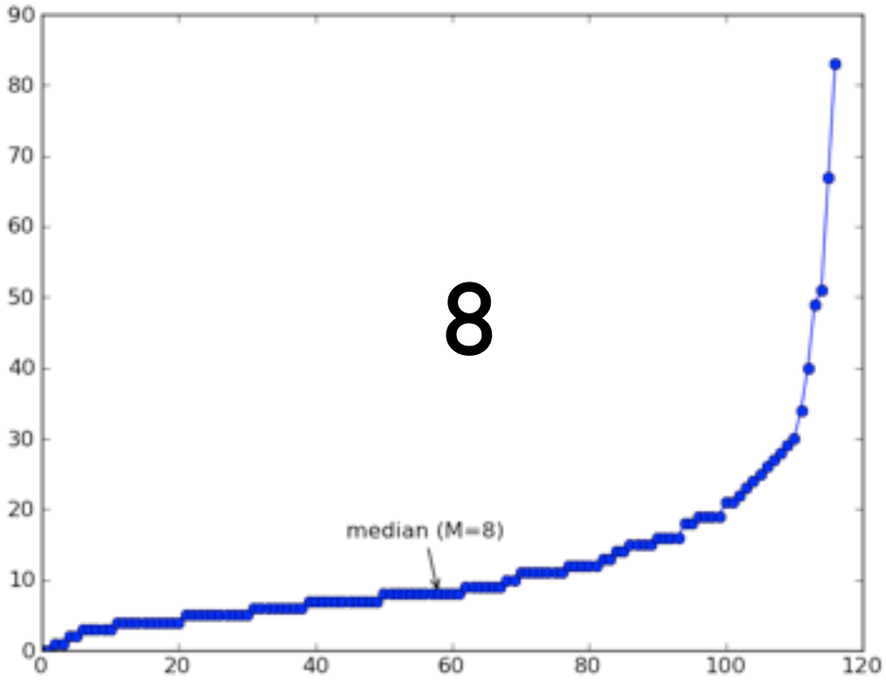
```
JSON.parse(localStorage.getItem('company')
```

Save company to local storage

```
localStorage.setItem('company', JSON.stringify(company))
```

[contributions/html5local/company.js]

The 101companies **Repository**



#Files per implementation	Technologies
Languages	LOC per implementation

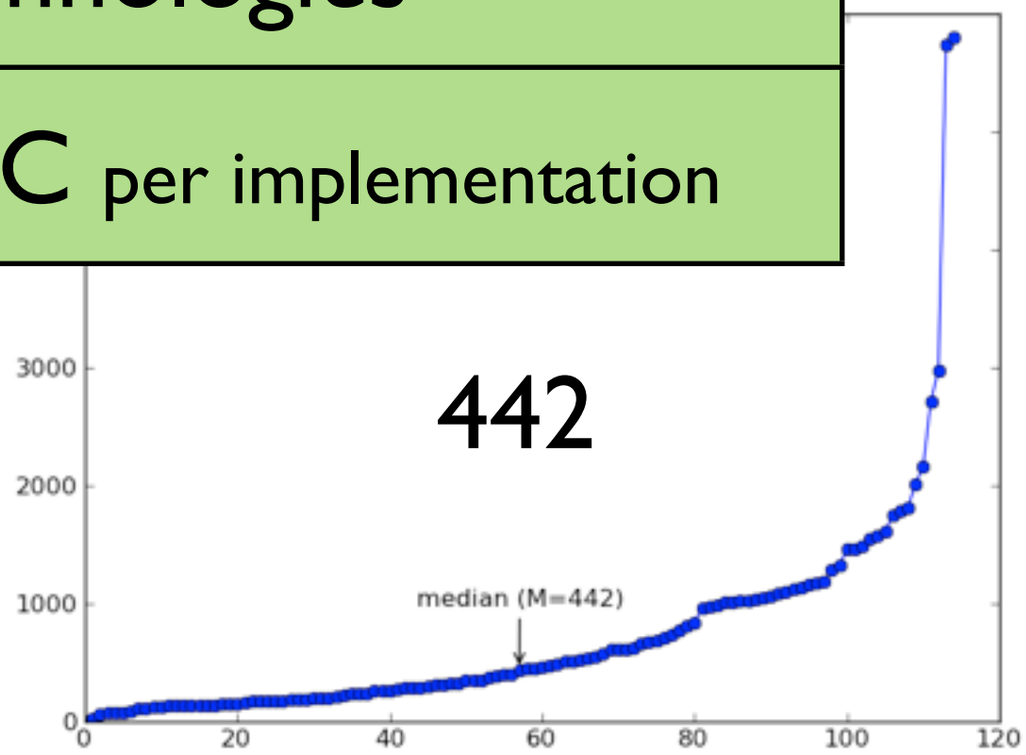


Table of contents

Thanks for your interest.
Questions? Comments?