Structure

- Garlik – the background and origins
- Going down the Semantic technology route
- Raising VC investment for a Semantic Tech company
- The Challenge of the Consumer
- Challenges of building the platform
- The Challenge of the Semantic Web - Garlik Labs
- Lessons Learnt
My Journey to Garlik
Our CEOs Journey to Semantic Technology

CIO of an online bank with 4m customers

Conducted a 6 month review of emerging technologies

$15m AKT Semantic Web research project
Making the Final Decision

Did we really believe...

- That the Web is evolving from Documents to Data?
- That this evolution would happen over 5 years?
- That the solid background research had been done?
- That smart people are involved?
- That we could build a robust, scalable, secure platform?
- That this will give us a long term competitive advantage?
The Management Team

Tom Ilube, Chief Executive Officer
Previously, Tom was Chief Information Officer & Executive Committee member of Egg plc, the world's largest pure online bank. In 1996 Tom founded a software company where he was Chief Executive for 6 years. Tom's career spans 20 years with companies including Goldman Sachs, PwC & London Stock Exchange.

Mike Harris, Executive Chairman
As founding Chief Executive Officer of Egg plc, Mike took Egg from concept to a £1bn public company within 3 years. Previously, Mike was Chief Executive of Mercury Communications & board director of Cable & Wireless and he also created First Direct, the world's first telephone banking operation.

Nigel Shadbolt, Chief Technology Officer
Nigel is a Professor in the School of Electronics & Computer Science at University of Southampton, with an international reputation in knowledge management and artificial intelligence. His current research focus is the Semantic Web. He was President of the British Computer Society in its 50th Anniversary Year.
Wendy Hall is Chair of the Garlik Advisory Board. She is Professor of Computer Science at University of Southampton and was Head of the School of Electronics and Computer Science. Wendy is a member of the Prime Minister's Council for Science and Technology and Senior Vice President of the Royal Academy of Engineering.

Simon Davis is a Visiting Fellow at the London School of Economics and a key author of the LSE report on ID Cards. As founder & director of watchdog group Privacy International, he is one of the world's leading figures in privacy & data protection. Simon's work in this area, consumer rights, policy & technology analysis spans 20 years & 30 countries.

Sir Tim Berners-Lee, KBE, OM inventor of the World Wide Web, is Director of the World Wide Web Consortium (W3C), Senior Researcher at Massachusetts Institute of Technology (MIT) and Professor of Computer Science at University of Southampton.

Daniel Cooper is a well-known specialist in European privacy & data protection matters. He is a regular contributor to World Data Protection Report & member of the Law Society's data protection group. Dan is Special Counsel for Covington & Burling, CASRO's General Counsel and is on Privacy International's Advisory Board.
The Opportunity

The internet is part of everyday life

- 64% UK adults online
- 19 MM have broadband access

Web risks becoming a victim of its own success

- 25 billion web pages & 100 billion by 2010
- Will we be able to find what or who we are looking for?

Private information exposed for others to use or abuse

- Information you share or others share about you
- Company or government data

Big – and growing – issue for consumers

- 39% concerned about their privacy on the internet
- 35% don’t trust companies with the personal data they collect
Every 12 minutes, an attack on your home PC

Thousands exposed to danger of identity theft

Fear of crime stops people going online

How to protect yourself from identity fraud

My identity was stolen 52 times

If you've ever thought that you've come to all together to sound too clever to be had.

The chatroom charmers who will steal more than your heart

HSBC and Tesco in new identity fraud scares

Tube boss swindled of £100k

Threat to your PC as you sleep
This is just what HMG has on us - not known for their ability to keep this data secure
A selection of sources on the Web that hold information on us
A selection of sources on the Web that hold information on us

Flickr’s “ecology”

Diagram by Bryce Glass, http://soldierant.net
A selection of sources on the Web that hold information on us.

Map of Blogosphere from Matthew Hurst
Garlik’s Purpose

we give individuals and their families real power over the use of their personal information in the digital world
Garlik Today

A great team
Management
Advisors
Technologists
Sales, marketing, operations

Blue chip investors
$18.5m from 3i, Doughty Hanson
$5m from Noble Capital

Growing customer base
65,000 + in the UK and growing fast
US Product about to launch

A growing reputation
Red Herring Top 100 Europe
The Register UK #1 Web 2.0 company
Davos 2008 World Economic Forum Technology Pioneer

Breakthrough technology
Large scale semantic technology platform
Structure

- Garlik – the background and origins
- Going down the Semantic technology route
- Raising VC investment for a Semantic Tech company
- The Challenge of the Consumer
- Challenges of building the platform
- The Challenge of the Semantic Web - Garlik Labs
- Lessons Learnt
Business Criteria in Technology Selection

We had

- Heterogeneous Data
- Ragged Data
- New Data sources arising all the time

We Wanted

- Opportunistic reuse
- Large-scale
- People centric
- Rich semantics
- Provenance
On the order of $10^8$ users - used in every country on Earth

A tiny percentage is “trained” in any way

Billions indexed web resources (text) in Google etc

Massively distributed and open-a set of protocols and languages driven by a strong standards approach

The next stage is to move beyond a web of documents to a web of data - this is our current vision of the Semantic Web

**The Semantic Web Revisited**

Nigel Shadbolt is a professor of artificial intelligence in the School of Electronics and Computer Science at Southampton University. Contact him at nrs@ecs.soton.ac.uk.

Tim Berners-Lee is the director of the World Wide Web Consortium, a senior researcher at the Massachusetts Institute of Technology’s Computer Science and Artificial Intelligence Laboratory, and a professor of computer science in the Department of Electronics and Computer Science at Southampton University. Contact him at timb@w3.org.

Wendy Hall is a professor of computer science in the School of Electronics and Computer Science at Southampton University. Contact her at wlb@ecs.soton.ac.uk.
Technical Architecture

- Unstructured data
- Semi-structured data
- WWW
- Structured data
- URLs
- Harvested Pages
- NUTCH
- Ranked Searchable Segment
- Natural Language Engineering
- NPS
- RDS
- JXT smush
- Front End

JXT store
Garlik’s Semantic Platform

- Ontologies
- Semantically informed search and harvesting
- NLP capabilities
- Smushing
- Large-scale clustered RDF triple store
- SPARQL data access standard
Structure

- Garlik – the background and origins
- Going down the Semantic technology route
- Raising VC investment for a Semantic Tech company
- The Challenge of the Consumer
- Challenges of building the platform
- The Challenge of the Semantic Web - Garlik Labs
- Lessons Learnt
Pitching Semantic Technology to the VCs
Structure

- Garlik – the background and origins
- Going down the Semantic technology route
- Raising VC investment for a Semantic Tech company
- The Challenge of the Consumer
- Challenges of building the platform
- The Challenge of the Semantic Web - Garlik Labs
- Lessons Learnt
DataPatrol

A simple & effective way to monitor what is out there
Your garlik credit report has been compiled by Checkmyfile using extracts from the information held at Creditreform, alongside additional data held by Checkmyfile.

This page shows you at a glance how each type of personal data impacts upon your credit rating. It also helps you to understand how the various types of information can be viewed in terms of fraud risk.

**Other Reports**
- [05.11.07 report](#)
- [19.06.07 report](#)

To compare 2 or more reports, select them then [click here](#).

**Your overall rating**

**Your credit score**

Your credit score is [4 stars]. This gives you a 4 star rating.

Your credit score - within a range of 0 - 1000, where 0 is very bad and 1000 is excellent - is a statistical measure of the likelihood of you paying on time, so is used by lenders to assess whether to lend you money. A high score indicates you are a low credit risk, conversely a low score indicates you are a high credit risk. The average credit score is 750.

Your credit rating - within a range of 1-5 stars, where 1 is very bad and 5 is excellent - is a more broad measure of your creditworthiness. It is used by lenders to help customer service staff understand what sort of payer you are, and to divert calls and enquiries to staff that specialise in different types of customer. It is also used to help match financial products with typical customer needs. The average credit rating is 3 stars.

**Report specifics**

**Accounts**

Your worst payment status is 1 month late.

Information about how you pay your bills is shared by around 260 banks, credit card companies, mobile phone providers, home shopping companies, mail order firms and mortgage companies. Regular and timely payments of bills will help maintain a good credit rating.

It makes sense to keep a separate credit card for use in petrol stations and restaurants, where the risk of ‘skimming’ card details is higher — you’ll be able to spot rogue transactions much easier. Close any credit cards and store card accounts that you don’t need any more — if someone stole details of any dormant cards you may have, you probably wouldn’t notice for some time. If you can, use lenders who provide online statements rather than paper ones. Mail interception is one of the most common methods of identity fraud, so the fewer statements you receive by post, the less the risk.
## Court proceedings

<table>
<thead>
<tr>
<th>Type</th>
<th>Status</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attachment of Earnings Order</td>
<td>No record held on credit file</td>
<td>1</td>
</tr>
<tr>
<td>Bankruptcy (Channel Islands)</td>
<td>Checked and clear – no record found</td>
<td>1</td>
</tr>
<tr>
<td>Bankruptcy (England and Wales)</td>
<td>Checked and clear – no record found</td>
<td>1</td>
</tr>
<tr>
<td>Bankruptcy (Isle of Man)</td>
<td>Checked and clear – no record found</td>
<td>1</td>
</tr>
<tr>
<td>Bankruptcy (Northorn Ireland)</td>
<td>Checked and clear – no record found</td>
<td>1</td>
</tr>
<tr>
<td>Charging Order</td>
<td>No record held on credit file</td>
<td>1</td>
</tr>
<tr>
<td>Child Support Order ‘CSA order’</td>
<td>No record held on credit file</td>
<td>1</td>
</tr>
<tr>
<td>County Court Judgment ‘CCJ’</td>
<td>Checked and clear – no record found</td>
<td>1</td>
</tr>
<tr>
<td>County Court Summonses</td>
<td>No database currently available</td>
<td>1</td>
</tr>
<tr>
<td>Debt Management Arrangement</td>
<td>No database currently available</td>
<td>1</td>
</tr>
<tr>
<td>Deed of Arrangement</td>
<td>Checked and clear – no record found</td>
<td>1</td>
</tr>
<tr>
<td>Garnishee Order</td>
<td>No record held on credit file</td>
<td>1</td>
</tr>
<tr>
<td>High Court Judgment</td>
<td>No record held on credit file</td>
<td>1</td>
</tr>
<tr>
<td>High Court Writ</td>
<td>No database currently available</td>
<td>1</td>
</tr>
<tr>
<td>Individual Voluntary Arrangement</td>
<td>Checked and clear – no record found</td>
<td>1</td>
</tr>
<tr>
<td>Liability Order</td>
<td>No record held on credit file</td>
<td>1</td>
</tr>
<tr>
<td>Magistrates’ order</td>
<td>No record held on credit file</td>
<td>1</td>
</tr>
<tr>
<td>Receiver appointed under Law of Property Act 1925 ‘LPA Receiver’</td>
<td>No record held on credit file</td>
<td>1</td>
</tr>
<tr>
<td>Scottish Decree</td>
<td>Checked and clear – no record found</td>
<td>1</td>
</tr>
<tr>
<td>Sequestration (Scotland)</td>
<td>Checked and clear – no record found</td>
<td>1</td>
</tr>
<tr>
<td>Statutory Demand</td>
<td>No record held on credit file</td>
<td>1</td>
</tr>
<tr>
<td>Summary Court Judgment (Isle of Man)</td>
<td>Checked and clear – no record found</td>
<td>1</td>
</tr>
<tr>
<td>Trust Deed</td>
<td>Checked and clear – no record found</td>
<td>1</td>
</tr>
</tbody>
</table>
Turning information into knowledge: researchers unite to develop new technologies

March 22, 2000

Some of the UK’s leading scientists will be joining forces with major industrial companies in a multi-million pound project to develop computer methods and software for the management and manipulation of knowledge in the information age.

The University of Southampton is the lead partner in an Interdisciplinary Research Collaboration (IRC) to be awarded major funding by the Engineering and Physical Sciences Research Council (EPSRC). The project, known as AKT (Advanced Knowledge Technologies), will be involve the Universities of Southampton, Aberdeen, Edinburgh, Sheffield and the Open University. Its director is Professor Nigel Shadbolt of the University of Southampton’s Department of Electronics and Computer Science.

South Africans
What kind of country do we want SA to be in 2010?

www.youtube.com/watch

Point Of Sale Technology
Get a free trial of our software & see how we simplify your business!

www.AccurPOS.co.uk

MBA New Media Technology
Study at MODUL University Vienna 2-Year Executive Degree in Austria

www.modul.ac.at/mba/newmed

Green IT solutions
We will find a “Green Solution” for your ICT Processes and Problems

www.MNCD.co.uk

The researchers in the AKT consortium-Advanced Knowledge Technologies- include specialists in computer science, artificial intelligence, psychology, linguistics, multimedia and internet technology. Knowledge is clearly the most creative force in the modern economy. Its exploitation has revolutionised communications, services, industries-everything, from the food we eat, to the medicine we take,’ explains Professor Shadbolt. ‘With developments in telecommunications and the Internet, we have faster and better access to more and more of it.’

‘Despite the importance of knowledge, one of the problems we face is that well-known 21st century malady “information overload”.

Information is plied upon us, in databases, on the World Wide Web, and getting the right knowledge-useful information applied to solve a specific problem is like looking for a needle in a haystack.’

‘If knowledge is to be an asset in the knowledge economy, it has to be managed efficiently. The ability to use this knowledge successfully will be the chief source of the UK’s competitive advantage in a global economy.’

The AKT consortium will use its funds to undertake fundamental research to investigate the knowledge ‘life-cycle’, and developing-edge technology to support it. The research will focus on six core challenges:
Data Patrol CPP

Monitor/Alert
Help
Insure
### Items needing your attention

<table>
<thead>
<tr>
<th>Name</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kate Bolton</td>
<td><a href="http://www.myspace.com/abrandoffree">http://www.myspace.com/abrandoffree</a></td>
</tr>
<tr>
<td>Kate Bolton</td>
<td><a href="http://www.zoominfo.com/people/Bolton_Kathleen_(">http://www.zoominfo.com/people/Bolton_Kathleen_(</a>...)</td>
</tr>
<tr>
<td>Kate Bolton</td>
<td><a href="http://www.hebo.fi/anvoskelut.htm">http://www.hebo.fi/anvoskelut.htm</a></td>
</tr>
<tr>
<td>Kate Bolton</td>
<td><a href="http://www.somersetcountygazette.co.uk/mostpopular">http://www.somersetcountygazette.co.uk/mostpopular</a></td>
</tr>
<tr>
<td>Kate Bolton</td>
<td><a href="http://artellacafe.com/blogs/aunt_bobbys_miscel">http://artellacafe.com/blogs/aunt_bobbys_miscel</a></td>
</tr>
<tr>
<td>Kate Bolton</td>
<td><a href="http://sellerssource.net/sites/category/product/page/2/">http://sellerssource.net/sites/category/product/page/2/</a></td>
</tr>
<tr>
<td>Kate Bolton</td>
<td><a href="http://www.flickr.com/photos/dalimur/223304509/">http://www.flickr.com/photos/dalimur/223304509/</a></td>
</tr>
<tr>
<td>Kate Bolton</td>
<td><a href="http://www.trampolining-online.co.uk/trampolint/(">http://www.trampolining-online.co.uk/trampolint/(</a>...)</td>
</tr>
<tr>
<td>Kate Bolton</td>
<td><a href="http://www.tripadvisor.com/ShowUserReviews-g147-">http://www.tripadvisor.com/ShowUserReviews-g147-</a> (...)</td>
</tr>
<tr>
<td>Kate Bolton</td>
<td><a href="http://www.linkedin.com/find/b/b7f671_49.html">http://www.linkedin.com/find/b/b7f671_49.html</a></td>
</tr>
<tr>
<td>Kate Bolton</td>
<td><a href="http://vcfa.com.au/Pages/library.htm">http://vcfa.com.au/Pages/library.htm</a></td>
</tr>
<tr>
<td>Kate Bolton</td>
<td><a href="http://www.doweb.co.uk/action_listing_urlid_sHR">http://www.doweb.co.uk/action_listing_urlid_sHR</a> (...)</td>
</tr>
<tr>
<td>Kate Bolton</td>
<td><a href="http://www.the-lidings.com/2005/02/25/looking.htm">http://www.the-lidings.com/2005/02/25/looking.htm</a></td>
</tr>
</tbody>
</table>

### Items you've stored to keep track of.

<table>
<thead>
<tr>
<th>Name</th>
<th>Delete</th>
<th>Matches</th>
</tr>
</thead>
<tbody>
<tr>
<td>kbb2007</td>
<td></td>
<td>no matches</td>
</tr>
<tr>
<td>Kate Bolton</td>
<td>X</td>
<td>3 matches</td>
</tr>
<tr>
<td>kateki</td>
<td>X</td>
<td>no matches</td>
</tr>
</tbody>
</table>

---

Census, Property and Birth data is © Crown Copyright. © 2008 DataPatrol powered by garlik

---

**Local Area Report**

Prepared: 14.05.08

Average house price: £455,000

View more...
easy for kids these days. Boys and girls have curfews. Kids don’t always have transportation. Many teens are just overscheduled.

“Teens are seeking this sort of access to public,” Boyd explains. “And they’re being forbidden from them in the physical space. So a certain degree of isolation is a public, right? But it’s a public that’s very much controlled by adults. The value of social network sites is that the control is not so obvious. And so it feels like some form of freedom among this culture where they don’t have much freedom.”

That’s what concerns parents who say there’s too much freedom online. Parents fear there aren’t enough controls on sites like MySpace and Facebook to protect their children from those who mean them harm. Boyd agrees that kids need to be careful to use online social networks safely. But she argues that kids need to practice their social skills somewhere before they venture out on their own. Online social networks are one of the public spaces where they can do that.

A Second Chance for a First Impression

Three weeks later, Lily is back in the United States after her vacation in Germany with her friend Dan. It’s the day freshmen arrive at the University of Vermont. Lily has just walked through the front doors of her dormitory. Parents and kids squeeze past her carrying duffle bags and plastic storage crates. She pauses in the hallway for a moment. Her new roommate Meagan has come down the stairs looking for her. And soon Meagan spots her.

Meagan rushes over to greet Lily. The two girls recognize each other instantly. After finding each other on Facebook, they became official Facebook friends. They’ve been messaging each other by writing on each other’s Facebook walls. They’ve even talked on the phone. That’s when Lily learned that Meagan pronounces her name “Megan.” Lily also discovered that they’d be reuniting with two other girls, Louise and Kate. Lily promptly “friends” them on Facebook.

Now Meagan and Lily hug and say how nice it is to finally meet in person.

Meagan asks if Lily needs help bringing her stuff up to their room. As she lifts one of Lily’s bags, she notices the skateboard propped against it.

“Oh my God, you skateboard?” Meagan exclaims. Their other roommate Louise has a skateboard too.

Lily hadn’t mentioned skateboarding among her Facebook interests. She says it’s been a while and she’s out of practice. It might have seemed like false advertising to claim it as a hobby.

The girls head Lily’s boxes up to their dorm room. The other kids on their floor tripped in and out and help each other with the heavy loads. Jann Koplif and Jimm Hendrix songs float down the hallway from another open door.

Most of the freshmen have been on Facebook in the days leading up to their arrival. Some of them even recognize each other in the hallways from their profiles. Meagan and Lily agree that they both really resemble their Facebook profiles. They say, no surprises there.

We All Have One

As moving day at the University of Vermont ends and parents go home, the freshmen start to gather in the way college freshmen usually do: at a party in someone’s dorm room.

This room is fairly typical. It has white walls and a gray linoleum floor. Just one fluorescent light is turned on for better party ambiance. Kids stand in clusters or sit on the floor. They’ve known each other for mere hours, but they all use Facebook. And they start to compare notes.

“I met my roommates on Facebook,” says Kate Reich, a freshman from Minnesota. “But they were different than they seemed on Facebook.”
Kathleen Bolton
This is Me
President (Past)
The Macomb Land Conservancy (Past)
Clinton Township, MI

Employment History

President
The Macomb Land Conservancy
Headquarters Address:
Macomb Land Conservancy 18875 Harper
Clinton Township, MI 48036
USA
Website: www.savingplaces.org
Phone: (586) 783-2385
Fax: (586) 783-4033
Macomb Land Conservancy home page

Web References

1. MLC Saving Places News Letter
www.savingplaces.org/archives/ - [Cached]
Published on: 2/23/2009 Last Visited: 10/20/2007
By Kathleen Bolton, MLC President

2. MLC Saving Places News Letter
www.savingplaces.org/archives/ - [Cached]
Published on: 1/1/2001 Last Visited: 10/20/2007
Pankow's plans to construct a "much-needed" golf course in the area have been
further slowed due to opposition from Kathleen Bolton, president of the Macomb
Land Conservancy. Bolton said the proposed golf course development would not
be wetland-dependent and "prudent and feasible alternatives" were not considered
for the property. "The Dean Creek is a parcel now to the Clinton River," Bolton
said. "Our studies show that the creek is a very healthy stream and protecting it is
important to the Clinton River Watershed."

Bolton also formed the Macomb Land Conservancy in March to help preserve
the kind of development for future generations."

"Macomb County is growing rapidly," Bolton said, "and our wooded habitats,
streams, and farmlands are rapidly disappearing."
Garlik: The Nature of our Data (i) Structured Data Relationships changing order of few times a year
Garlik: The Nature of our Data (i) Semi and unstructured Data Relationships changing order many times a year
Garlik: The Nature of our Data (i) Semi and unstructured Data
Relationships changing order many times a year
Garlik: The Nature of our Data (i) Semi and unstructured Data
Relationships changing order many times a year
## Data Matching Issues - Examples

<table>
<thead>
<tr>
<th>Category</th>
<th>Source(s)</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Misspellings</td>
<td>192.com</td>
<td>Farneti vs. Farnetti in Electoral Roll</td>
</tr>
<tr>
<td></td>
<td>Land Registry (via</td>
<td>&quot;Curly Way&quot; vs &quot;Curley Way&quot; in house price paid (can use PAF for address</td>
</tr>
<tr>
<td></td>
<td>nethouseprices)</td>
<td>hygiene)</td>
</tr>
<tr>
<td></td>
<td>192.com</td>
<td>Tom's middle name - Segun in CH vs. Shegun in Electoral Roll</td>
</tr>
<tr>
<td>Transposition</td>
<td>Companies House</td>
<td>E.g., example of birth date at CH month/day transposed compared to other</td>
</tr>
<tr>
<td></td>
<td></td>
<td>sources</td>
</tr>
<tr>
<td>Name Ordering/Usage</td>
<td>Companies House</td>
<td>Jim Stanfield (Lostwax) is actually &quot;John James Stansfield&quot;</td>
</tr>
<tr>
<td>Inconsistent usage</td>
<td></td>
<td>Use of middle initial, full middle name(s) or not e.g., for Nigel Grierson:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GRO - BMD Nigel P</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CH Nigel</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FSA Nigel Peter</td>
</tr>
<tr>
<td></td>
<td>192.com (Electoral Roll)</td>
<td>Nigel P.</td>
</tr>
<tr>
<td>Non-uniqueness</td>
<td>BMD</td>
<td>10 Michael J Harris births registered in Q1'49</td>
</tr>
<tr>
<td></td>
<td>192.com</td>
<td>&gt;&gt;200 &quot;Michael Harris&quot; in Electoral Roll and Directory Enquiries</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&gt;&gt;200 &quot;Michael J* Harris&quot; in Electoral Roll</td>
</tr>
<tr>
<td></td>
<td></td>
<td>63 Directors with same name &quot;Michael John Harr Oxfordshire&quot; = 3 Directors, 10 Electoral Roll</td>
</tr>
</tbody>
</table>
QDOS

A new sociometric

Your networth
Nigel Shadbolt

QDOS is a new way to measure your digital status. Check out Nigel Shadbolt's QDOS and compare it to yours.

Q3002

Overall position in QDOS rankings: 28,926

Your QDOS breakdown

Popularity
Impact
Activity
Individuality

Where are you on the internet?

- FOAF
- zidnet.co.uk
- bbc.co.uk
- bcs.org
- informatik.uni-trier.de
- news.com
- webscience.org
- ft.com
- gow.epsrc.ac.uk
- de.scientificcommons.org
- Facebook
- ukrcn.org.uk
- garlik.com
- themakedscientists.com
- rciuk.ac.uk
- brightmindgo.com
- raeng.org.uk
- computer.org
- users.esc.soton.ac.uk
- amazon.com
- Flickr
- nesc.ac.uk
- news.bbc.co.uk
- memoriesforlife.org
- anthemmedia.org
- www2006.org
- ontoworld.org
- epistemitas.co.uk
- meerkatting.com
- scrawford.blogware.com
BCS: UK Tech Industry Faces Skills Crisis
**News** The UK technology industry faces a looming skills crisis, according to the president of the British Computer Society, Professor Nigel Shadbolt. A fall of a third in the intake for full-time undergraduate computer science degrees since 2001...

Tags: BCS
[November 17, 2006, 15:42] [Find Related Articles]

BCS Calls For UK Net Neutrality Debate
**News** Professor Nigel Shadbolt said on Friday that, because so much of the internet's content is derived from the US, the UK and Europe would be affected by any net neutrality-related decisions made across the Atlantic.

Tags: Eu, Bandwidth, Legislation, Traffic
[September 17, 2007, 16:39] [Find Related Articles]

Judges Reveal Best (and Worst) Politicians' Websites
**News** BCS president Nigel Shadbolt chaired an experienced team of judges, including political journalists Patrick Wintour from The Guardian, and the Daily Mail's Quentin Letts. The inaugural MP Website Awards, an event spearheaded by the British...
Structure

- Garlik – the background and origins
- Going down the Semantic technology route
- Raising VC investment for a Semantic Tech company
- The Challenge of the Consumer
- Challenges of building the platform
- The Challenge of the Semantic Web - Garlik Labs
- Lessons Learnt
Technical Architecture

Unstructured data
WWW

Semi-structured data

Front End

JXT store

NUTCH

Harvested Pages

Ranked Searchable Segment

Natural Language Engineering

NPS

RDS

Structured data

JXT smush

ESWC

June 2008
Garlik and the design of Semantic Databases: Why Garlik uses RDF storage

- **Ragged Data**
  - Some existing data is intrinsically ragged - address data, companies house
  - Whilst this can be done in RDBMS query construction and maintenance is very awkward

- **Natively storing graph structures**
  - Easy representation of relationships
  - Build more complex data objects easily
  - Optimisation on graph queries (in particular self-joins)

- **Inclusion of new data**
  - Schemaless Approach; amend the ontology and add the data! V change request to DBA, change schema; retest all the applications;
  - E.g. unanticipated changes, people can have more than one name; RDMS column has to change to foreign key, create a new relations table to relate key to name, all the queries have to be checked and some changed
  - Much less house keeping and refactoring in RDF

- **Standards**
  - There are some
  - High levels of interoperability
Garlik and the design of Semantic Databases: Ontology “Engineering”

- **DPA**
  - Reasonably large and expressive people and interval centric ontology
  - Heavily reified
  - Has been extended continuously
  - Used no DL inference
    - No issues wrt data integrity - delete a graph all the triples go away
    - No surprising query results
    - Some performance benefit
  - Separate out domain and application ontologies
  - Engineered in Protégé and scope notes throughout

- **QDOS**
  - Very lightweight FOAF-like ontology
  - Built as we went along
  - Single domain and application ontology
Garlik and the design of Semantic Databases: Data access

- Use SPARQL and http PUT
  - SPARQL has no UPDATE syntax
  - SPARQL widely adopted and a W3C rec
  - For some applications http PUT is inconvenient
- Both well defined standards
- In principle decentralised, universal access

- We harvest RDF to central stores
- Alternative is lots of distributed stores
- We believe the former approach will be with us for some time
Garlik and the design of Semantic Databases: The challenge of URI Management

- Minting URIs for DPA and QDOS
  - Generously/promiscuously then link them up later

- QDOS
  - Have explicit link ups

- Management
  - Name conventions - need long term persistence and stability
  - Careful not to confuse URI of a persona and a person
  - These are long standing issues about nature and persistence of identity
Performance Measures and Benchmarks

- Would be nice…

- Lehigh University BM
  - Designed to benchmark inference behaviour
  - Very repetitive data
  - Queries are trivial

- Original Southampton University sets
  - Systems have evolved on
  - Decent range of query types was evaluated

- Internal to Garlik we use
  - TIGER Line
  - FOAF data from QDOS
Scale and Growth of our RDF data

- Running currently at 10 GT (1GT is $1 \times 10^9$ triples)
- Capable of currently sustaining 60 GT
- Designing out to the Tera Triple range
- The linked Web of RDF is likely to be very large - currently 100 GT and up
- E.g 3% of FOAF is 1 GT
Why did we build our own DBMS?

- We had no choice if we wanted an RDF solution

  Nothing working at the scale we needed

  Availability
  - Uptime, online backup, replication, access to source

  Even now commercial offerings e.g. Oracle 11g only up to order 10^7 triples, Jena order 10^6, OpenLink Virtuoso 10^7, DERI YARS2 7GT

  We needed different response modes for different applications
  - Import time
  - Update capability

  Wanted to control our own S/W destiny
Cluster Architecture

External services

Processing nodes

Gigabit switch

Storage nodes
Software versions
## Application Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DataPatrol</th>
<th>QDOS (+ FOAF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>JXT version</td>
<td>N</td>
<td>R</td>
</tr>
<tr>
<td>Size (triples)</td>
<td>10 GT</td>
<td>1 GT</td>
</tr>
<tr>
<td>Size (graphs)</td>
<td>100</td>
<td>2.7M</td>
</tr>
<tr>
<td>Cluster (NxCxR)</td>
<td>8x4x8</td>
<td>8x8x8</td>
</tr>
<tr>
<td>Query rate (peak)</td>
<td>200/s</td>
<td>1500/s</td>
</tr>
<tr>
<td>Import rate (avg)</td>
<td>75 kT/s</td>
<td>3 kT/s</td>
</tr>
</tbody>
</table>
Structure

- Garlik – the background and origins
- Going down the Semantic technology route
- Raising VC investment for a Semantic Tech company
- The Challenge of the Consumer
- Challenges of building the platform
  - The Challenge of the Semantic Web - Garlik Labs
- Lessons Learnt
Is it time for Semantic Tech Community to be bolder?

- Why are we not in the centre of the big debates?
  - E.g. data portability, the social graph, identity databases

- What are our strengths?
  - Open, standards based, understand & leverage linked-data

- Can we target some important emerging challenges to prove our value?

- Garlik is looking at emerging challenges in the identity arena

Gartner Inc.
Does online identity matter to people?

We asked 2,000 Britons whether they search online...

- Before hiring a lawyer, accountant etc – 32%
- Before hiring an employee – 11%
- Find out about a prospective boss – 20%
- Find out about prospective neighbours – 16%
- Before going on a date – 12%
Online dating is an emerging identity battlefield

Fake identities on online dating sites
Who is Robert Schock?

From: Mr. Robert Schock <robert@bank.com>
Sent: Tuesday, June 19, 2007 7:02:31 PM
Subject: Next Of Kin Claim

Dear Friend

I am Mr. Robert Schock, I work with the Chartered National Bank as an account officer in the Treasury/Credit Control Unit. I was the accounting officer of a national of your country, who died with his wife and their only son, after they were involved in a car accident.

After these several unsuccessful attempts, I decided to track his last name over the Internet, to locate any member of his family hence I contacted you. I have contacted you to assist in repatriating the assets and Capital valued at 25,540,000.00 Million Pounds left behind by my client before they get confiscated ....

Oakland California Law Firm

His reputation speaks for him. Let him speak for you.

Robert G. Schock is the person to contact if you, a friend or a family member have suffered a severe, disabling injury or the wrongful death of a family member.

Robert G. Schock has dedicated 40 years to fighting for the legal rights of injured people. He is known and respected as one of the top personal injury attorneys in the East Bay.
How a minor irritation becomes a full blown crisis

(what “Web Science” dynamics teach us about Email)

Huge anti-spam Industry emerges

~70% of email Is spam

Open email On “web” scale

Innovation leads To email

closed email groups On small scale

Transition leads To spam industry
What do “Web Science” dynamics point to about Social Networks?

What if ~70% of all online identities were fake?

Transition leads To identity spoofing Industry “for fun & profit”

Small trusted groups create online identities “for fun & profit”

Innovation leads To social networking industry

Social networking taken Up on massive, web scale And real-world implications emerge

We Are Here
What might a solution look like?

- Lightweight & Scalable
  - Can’t expect users or companies to work too hard
  - Needs to scale to hundreds of millions of people

- Sufficient
  - Needs to be “good enough” and adaptable, not perfect

- Understandable
  - Needs to fit users mental model; what do they do in real-life?

- Open and Standards based
  - (that’s what happens when TimBL sits on your advisory board!)

- This led us to the concept of Social Verification
Social Verification: The Concept
FOAF: A start point for social verification?

- **A handy, flexible way to manage Blog comments**
  - A busy blog site, some great comments, lots of random comments
  - Blogger wants to be flexible but doesn’t want to have to manage comments manually
  - So he wrote a bunch of perl scripts to accept comments from people in his FOAF file and people in their FOAF files (2 degrees out)

- **Could this solution be generalised?**
  - For blog comments? For email whitelists? For “friend requests”? Others?
  - Millions of FOAF files now
  - But not everyone can write their own perl scripts (I know, hard to believe!)
  - So, accessible, reliable FOAF reverse-index needed as starting point

- **Garlik has created this application**
  - 3m+ FOAF files (searchable, who you know, who knows you) and growing
  - Demonstrator app showing 2-degree OpenID verification given a FOAF file
Search for a profile

Simply enter an email address, or the URL of someone's homepage or blog in this box and click 'go'.

What is FOAF?

FOAF (Friend of a Friend) is a way of describing people, their activities, and their relationships to other people and objects. FOAF allows groups of people to truly open social networks without the need for a centralised database.

The FOAF project (http://www.foaf-project.org) was started in 2000 by Libby Miller and Dan Brickley. Today there are many millions of FOAF profiles published on the Web and it is growing daily. FOAF is a great way to build large scale, open social networks that are controlled by individual users and not by any one single company. To create your own FOAF file click here.

One of the challenges with FOAF is to make it more accessible, and Garlik's contribution is to make available this new QDOS FOAF service that allows people to search and browse millions of FOAF files in a user friendly way. Our aim is to index and make visible the entire FOAF universe and we will continue to extend, update and publish this index for free, to encourage wider use of FOAF for building social networks.
In Conclusion

- The Semantic Tech Community needs to step forwards and tackle some key emerging problems

- Garlik’s interest is consumer online identity, and we are focusing on emerging problems in that domain

- “Identity spam” is an inevitable consequence of the continued success of social networking

- Social verification is a lightweight, adaptable, scalable approach to tackling this problem, leveraging semantic web principals

- Garlik has created a FOAF based service (beta) for the community to explore and extend
Structure

- Garlik – the background and origins
- Going down the Semantic technology route
- Raising VC investment for a Semantic Tech company
- The Challenge of the Consumer
- Challenges of building the platform
- The Challenge of the Semantic Web - Garlik Labs

- Lessons Learnt
Lessons Learnt

- Importance of the Agile Programming approach
- Building on top of a non-finalised standard was painful (SPARQL)
- Balance of effort on various components - Harvesting v NLP v Database v Presentation
- Patience in waiting for the Linked Web of Data
- Indexing structures for RDF are quite different than RDBMS
- Stuff happens…
Things change...
Things change...

Nasdaq, 1996-2006
Phase 1 decline
"this must be the bottom"
Phase 2 decline
full retrace to pre-bubble level
post-bubble capitulation

Nasdaq 1997 - 2002: classic bubble symmetrical retrace

Working Daze

See, our stock options are not TOTALLY worthless. They make great PAPER AIRPLANES!
Some things are the same...
Some things are the same...
nigel.shadbolt@garlik.com
www.garlik.com
www.qdos.com
foaf.qdos.com
tomilube.blogspot.com

June 2008