Social Web! ... and now?

The Fifth Summer School on Ontological Engineering and the Semantic Web (SSSW'07)

Stephan Baumann
Competence Center Computational Culture
German Research Center for AI
Prologue
overwhelming

too many copycats

Semantic Web Ph.Ds MUST help!
Setting the scene (I)

- Devices
  - More computing power
  - More storage capacity
  - Less size
  - Connectivity, always-on, ad-hoc networks
- Software
  - Open Source, freeware, even frameworks
  - 1-click-away, even advanced academic stuff
Setting the scene (2)

- People
  - Digital Natives
  - Free-time programmers
  - Coding as hobby
  - Coding as art
  - Powerful communities (sourceforge, processing, generator-x, etc.)
  - APIs of the powerhouses (Google, Ebay, etc.)
Setting the scene (3)

• ... and as a result?!

• ripping
• sharing
• remixing
• mashing
• always
• everywhere

• transmedia: from evil ringtones to Web2.0 mash-ups
HOW TO HAVE A SUMMER HIT
THERE'S A SCIENCE TO IT. WE'VE GOT THE FORMULA

For many years, musicians have made records aimed at being summer hits, writes Dr Rupert Till. The Beach Boys realised that music which combines talk of summer, the beach and sunshine was a winning formula.

But for the first time, this formula has been scientifically expressed. The 'catchiness quotient' (outlined above) quantifies how likely we are to fall for a summer song, despite the fact they are mostly terrible.

First popularised in the 1970s, when British holiday-makers began travelling abroad on package deals, these hits all follow the same formula. From classics such as 'Un Paloma Blanca' to Kaoma's 'Lambada' — by way of Demis Roussos's 'Happy to Be on an Island in the Sun' — they each comprise a mixture of holiday and foreign-language references, exotic beats and simple sing-along lyrics. The formula works, until another popular song is ransacked into chart高位.

The Ketchup Song' scores high (84,240), as it has a well known dance and generous amounts of flesh in the video — a phenomenon in its infancy when 'Y Viva España' (score: 7083) was released.

Dr Rupert Till is a tutor at the University of Huddersfield, specialising in postmodernity in music. He is also a member of electronica group, the Chillagge.

WORKING OUT THE HIT FORMULA

CQ (catchiness quotient) The higher the figure, the more irresistible the summer hit.

B (budget) Amount in sterling the record company has to spend on marketing, record pluggers and PR companies.

F (flesh tone) Amount of flesh flashed in the video, on CD covers, flyposters and other publicity material. The video to Las Ketchup's 'Asereje' (which scores a total of 84240) features a different sample of winners until slavery.

hr (musical holiday references) Number of times the song uses elements of musical styles associated with British holidaymakers with holidays, such as Spanish guitars, castanets and samba rhythms.

L (foreign lyrics) Number of words sung in a foreign language.

D (dance routine) Number of steps in the routine.

Co (promo copies) Number of copies of the single given to DJs and holiday reps abroad.

Sp (chorus span) Distance in semi-tones from the lowest to the highest note. The lesser the range, the more singalongable.

T (average UK temperature in °C) The worse the British summer, the more that people want to remember the glorious weather on holiday, and the better a record sells.

Cultural impact

SUMMER HITS ALL FOLLOW THE SAME SCIENTIFIC FORMULA

Would You Download Music From This Man?

Everyone else does. Meet the one-stop shop for everything from MP3s to A Beautiful Mind and every Game Boy title on the market.

By Jeffrey M. O'Brien
Photograph by Michele Asselin
so what?!
Exciting area for scientific engagement :-}
He shares my musical taste ... handsome guy
• Freeware supporting common mobile phones
• Sources at Sourceforge (join in!)
• Import of Last.fm profiles
  • Last.fm „scrobbles“ 15 mio times daily
• CBS acquired Last.fm for $280 million

bluetuna.opendfki.de

Person, Context, Artifact, Relation, Network
I. People
User models for music

<User>
<generalbackground>
  <name>John White</name>
  <education>MS</education>
  <citizenship>US</citizenship>
  <birthdate>9/7/1974</birthdate>
  <sex>male</sex>
  <occupation>student</occupation>
</generalbackground>
</user>

<musicbackground>
  <education>none</education>
  <instrument>piano</instrument>
  <instrument>vocal</instrument>
</musicbackground>

<generalpreferences>
  <color>blue</color>
  <animal>dog</animal>
</generalpreferences>

<musicpreferences>
  <genre>classical</genre>
  <genre>blues</genre>
  <genre>rock/pop</genre>
  <composer>Mozart</composer>
  <artist>Beatles</artist>
  <sample>
    <title>Yesterday</title>
    <artist>Beatles</artist>
  </sample>
</musicpreferences>

<habit>
  <context>I’m happy
  <tempo>very fast</tempo>
  <genre>pop</genre>
</context>
  <pfeature>romantic
  <tempo>very slow</tempo>
  <softness>very soft</softness>
  <title>*love*</title>
</pfeature>
  <context>bedtime
</context>
</habit>

[Chai, Vercoe: ISMIR2000]
FOAF

- http://www.foaf-project.org

- FOAF documents describe the **characteristics and relationships amongst friends of friends**, and their friends, and **the stories they tell**.

- „So FOAF is quite pluralistic in its approach to representing relationships between people. FOAF is built on top of a general purpose machine language for representing relationships (ie. RDF), so is quite capable of representing any kinds of relationship we care to add. **The problems are generally social rather than technical**; deciding on appropriate ways of describing these interconnections is a subtle art.“

- „Perhaps the most important use of **foaf:knows** is, alongside the **rdfs:seeAlso** property, to **connect** FOAF files together. Taken alone, a FOAF file is somewhat dull. But linked in with 1000s of other FOAF files it becomes more interesting, with each FOAF file saying a little more about people, places, documents, things... people they mention, and so on.“

[Dan Brickley, Libby Miller: foaf-project.org, 2000-2005]
FOAF-related applications

- FOAFme (http://foafme.opendfki.de)
- FOAFNaut
- Gnowsis

[Sauermann, Lee: foafcamp.asemantics.org 2004]
Users & Context

1. **Environmental context:** This part captures the users surroundings, such as things, services, light, people, and information accessed by the user.
2. **Personal context:** This part describes the mental and tuples physical information about the user, such as mood, expertise, disabilities and weight.
3. **Social context:** This describes the social aspects of the user, such as information about friends, relatives and colleagues.
4. **Task context:** the task context describe what the user is doing, it can describe the user’s goals, tasks, activities, etc.
5. **Spatio-temporal context:** This type of context is concerned with attributes like: time, location and movement. The different aspects of the contexts are attribute-value tuples that are associated with the appropriate contexts.

[Kofod-Petersen&Cassens: MRC2005, FLAIRS2006]
• GUMO - General User Model Ontology (OWL)

• UbisOntology (RDF-S)

II. People–Networks
Friendster and Vizster

- 2003: Friendster, very popular social software
  - descriptive personal profiles
  - + explicit friendship links
  - attracted millions of users -> massive network
- Hard to navigate user-interface
- Boyd & Heer decided to design a visualization system for online social networks [Heer, Boyd: Vizster, InfoVis2005]
  - Danah Boyd: well-known ethnographer, UC Berkeley
  - Jeffrey Heer: user-interface design expert, UC Berkeley
Friendster and Vizster

- Data feed
  - crawl of 1.5 million members
- storage: mySQL database
- Lucene for keyword search
- [Newman‘s2004] community identification algorithm
- Custom visualization of diverse network characteristics based on prefuse [http://prefuse.sourceforge.net]
- Several visual interaction modes based on ethnographic findings
Analysis and Visualization of Popular Networks

- Source: Top 100 Blogs in Germany
- Analysis of blogroll structure
- Means for trust, quality?
- Blogsono Student Project: Visualization and Sonification
Analysis and Visualization of Popular Networks

- Information Architecture Lab, FH Kaiserslautern, Germany [Prof. Speck et al]
- Several Student Projects
  - Vinex: Visualization
  - Nepotist Analyzer: Economic Networks Analyzer
  - spamCAN: Spam Analyzer
  - PeerLo: BitTorrent Network Analyzer
  - B.A.S.S: MySpace Music Analysis
[Speck, Hendrik and Frédéric Philipp Thiele. Social Network Analysis. Business models, Explanations, Attributes and Features of Social Networks. CeBIT. WeblogCamp@CeBIT. March 18th, 2007, Hanover (Germany)]
[Speck, Hendrik and Frédéric Philipp Thiele. Social Network Analysis. Business models, Explanations, Attributes and Features of Social Networks. CeBIT. WeblogCamp@CeBIT. March 18th, 2007, Hanover (Germany)]
III. People-Artifacts-Networks
Social network analyzer for Enron's emails

When Enron was hit with lawsuits, the prosecution required the company to turn over the incriminating emails it had in its database through a legal process called discovery. Usually a company undergoing this kind of discovery carefully combs through its email and publishes only those messages that are relevant to the case, eliminating its employees' private messages and confidential material otherwise not related to the case.

However, Enron opted to save money by just dumping all of its email online, employee love-letters and all. This was bad for employee privacy, but good for social network researchers.

Jeffrey Heer at UC Berkeley has produced "enronic," a system for visualizing the relationships between Enron employees based on who emailed whom and how often. It's a Java applet that you can download and mess around with -- you can even take a turn at sleuthing out hidden participants in the scam by trying your own social network analysis.
Enronic

- Mining the social network from the data
  - e.g. by using Applied Natural Language Processing
- Email is straightforward
  - sender, recipient
  - direct edge in case of email
  - existing categorization of email body into a given taxomony (business, strategy, etc.)
  - further aspects of email body could be used for clustering
- Interactive visualization of these properties
I understand that we may have no choice in the matter, but having witnessed Lorett Lynch’s inability to define market power when asked to do so by Joe Barton last Monday, and having heard Carl Wood declare war on capitalism at FERC’s hearing on Tuesday, and knowing that the likes of Harvey Morse (PUC FERC attorney), who has never come down on the side of “no market power here” in any analysis, is “heading up” the PUC investigation, and seeing Gov. Davis’ press releases demanding refunds from the faceless, out-of-state suppliers, I’m well, maybe a little nervous about them getting their paws on our...
Flink: The Who is Who of the Semantic Web

- 1st prize Semantic Web Challenge 2004
- Data feeds:
  - Webpages
  - Emails (8185)
  - Publications (5147)
  - FOAF profiles
- Extraction of underlying social network + generation of an ontology + visualization
  - Co-occurrence analysis, inferences, identity reasoning, etc.
  - Social Network Analysis (SNA)
Flink: Online
Spam Analyzer

[Speck, Hendrik and Frédéric Philipp Thiele. Social Network Analysis. Business models, Explanations, Attributes and Features of Social Networks. CeBIT. WeblogCamp@CeBIT. March 18th, 2007, Hanover (Germany)]
IV. Tools
Social Network Analysis

- Specialization of network analysis
- Well-known field (e.g. sociologists, consultants, etc.)

**The network**

- Nodes (here people or groups)
- Relations (directed or undirected)
- Properties of relations (name, type, weight, etc.)

- SNA deals with **properties of the network**
  - Performance of subsets
  - Location of actors (implies possibilities and constraints for their actions)
Social Network Analysis

- Most popular measures
  - Degrees
  - Betweenness
  - Closeness
- Further properties
  - Boundary Spanners
  - Peripheral Players
  - Network Centralization
Social Network Analysis

- Network analysis packages
  - JUNG [http://jung.sourceforge.net/]
  - Pajek [http://vlado.fmf.uni-lj.si/pub/networks/pajek/]
  - UCINET [http://www.analytictech.com]
Self Organizing Map (SOM)

- Unsupervised, self-organized processing of data inspired by cortical maps in the human brain
- Non-linear projection of high dimensional data to low dimensional grid (usually 2D)
- Preservation of input space topology: data points close in input space are close on the map
- In contrast to
  - MultiDimensionalScaling (MDS)
  - PrincipalComponentAnalysis (PCA)
    - the original data space distances can be shown.
    - entangled clusters can be separated.
    - projection and clustering are provided
**Artists:** 398  
**Input:** 398 * 3313  
**SOM:** 49 units (7*7)
Learning from music-related Networks

- Large-scale experiment: AMAZON reviews as feeds
  - 33879 artist reviews
  - 33879 x 242 (dim. reduction PCA)
  - SOM: 3000 units (60 x 50) representing genre/style concepts

- Interactive music station for Netaudio: archive.org MP3 + XML feeds
  - (optional audio analysis)
  - 1125 artist reviews
  - 1125 x 1139 features
  - SOM: 180 units (15 x 12) representing genre/style concepts
ANLP + Visualization

- “Decoding the (textual) artifacts“
- JUDGE

- Just another Java-based Text Retrieval & Machine Learning Framework
  - automatic feature extraction
  - feature preprocessing and attribute selection methods
  - document classification with several classification methods
  - document clustering with several clustering methods
  - full support of almost all WEKA Algorithms
  - easy usage as a webservice with a special interface
  - http://www3.dfki.uni-kl.de/judge

- DYNAQ: interactive, dynamic queries

- Visualization: processing.org
Semantic Media: Tagging

Instance browser

Media player

List of annotations

SmartMedia ontology

Media instance
Bottom-up meets top-down

V. Networks-Interaction
(„Games, Chats, Tags ...“)
Related work at CHI2007

- S. Ahern, D. Eckles, N. Good, S. King, M. Naaman, R. Nair, Yahoo! Research Berkeley "Over-Exposed? Privacy Patterns and Considerations in Online and Mobile Photo Sharing"

- flickr upload study: private vs. public photos

- 81 users, 37k photos

- privacy determination by location and content (tags)

- demand for privacy policies!

- no privacy concerns by users yet
Related work at CHI2007

- C. Lampe, Michigan State University, E. Johnston, University of Michigan "Follow the Reader: Filtering Comments on Slashdot"

- slashdot comment rating

- critical mass vs. information overload

- analysed 2.4M page hits by 90k users

- interviewed 8k users

- ratings considered useful!

- filter customization is done rarely

- importance of defaults!

[Obradovic, Darko: State of the Art in SNA & CHI / preliminary work for Ph.D, June2007, DFKI Kaiserslautern, Germany]
Related work at CHI2007

- N. Ducheneaut, E. Nickell, R. Moore, PARC, N. Yee, Stanford University
  "The Life and Death of Online Gaming Communities: A Look at Guilds in World of Warcraft"

- max. size for creative & cooperative social groups: 45 (Dunbar number: 150, current WoW guilds: <35)

- social factors for group growth & death

- monitored 5k guilds on 5 WoW servers

- "small is beautiful"

- importance of social group management

[Obradovic, Darko: State of the Art in SNA & CHI / preliminary work for Ph.D, June2007, DFKI Kaiserslautern, Germany]
Related work at CHI2007

- M. Ames, Stanford University, M. Namaan, Yahoo! Research Berkeley
  "Why We Tag: Motivations for Annotation in Mobile and Online Media"

- evaluation with flickr/zonetag (direct upload from cameraphone) with 172 users

- taxonomy of tagging motivations

- tagging for social purposes is more motivating

- tagging at creation time is favoured, won't be done later

- tag suggestions help a lot

[Obradovic, Darko: State of the Art in SNA & CHI / preliminary work for Ph.D, June2007, DFKI Kaiserslautern, Germany]
Related work at CHI2007

- A.W. Rivadeneira, University of Maryland, D. M. Gruen, M. J. Muller, D. R. Millen, IBM Research "Getting Our Head in the Clouds: Toward Evaluation Studies of Tagclouds"

- cognitive research

- spatial algorithms (size)

- clustering algorithms (similar, occur together)

- tested: font size, quadrant and proximity-to-largest-word

- experiment with 13 subjects

- tagcloud designers should focus on layout

- simple lists by frequency are ok!

[Obradovic, Darko: State of the Art in SNA & CHI / preliminary work for Ph.D, June2007, DFKI Kaiserslautern, Germany]
Figure 2. (a) Tagcloud depicting Quadrants (UL=Upper-Left, LL=Lower-Left, UR=Upper-Right, LR=Lower-Right). (b) Tagcloud depicting Proximity-to-largest-font (AD=Adjacent, N-AD=Non-Adjacent).
Related work at CHI2007


- interactive, more social television experience

- experiment: cartoon watching with simultaneous chat (image attached)

- settings: real-time, chat breaks, chat after cartoon

- is fun, has social benefits, but distracts

- use intermissions for appropriate content

- future work: evaluate different content

[Obradovic, Darko: State of the Art in SNA & CHI / preliminary work for Ph.D, June2007, DFKI Kaiserslautern, Germany]
[J. D. Weisz, S. Kiesler, H. Zhang, Y. Ren, R. E. Kraut, Carnegie Mellon University, J. A. Konstan, University of Minnesota]

"Watching Together: Integrating Text Chat with Video"
Related work at CHI2007


- the "mobile factor" changes video experience, especially in the social context

- behaviour study with 28 participants

- mobile video is more than passive consumption due to boredom

- social importance of owning content and getting credit for recommending content

- solitary activity has social underpinnings (watch video at a time it won't annoy your family)
The End
Hot topics (2006 talk!)

- SPACE: Locative media
- TIME: Long-term (lifelong?) changes
  - Development of „taste“ (with respect to objects)
  - Liu, Maes & Davenport [Unraveling the Taste Fabric of Social Networks, IJSWIS2005]
  - Dependencies on the development of social network preferences?
Hot topics (2006 talk!)

- DIGITAL IDENTITY
  - Sherry Turkle [The Second Self, 2005] [Life on the Screen, 1995]
  - Danah Boyd [Identity Production in a Networked Culture, AAAS2006]
  - Liberty Alliance [projectliberty.org]
  - ID 2.0 [Dick Hardt: OSCON 2005 Keynote]
Identity matters!

- Privacy vs. Impression Management (=Exhibitionism?)

- OpenID as a supporting open standard!?

- Research issues
  - Aggregation of „FACETED IDENTITY“
  - Measures for TRUST

- Political implications
  - e.g „Vorratsdatenspeicherung“ in Germany

- Projects with CCTV and RFID technology
Hot topics (2006 talk!)

- Bridging the gap between Web2.0 and SemanticWeb?
- GRDDL (Gleaning Resource Descriptions from Dialects of Languages)
- 2007: From a commercial / startup perspective
  - Semantic social networks are „hot“, some rely on microformats
  - yowhassup.com
Think...Post...Get in touch.

Characters available: 65

http://www.solution-media.de/blog Thomas Huhn on semantics and Identity 2.0

This is my... ergn education career work goal belief appearance nutrition spare-time property lifestyle sexuality bbq/website RSS-feed follower colleague co-worker co-resident neighbor friend acquaintance contact child parent sibling spouse kin sweetheart date crush muzel self back

Submit post to "professional all" tagged as "my blog/website"

attach file:

Maximum size for all uploaded files is 5.0 MB; images will automatically be reduced to thumbnail and 800x600 size.

Livefeed from all

Stephan Baumann - stephan.baumann@ywhassup.com

Bio

Von Beruf: Informatiker Ansonsten: Musiker

Networks

Tags

Stephanie Huhn - stephanie.huhn@ywhassup.com

Bio

Me and my dog Max!

Networks

Tags

Thomas Huhn

Is male. is born 1965-11-06, is located in 67527 DE (timezone Europe/Berlin), speaks German, OpenID is thomas.huhn.ywhassup.com

About

Entrepreneur, Business-Angel and OpenID Evangelist.

Your own words

upcoming cinema

Local weather

@stephan.baumann can look into our garden and see the window

@stephan.baumann is brilliant in artificial intelligence

@stephan.baumann is my older daughter
Problems

• 100 onliners =
  • 1 creator +
  • 10 interactions +
  • 89 viewers
• Slots of attention decrease
• Cultural differences
Future trends?

- „Precision over Recall!“
- StumbleUpon.com
- „Semi-closed and closed“ networks
- Quality, trust matters!
Future trends?

- „Generation Mobile“ (e.g. Twitter)
- „Generation Game“
- I still wait for efficient reasoning :)
  - large-scale RDF stores
  - full-fledged „semantified“ media
Mind the gap(s)!

- Try to work on an application you are missing personally! (intrinsic motivation)
- I will start to work on Music Information Retrieval again!
- Work on real-world data!
- Show it to your friends!
- Talk about it!
thanks for listening to my few bits ... 
send emails(!) to: 
stephan.baumann@dfki.de 
read our blog 
www.computationalculture.org