

Semantic Computing: From Patterns to Meaning

Stanford, September 20, 2011

Industry Panel, ICSC'11

Natasa Milic-Frayling

Principal Researcher



Microsoft Research Cambridge

<http://research.microsoft.com/>

Microsoft Research (MSR) Worldwide

MSR New England



MSR Asia



MSR Cambridge



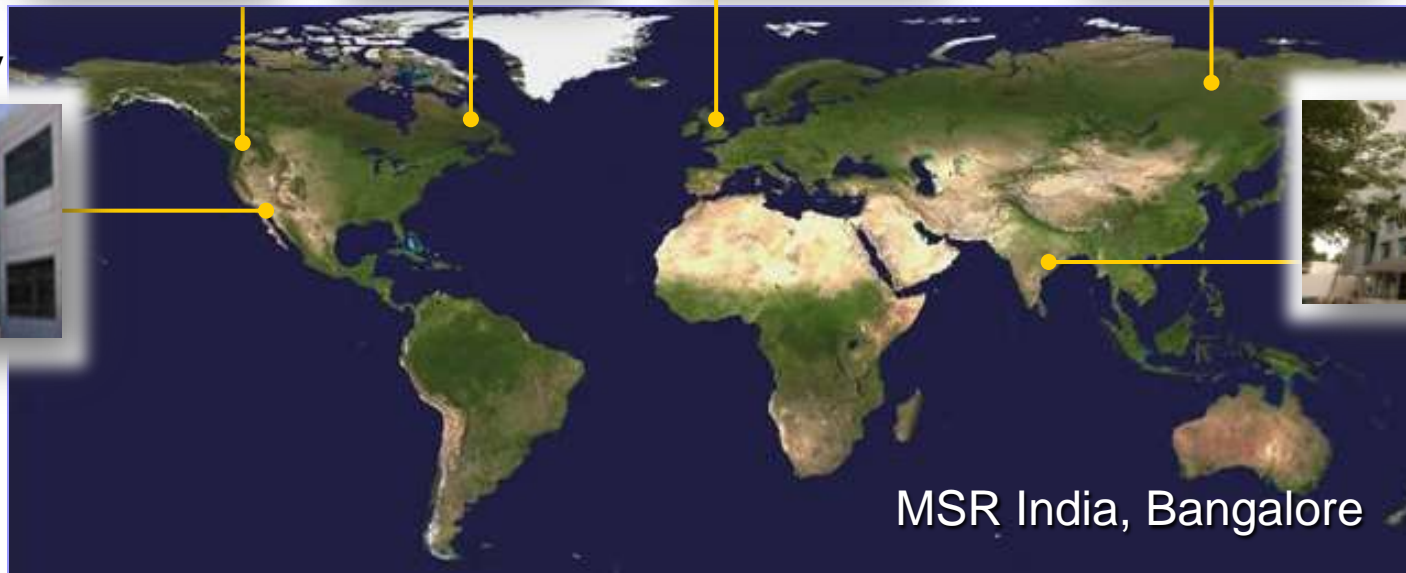
MSR Redmond



MSR India



MSR India, Bangalore



Silicon Valley



Research in Semantics

- Analysis of language expressions
- Meaning of utterances
- Meaning of gestures
- Meaning of patterns in digital traces (user activities and user intent)

ARTICLE *from the Encyclopædia Britannica*

semantics, also called **semiotics**, **semology**, or **semasiology**, the philosophical and scientific study of meaning in natural and artificial **languages**. The term is one of a group of English words formed from the various derivatives of the Greek verb *sēmainō* ("to mean" or "to signify"). The noun *semantics* and the adjective *semantic* are derived from *sēmantikos* ("significant"); *semiotics* (adjective and noun) comes from *sēmeiōtikos* ("pertaining to signs"); *semiology* from *sēma* ("sign") + *logos* ("account"); and *semasiology* from *sēmasia* ("signification") + *logos*.

Analysis of Language Expressions: Paraphrasing in English

<http://labs.microsofttranslator.com/thesaurus/>

The screenshot shows the Microsoft Translator Contextual Thesaurus interface. The browser address bar displays the URL <http://labs.microsofttranslator.com/thesaurus/>. The page title is "Contextual Thesaurus" and the subtitle is "Translate from English to English to explore alternate ways of expressing the same idea. (Learn more...)"

The interface includes a search box with the text "rephrasing is a hard problem for the computer." and a list of suggestions: "rewording", "reformulation", "rewriting", and "rephrasing". A "Suggestions" button is visible.

Below the search results, a word network diagram is displayed. The diagram shows the following structure:

- Input: "rephrasing is a hard problem for the computer."
- Word "rephrasing" is connected to "rewording", "reformulation", "rewriting", and "rephrasing".
- Word "is" is connected to "rewording" and "is".
- Word "a" is connected to "is" and "a".
- Word "hard" is connected to "a" and "hard".
- Word "problem" is connected to "hard" and "problem".
- Word "for" is connected to "problem" and "for".
- Word "in" is connected to "problem" and "in".
- Word "on" is connected to "problem" and "on".
- Word "of" is connected to "problem" and "of".
- Word "the" is connected to "of" and "the".
- Word "your" is connected to "of" and "your".
- Word "a" is connected to "of" and "a".
- Word "computer" is connected to "a" and "computer".
- Word "computer" is connected to "of" and "computer".

The diagram uses arrows to show the relationships between words, with a red arrow highlighting the path from "rephrasing" to "is" to "a" to "hard" to "problem".

Contextual Thesaurus

Translate from English to English to explore alternate ways of expressing the same idea. ([Learn more ...](#))

enter your phrase

many are dismayed by his
behaviour.

Hint: Enter short phrases (about 4-8 words) in a
business or formal style. To see another random
example, refresh your browser (F5).

clear

suggestions



many are shocked by his behavior.

many were disappointed by his behavior.

many were appalled by their behavior.

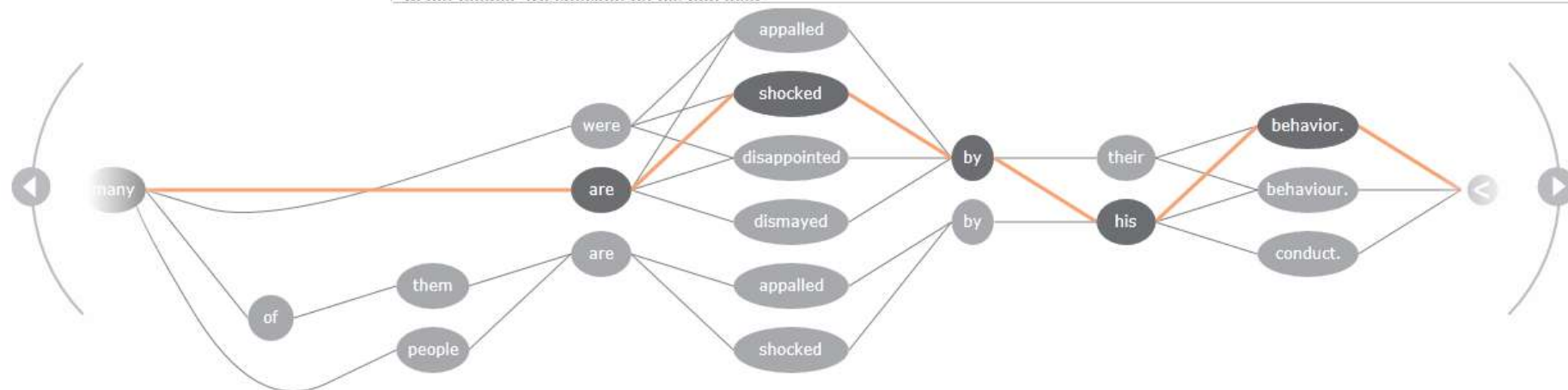
many of them are appalled by his behavior.

many are appalled by his behavior.

many are shocked by their behaviour.

many were shocked by their behaviour.

many people are shocked by his behavior.



Sentence:

“many are dismayed by his
behaviour”



many are dismayed by his behaviour

many were appalled by his behavior

many were shocked by his behavior

many are shocked by his behavior

many are shocked by their behaviour

many were shocked by their behaviour

many were appalled by their behavior

Meaning of Utterances: Search Over Audio

http://www.msra.com/audiosearch_demo/

The screenshot shows a web browser window displaying the Microsoft Video Web search results page. The search query is "semantic conferences". The page indicates that 1783 videos were searched from 741.1 hours in 0.64 seconds, showing results 1-20 of 21. Two video results are visible:

- Microformats and Semantic Markup** (Mix11 - 4/1/2011): A video discussing microformats and semantic markup. The description mentions that microformats are simple HTML design patterns for adding semantics to web content, making it more findable, extensible, standards-compliant, and usable. It also mentions Plain Old Semantic Markup (POSH). The video player shows a progress bar at 00:57:27.
- Expert to Expert: Natural Language and Computational Linguistics** (Channel 9): A video discussing natural language and computational linguistics. The description mentions that computational linguistics is an interdisciplinary field dealing with the statistical and/or rule-based modeling of natural language from a semantic knowledge base. The video player shows a progress bar at 00:57:32.

The browser's address bar shows the URL http://www.msra.com/audiosearch_demo/. The Microsoft Research logo is visible in the top right corner of the page.

Meaning of Utterances: Search Over Audio

http://www.msra.com/audiosearch_demo/

Search through the audio,
not just text.



Search through Videos from selected Microsoft sites

We search through the audio, not just text. Search through the sound tracks of over 20,000 videos on this site from different sources. Click on search-result text snippets to navigate directly to keyword matches in the video. Select the source content from the tabs and type a query into the search box above, or try one of these example queries:

[blob storage](#) [speech recognition](#) [mobile sdk](#) [parallel ling](#)

Powered by [MAVIS](#) and [Microsoft Research](#).

Meaning of Gestures: Kinect SDK

<http://research.microsoft.com/en-us/um/redmond/projects/kinectsdk>

Kinect for Windows SDK beta

home **download** documentation forums about

Download Kinect for Windows SDK beta



The Kinect for Windows SDK beta is a starter kit for applications developers that includes APIs, sample code, and drivers. This SDK enables the academic research and enthusiast communities to create rich experiences by using Microsoft Xbox 360 Kinect sensor technology on computers running Windows 7.

[Learn more >](#)

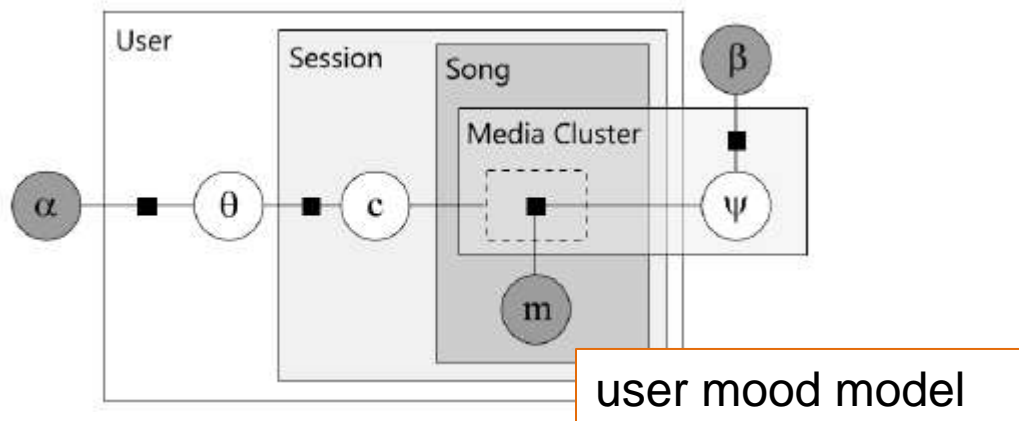
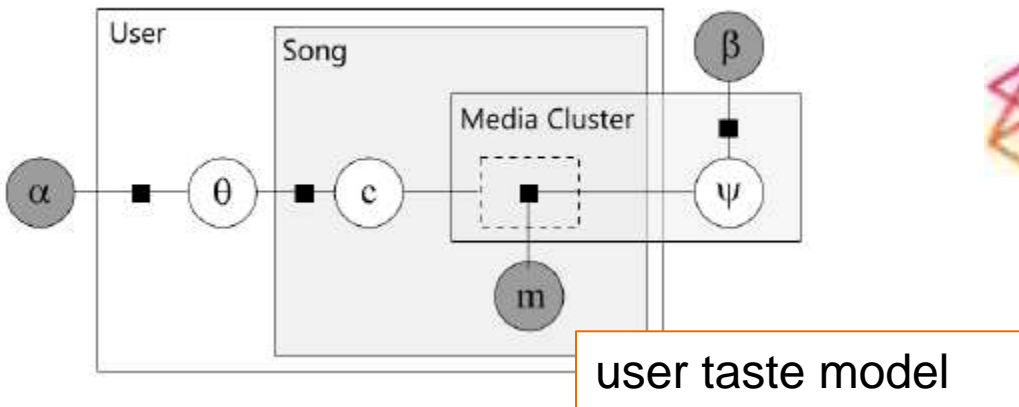
The Kinect for Windows SDK beta includes the following:

- Drivers, for using Kinect sensor devices on a computer running Windows 7.
- Application programming interfaces (APIs) and device interfaces, together with technical documentation.
- Source code samples.

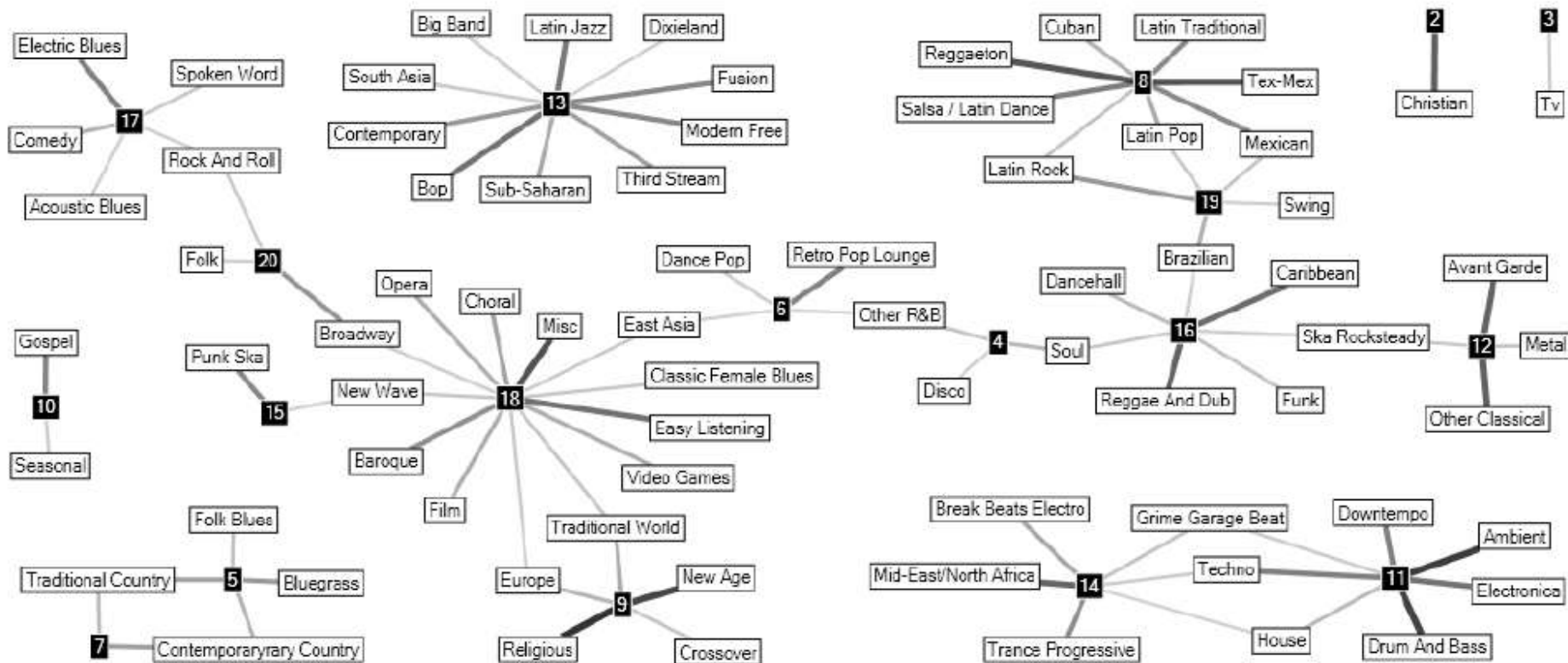
File name	File size	Version	Date published	Language
kinectsdk32.msi	21,320 KB	1.00.12	29 July 2011	English

32-bit download ↓

Meaning of Patterns in Digital Traces: Zune – Users' Mood and Taste



Pattern detection: Clusters from Bayesian Inference



Infer.Net

<http://research.microsoft.com/en-us/um/cambridge/projects/infernet>



infer.net

- Home
- Download

Documentation

- User Guide
- Tutorials & Examples
- API Documentation
- Resources & References

Support

- FAQ
- Forum
- Blog
- Contact Us

Team

- Meet the Team
- MLP Group
- MSR Cambridge

::: INFER.NET

Infer.NET is a framework for running Bayesian inference in graphical models. It can also be used for [probabilistic programming](#) as shown in [this video](#).

You can use Infer.NET to solve many different kinds of machine learning problems, from standard problems like [classification](#) or [clustering](#) through to [customised solutions to domain-specific problems](#). Infer.NET has been used in a wide variety of domains including information retrieval, bioinformatics, epidemiology, vision, and many others.

Infer.NET 2.4 beta 2 is now available [17th December, 2010].
 This release is a *minor update* to Infer.NET 2.4 beta 1. See the [release change history](#) for details.

Questions? Please use [the forum](#) to provide feedback and to share the ways in which you are using Infer.NET (or send e-mail to infersup@microsoft.com).

Suggestions?

Citing Infer.NET If you use Infer.NET as part of your research, please cite us as detailed in [the FAQ](#).



infer.net

User Traces in Social Network



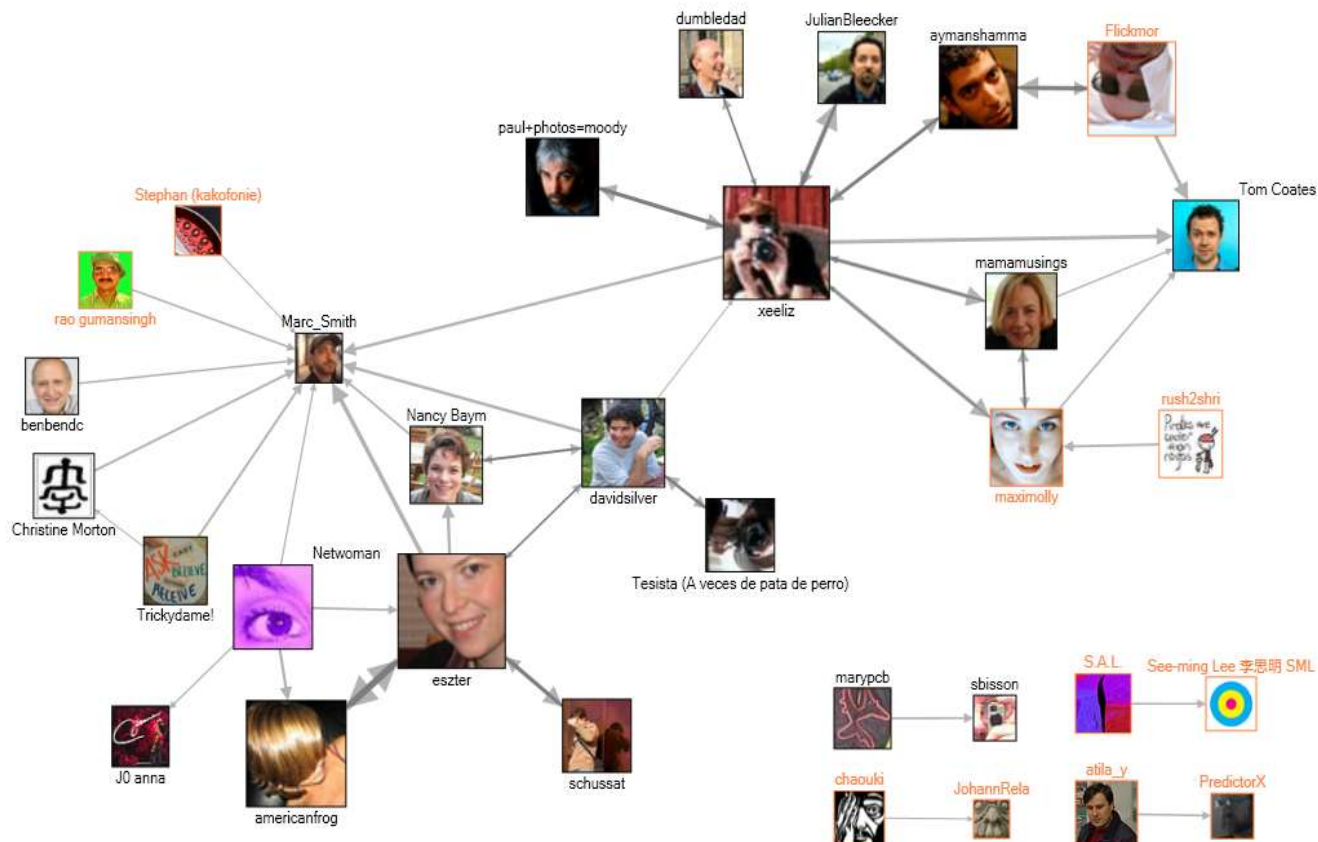
ENGAGEMENT: *Social network vs. managed crowdsourcing service*

INCENTIVES: *Social rewards vs. pay*

Pattern Discovery and Semantic Interpretation: Graph of Co-occurring Flickr Tags

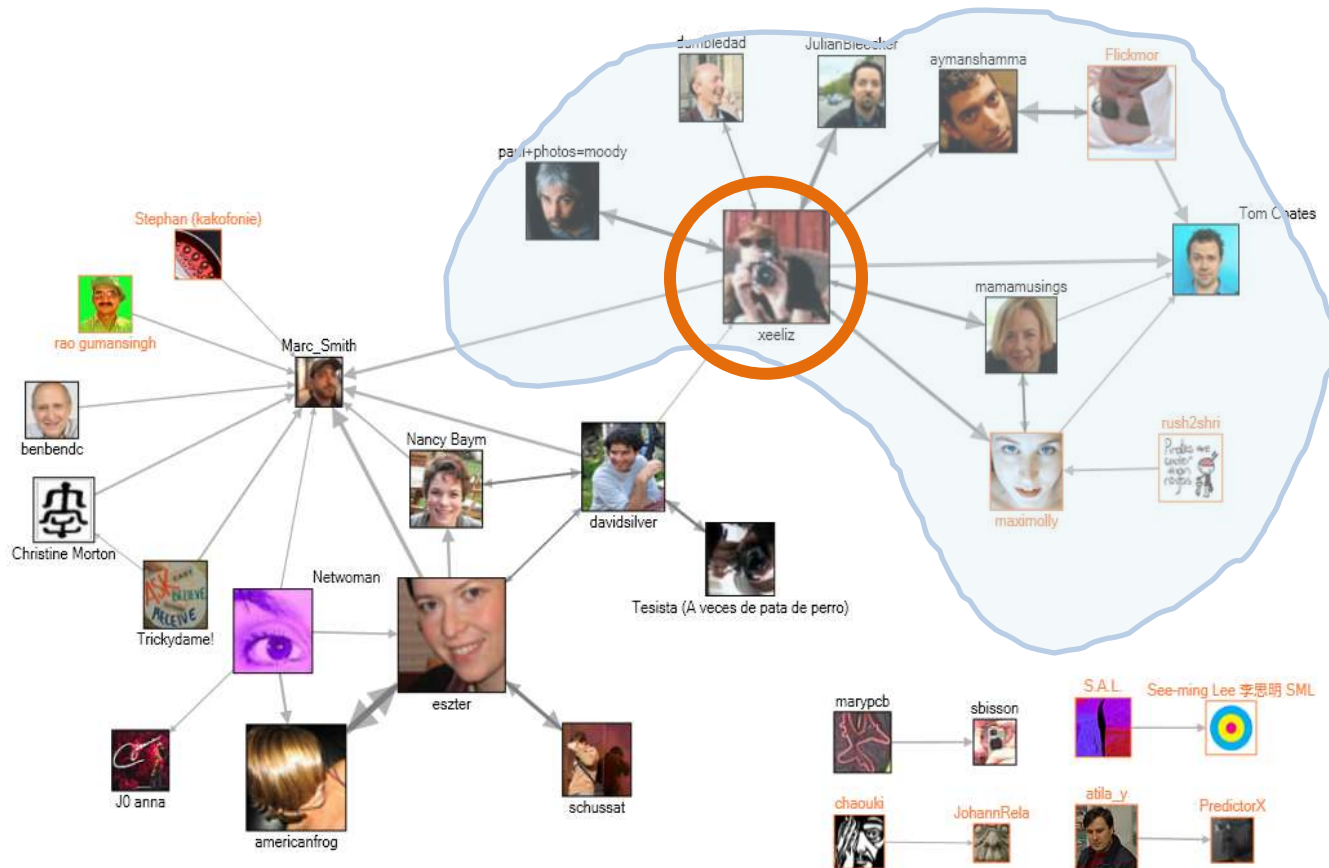


Pattern Discovery and Sociological Interpretation: 'Commenting' Activity on Flickr



Flickr users who commented on Marc_Smith's photos (more than 4 times)

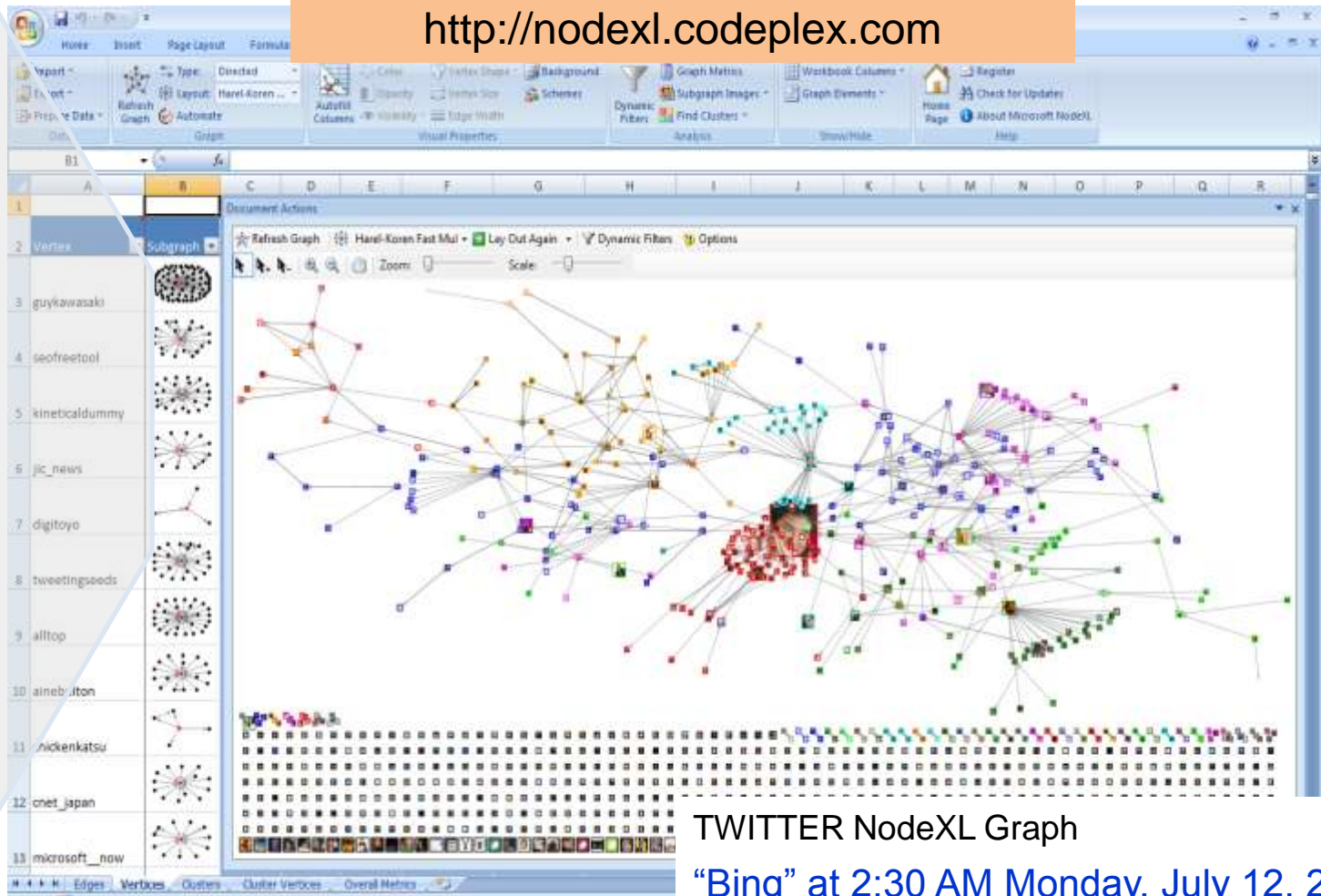
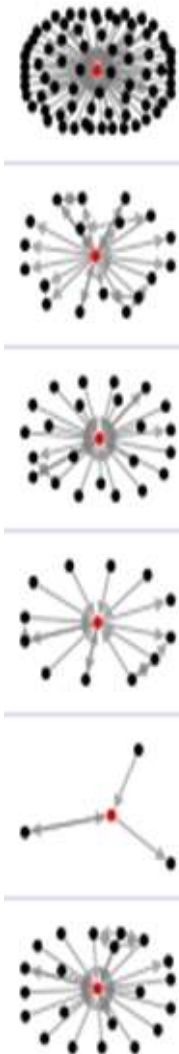
Pattern Discovery and Sociological Interpretation: 'Commenting' Activity on Flickr



Flickr users who commented on Marc_Smith's photos (more than 4 times)

Semantics of Network Patterns: NodeXL

<http://nodexl.codeplex.com>

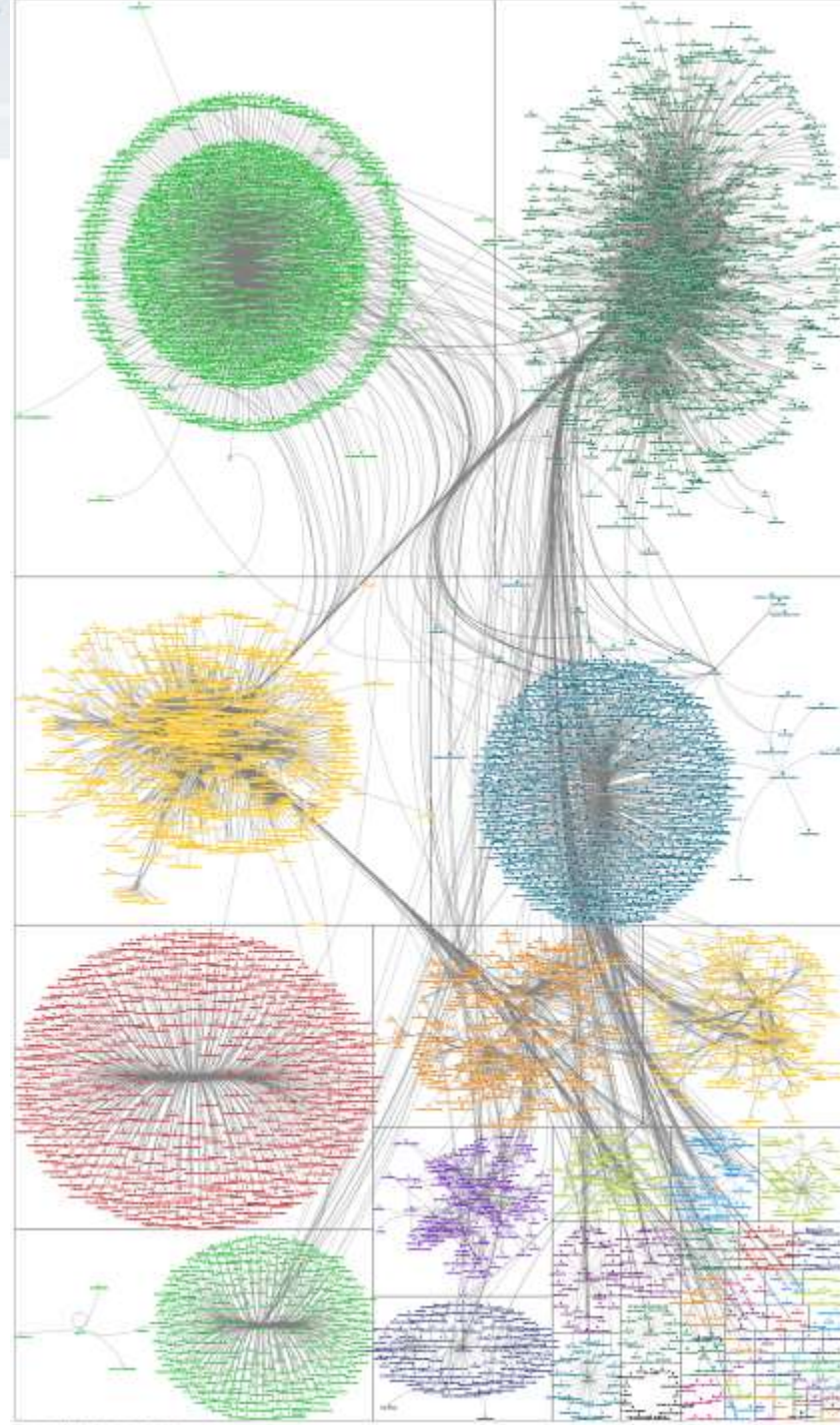


TWITTER NodeXL Graph

“Bing” at 2:30 AM Monday, July 12, 2010

From Pattern to Meaning: Email

- Validation of pattern analysis requires human input.
- Meaning can be considered globally accepted or strictly contextual, generally understood or individually constructed.



Enabling Users to Express the Meaning **Project Colletta**

<http://research.microsoft.com/en-us/um/cambridge/projects/ResearchDesktop/ProjectColletta>

The screenshot displays the Project Colletta interface. On the left, a sidebar lists various project folders such as 'ERMA-March31'11 Edinburgh', 'gesture paper', 'Karin', 'libya', 'MSRC', 'NodeXL Reports', 'NodeXL Study - UMD', 'NodeXL Visualizations', 'NP Study', 'Olton', 'SCAPE', and 'TAGtivity'. The main area shows a file explorer for the 'NodeXL Study - UMD' folder, listing sub-folders like 'Collins_SNA', 'DesignOfInformationMana...', 'Doran_SNA', 'Foss_SNA', 'Jones_SNA', 'Mathew.SNA', 'Seay_SNA', 'SNA_Final_Assignments', 'Stewart_SNA', 'Students' Assignments', and 'Thakar_SNA'. A preview window for 'Doran_SNA' is open, showing a network graph with red and green nodes and a 'PageCount: 3' indicator.



- Universal bookmarking
- Creating, naming, and describing collections of artefacts that are relevant to the user's tasks

Enabling Users to Express the Meaning Project Colletta

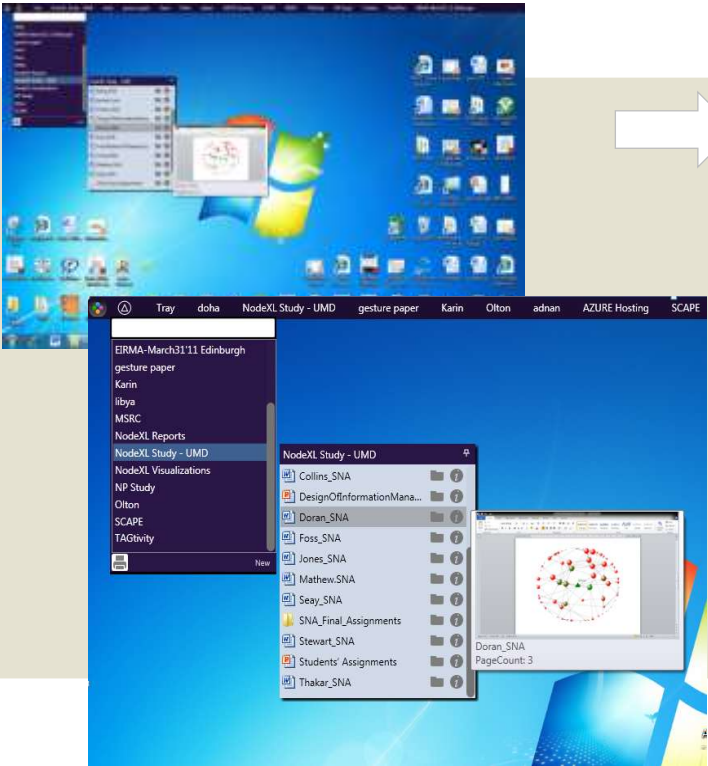
The image shows a presentation slide titled "Analysis of Language Expressions: Paraphrasing in English" from Microsoft Research. The slide features a URL <http://labs.microsofttranslator.com/thesaurus/> and a diagram of a neural network. A file explorer window is overlaid on the slide, showing a list of files including "SNA_Final_Assignments", "Stewart_SNA", "Doran_SNA", "Foss_SNA", "Jones_SNA", "Seay_SNA", "burles1_sna", "Collins_SNA", "Mathew_SNA", "Albrittain_SNA", "Bang_SNA", "Thakar_SNA", "Students' Assignments", "<File not found>", "DesignOfInformationMa...", and "NodeXL Study - UMD". The "Doran_SNA" file is selected, and a preview window shows a network graph with the text "Doran_SNA PageCount: 3". The presentation is on "Slide 4 of 27" and the system language is set to "English (U.K.)".

GATHERING content across application and storage silos

CONTEXTUALIZING – creating representations of data while authoring, sense making, etc.

DESKTOP TAGGING FACILITY

COLLECTION BROWSLETS



Crowdsourcing of Human Input and Feedback

amazonmechanicalturk Artificial Intelligence

Introduction | Dashboard | Status | Account Settings

Mechanical Turk is a marketplace for work.
We give businesses and developers access to an on-demand, scalable workforce. Workers select from thousands of tasks and work whenever it's convenient.
129,713 HITs available. [View them now.](#)

Make Money by working on HITs

HITs - *Human Intelligence Tasks* - are individual tasks that you work on. [Find HITs now.](#)

As a Mechanical Turk Worker you:

- Can work from home
- Choose your own work hours
- Get paid for doing good work

Find an interesting task → **Work** → **Earn money**

[Find HITs Now](#)

Get Results from Mechanical Turk Workers

Ask workers to complete HITs - *Human Intelligence Tasks* - and get results using Mechanical Turk. [Register Now](#)

As a Mechanical Turk Requester you:

- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITs completed in minutes
- Pay only when you're satisfied with the results

Fund your account → **Load your tasks** → **Get results**

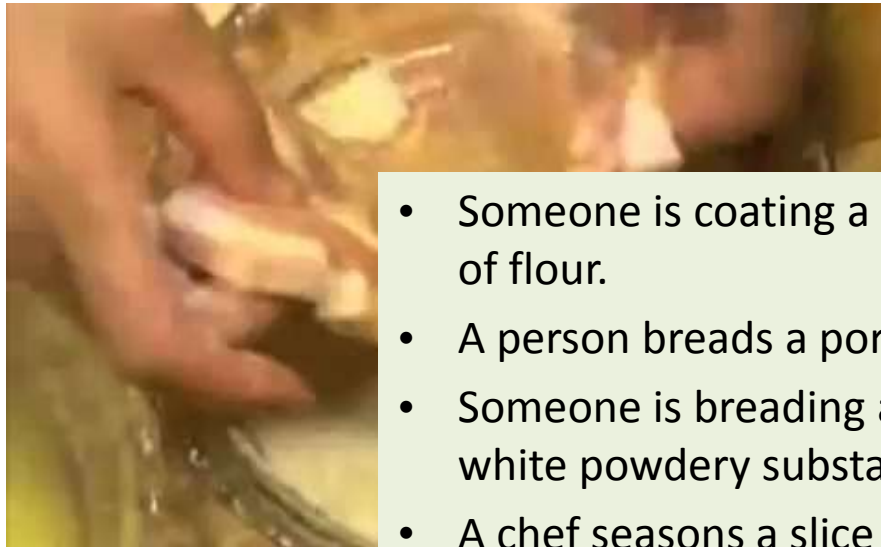
[Get Started](#)

Designing HITs to Collect the Meaning



- Task: describe the video in one sentence
- Use the collected sentence to train the paraphrasing algorithms

Designing HITs to Collect the Meaning



- Task: describe
- Use the collected paraphrasing

- Someone is coating a pork chop in a glass bowl of flour.
- A person breads a pork chop.
- Someone is breading a piece of meat with a white powdery substance.
- A chef seasons a slice of meat.
- Someone is putting flour on a piece of meat.
- A woman is adding flour to meat.
- A woman is coating a piece of pork with breadcrumbs.
- A man dredges meat in bread crumbs.
- A person breads a piece of meat.
- A woman is breading some meat.

Summary

- The challenge is not in the standards for representations and pattern discovery but in the interpretation and validation of that interpretation.
- ‘Meaning’ has different connotations in different context
 - The difficulty is determining and addressing the right level of granularity.



- The meaning of things may escape us by its very nature. It is relative, subjective, volatile, under-defined.

Users adapt to and adopt technology and its result. They reuse and apply it in unpredictable ways, finding utility where designers have not seen it.

Thank you!

Contact:

Natasa Milic-Frayling

Integrated Systems

Microsoft Research Ltd.

Cambridge, U.K.

natasamf@microsoft.com

<http://research.microsoft.com/is>