COLLEGE OF INFORMATION STUDIES

PopIT: Scalable Computational Analysis of the Diffusion of Technological Innovations

Ping Wang
Kenneth R. Fleischmann
Douglas W. Oard

Outline

- Who Project team and sponsors
- When History and project duration
- Why Project goals
- What Three thrusts of research
- How Interdisciplinary collaboration
- What Publications and working papers
- Where Next steps, locations, contact info

Who Are We?


Ping Wang
PI


Ken Fleischmann
Co-PI


Doug Oard
Co-PI


Philip Resnik
Senior Personnel


Emi Ishita
Visitor-Japan


Xin Jin
Visitor-China


An-Shou Cheng
Graduate Student


Asad Sayeed
Graduate Student


Chia-jung Tsui
Graduate Student


Tiffany Chao
Undergrad Student


Ederlyn Lacson
Undergrad Student


Samer Faraj
Advisor
McGill University


Rob Fichman
Advisor
Boston College


Burt Swanson
Advisor
UCLA


Allison Brochu
Alumna


Lidan Wang
Alumna


Yejun Wu
Alumnus

PopIT History

- 2006-11-22: Ideas at Thanksgiving dinner
- 2007-02-21: Proposal submitted
- 2007-09-15: Project funded by NSF
- 2007 Fall: eInnovation reading group
- 2007-12-13: Project officially launched
- 2010-08-31: Project ends (estimated)

PopIT Goals

- To understand the dynamic social system and processes underlying the development, diffusion, and use of IT innovations
  - Popularity of IT innovations (Wang)
  - Values and IT (Fleischmann)
- To integrate computational analysis of text with theory building and testing in social science research
  - 4th Generation Content Analysis (Oard/Resnik)

IT Innovation Is …

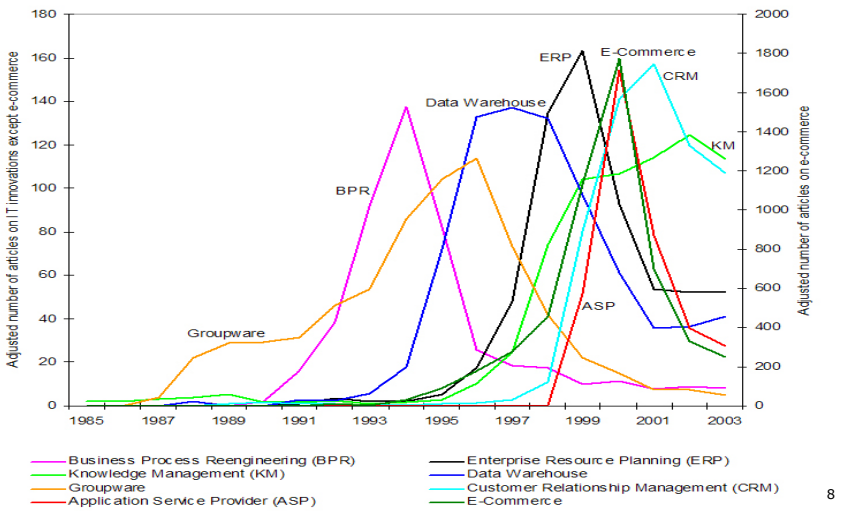
- Information technology perceived as new by adopter
- Vital to economy and society, especially in tough times (recession is fertile ground for innovations)
- Developed, disseminated, and used in and by social system and processes
  - Social structure: network, community, hierarchy
  - Social cognition: value, sentiment, learning

Portable Personality
SaaS
Chatbots
Ajax
RFID
Ultramobile Devices
BPO
Application Quality Dashboards
SOA
VoIP
Mashup
Identity Management
DRM
OSS
Thin Provisioning
Business Intelligence
Semantic Web
Cloud
SCM
Tera-architectures
Computing
Web2.0
Distributed Encryption
CRM

Why do some innovations become popular, but others don't?

Popularity of IT Innovations

IT Popularity



Popularity of IT Innovations

- Popular innovations address popular problems
- Differentiated social status in the networks of actors partially explains the popularity and success of innovations
- Different relationships among innovations have different influences on the trajectories of innovations
- Different social groups take turns to serve as opinion leaders either promoting or denouncing an innovation during the course of its life cycle

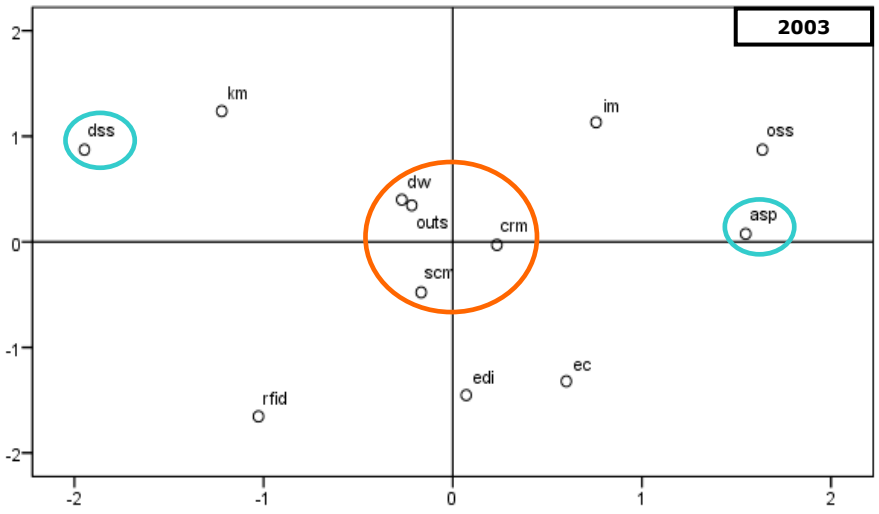
Wang, 2008; Wang & Swanson, 2008; Wang, 2009

IT Innovation Discourse

- Discourse both reflects and constructs reality
- Similar innovations → similar discourse → related popularity?
- Data: *InformationWeek* articles (1998-2007) Lexis/Nexis
- Discourse content
  - Apply Kullback–Leibler (KL) divergence to measure the difference between innovation discourses
  - Apply multidimensional scaling (MDS) to the KL divergence matrix for visualization
- Volume
  - Tally articles assigned to each subject in Lexis/Nexis

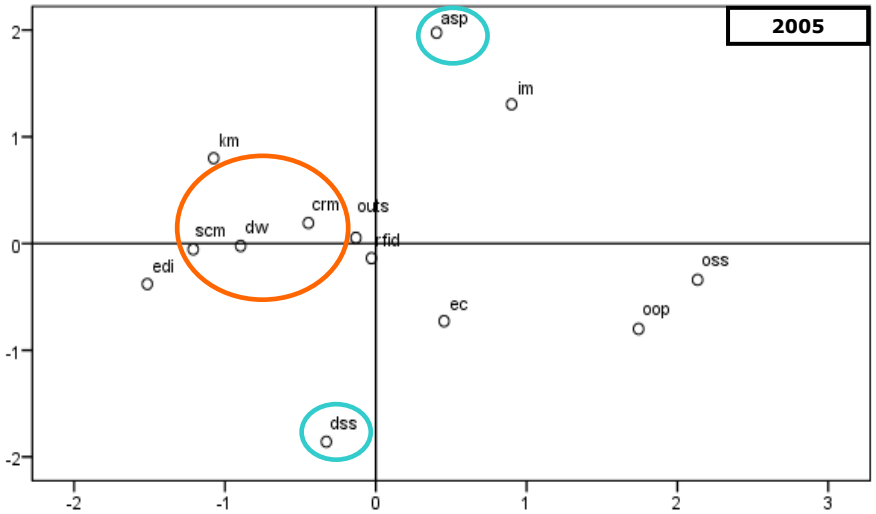
Discourse Content Similarity

IT Popularity



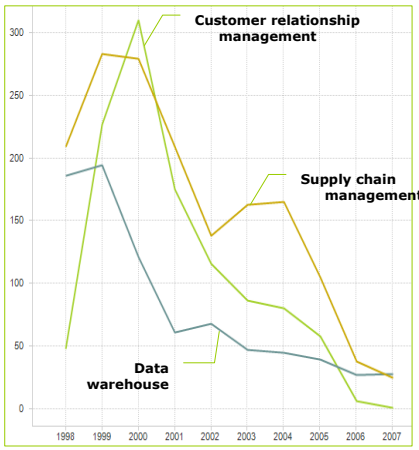
Discourse Content Similarity

IT Popularity

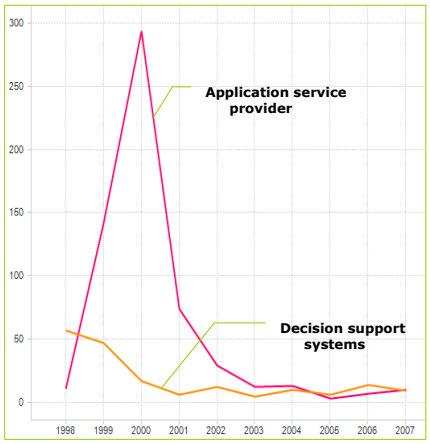


Popularity (Discourse Volume)

IT Popularity



H1: Trajectories of similar innovations are correlated



H2: Trajectories of dissimilar innovations are uncorrelated

Content Analysis

4G Content Analysis

- "… any technique for making inferences by objectively and systematically identifying specified characteristics of messages …" (Holesti, 1969)
- "… the study of recorded human communications such as books, Web sites, paintings, and laws …" (Babbie, 1975)
- "… a summarizing, quantitative analysis of messages that relies on the scientific method …" (Neuendorf, 2002)

Content Analysis Generations

4G Content Analysis

G1 Read and understand something

G2 Manually infer something, then count it

G3 Directly observe something, then count it

G4 Automatically infer something, then count it

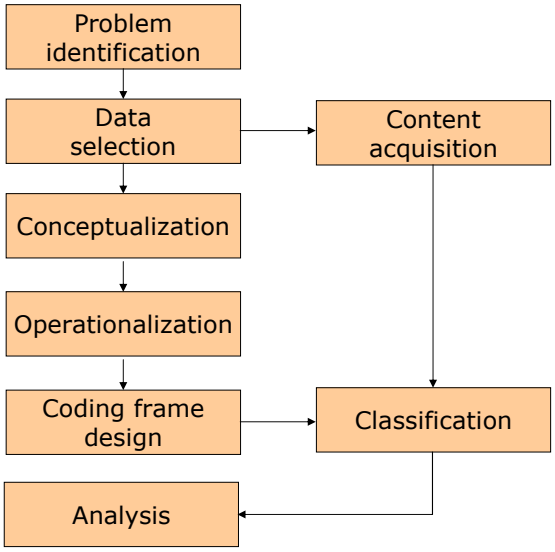
Build Tools, not "Solutions"

4G Content Analysis

Cartoon by Carroll Zahn, Original Artist

2G Content Analysis

4G Content Analysis

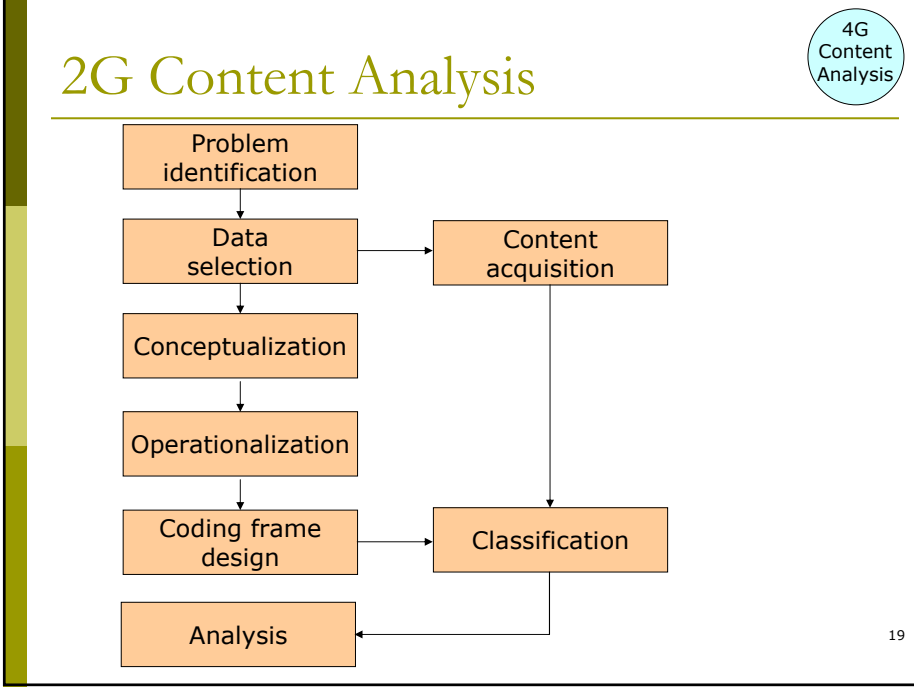


3G Content Analysis

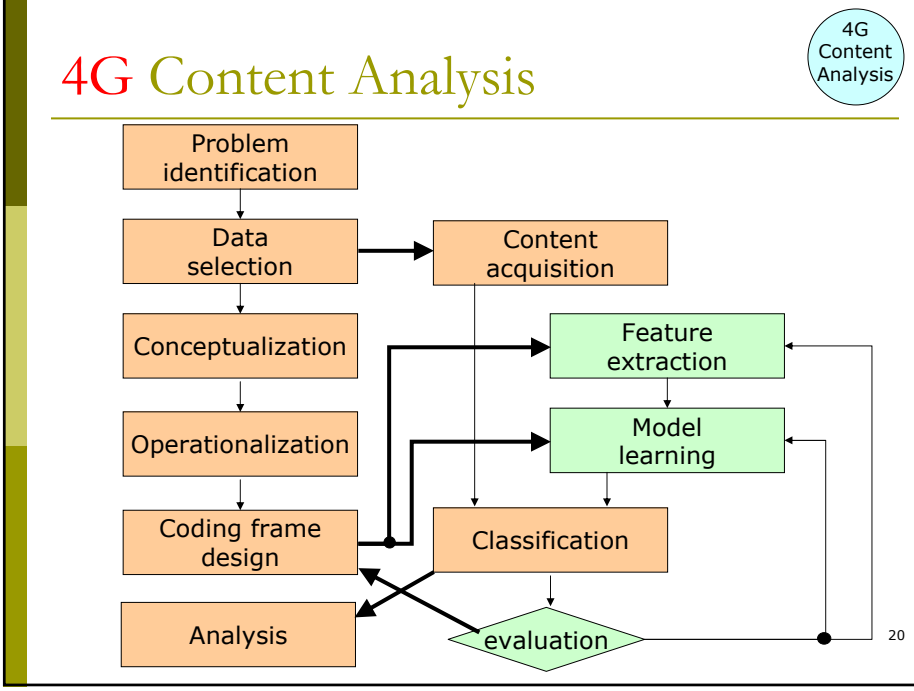
4G Content Analysis

- Computer-Assisted Text Analysis (CATA)
- "Dictionary-based" word counting
  - Person or organization names
  - Positive and negative sentiment terms
- Vastly more scalable than manual coding
  - Alias list can accommodate synonymy
  - Focused domain can limit homonomy effects
  - Regression models context-dependent effects

2G Content Analysis

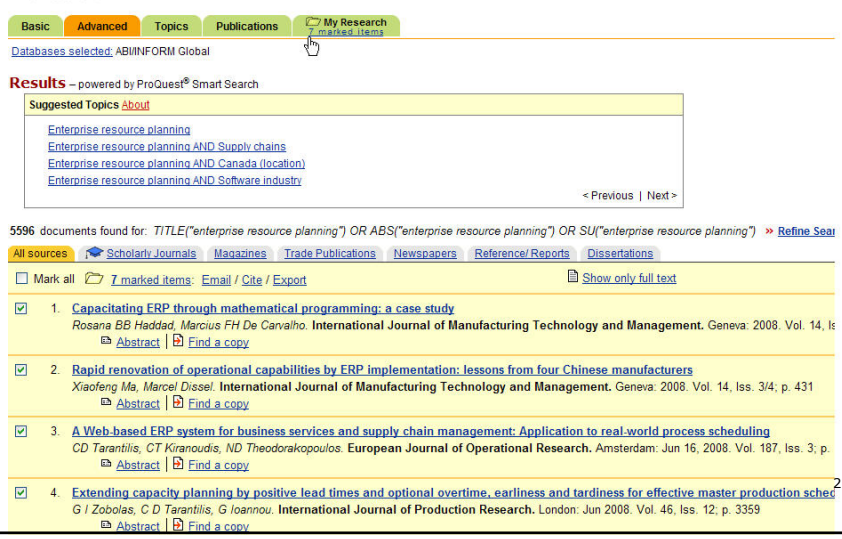


4G Content Analysis



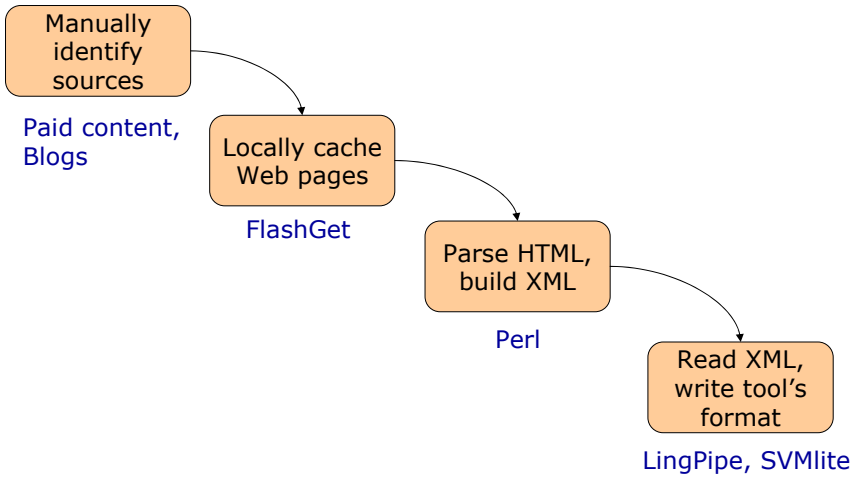
Manual Content Acquisition

4G Content Analysis



Automated Content Acquisition

4G Content Analysis



Trade Press Magazines

4G Content Analysis

- 6-month Pilot Study Collection
  - Computerworld — Jan 2005–Jun 2005
    - 1,193 documents
    - 26 issues
- 10-year Collection
  - Computerworld — Jan 1998–Jun 2008
    - 25,278 documents
    - 534 issues
  - InformationWeek: — Jan 1998–Jun 2008
    - 31,112 documents
    - 527 issues

Full LexisNexis Magazine Runs

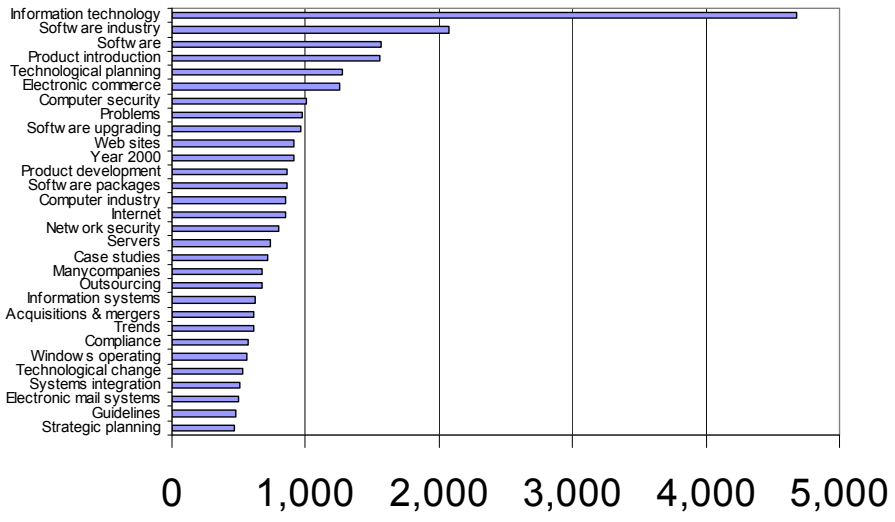
4G Content Analysis

3 genres, 2 pubs/genre, 1988-2008 (unless noted)

- General News
  - *Newsweek*
  - *U.S. News & World Report*
- Business News
  - *BusinessWeek*
  - *The Economist*
- IT Trade Press
  - Computerworld
  - InformationWeek (1991-2008)

ProQuest Subject Labels (2G)

4G Content Analysis



10-year Computerworld collection

Term-List Classifier (3G)

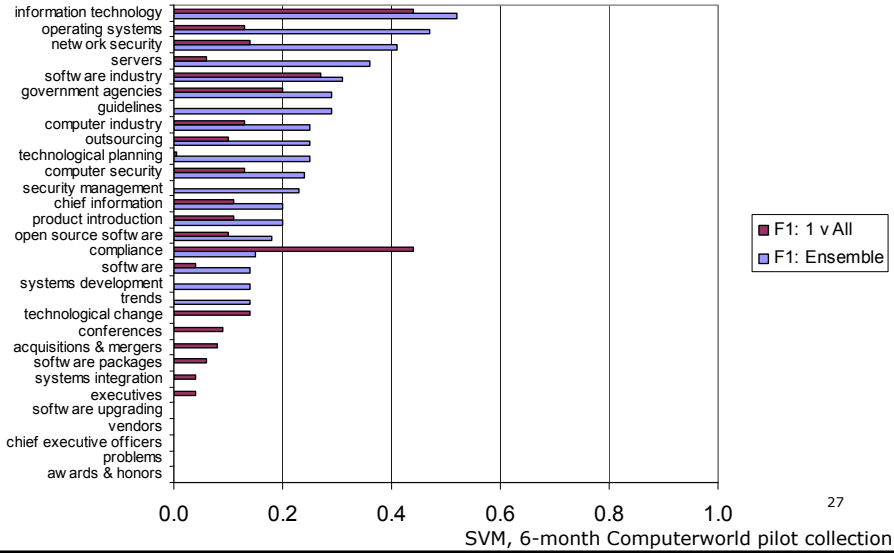
4G Content Analysis

	A	B	C	D
1	Label	Full Name	Alias 1	Alias 2
2	AI	artificial intelligence		
3	BPR	business process reengineering	business-process reengineering	
4	CAD	computer-aided design	computer aided design	
5	CAI	computer-assisted instruction	computer assisted instruction	
6	CAM	computer-aided manufacturing	computer aided manufacturing	
7	CASE	computer-aided software engineering	computer aided software engineering	
8	CIM	computer-integrated manufacturing	computer integrated manufacturing	
9	clntserv	client server	client-server	
10	datamine	data mining	data-mining	datamining
11	datawrhs	data warehouse	data-warehousing	data-warehouse
12	DBMS	database management system	data-base management system	database-manage
13	deskpub	desktop publishing		
14	DSS	decision support system	decision-support system	
15	ecom	e-commerce	electronic commerce	electronic-comme
16	EDI	electronic data interchange	EDI	
17	EFT	electronic funds transfer	electronic-funds transfer	
18	EIS	executive information system		
19	enducomp	end-user computing	end user computing	
20	ERP	enterprise resource planning		
21	expertsy	expert system	expert-system	
22	GDSS	group decision support system		
23	GIS	geographic information system		
24	grupware	groupware	group-ware	
25	GUI	graphical user interface		
26	intranet	intranet		
27	ISP	internet service provider		
28	ITF	integrated test facility		
29	JAD	joint application development		
30	JCL	job control language		

Coding frame: ProQuest subject labels

SVM Classifier (4G)

4G Content Analysis



SVM, 6-month Computerworld pilot collection

Automating Annotation

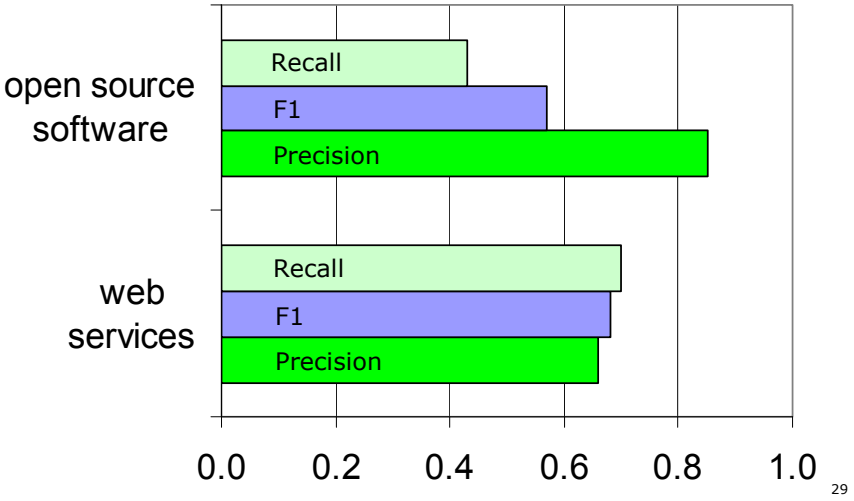
4G Content Analysis

"There has been a lot of buzz over the arrival of Firefox, the open-source browser published by the Mozilla Foundation… Sun Microsystems Inc. hopes that open-source Solaris will draw in new users and new growth opportunities."

Segmentation:	Classification:	Association and clustering: *(company, software)*
Firefox	Open Source Software	Firefox
Mozilla Foundation	Organization	Mozilla Foundation
Sun Microsystems	Organization	Sun Microsystems
Solaris	Open Source Software	Solaris

Automatic Mention Annotation

4G Content Analysis



LingPipe, 6-month Computerworld pilot collection

Values in IT Design, Use & Policy

Values & IT

- Study the role of values in
  - Design and use of IT
  - Telecommunication policy
- Survey method is insufficient
- Manual content analysis helps, but …
  - Sentence-level closed-coding of public hearings on net neutrality following the Schwartz Values Inventory (Schwartz 1992)
- Automatic detection and classification affordably enables macro-scale social science research
  - Build and apply value classification systems to study how values influence IT design, use, and policy

Fleischmann 2007; Fleischmann, Oard, Cheng, Wang, & Ishita, Working Paper; Cheng & Fleischmann, Working Paper

Manual Coding for Training

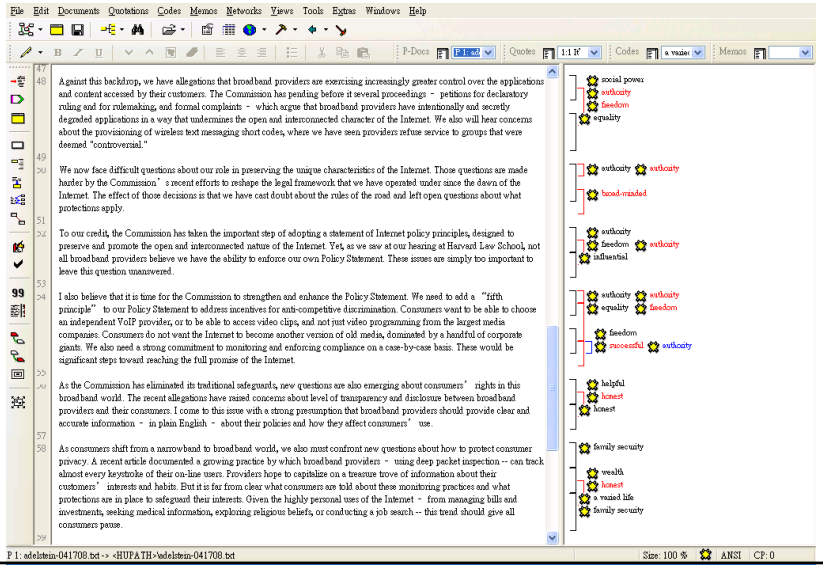
Values & IT

Against this backdrop, we have allegations that broadband providers are exercising increasingly greater control over the applications and content accessed by their customers. The Commission has pending before it several proceedings – petitions for declaratory ruling and for rulemaking, and formal complaints – which argue that broadband providers have intentionally and secretly degraded applications in a way that undermines the open and interconnected character of the Internet. We also will hear concerns about the provisioning of wireless text messaging short codes, where we have seen providers refuse service to groups that were deemed "controversial."

We now face difficult questions about our role in preserving the unique characteristics of the Internet. Those questions are made harder by the Commission's recent efforts to reshape the legal framework that we have operated under since the dawn of the Internet. The effect of those decisions is that we have cast doubt about the rules of the road and left open questions about what protections apply.

Recording Codes in ATLAS.ti

Values & IT



Weka "Training Data"

Values & IT

17	A0-Authority	We also need a strong commitment to monitoring and enforcing compliance on a case-by-case basis.
18	A1-Successful	These would be significant steps toward reaching the full promise of the Internet.
19	C1-Helpful	As the Commission has eliminated its traditional safeguards new questions are also emerging about consumers rights in this broadband world.
20	C1-Honest	The recent allegations have raised concerns about level of transparency and disclosure between broadband providers and their consumers.
21	C1-Honest	I come to this issue with a strong presumption that broadband providers should provide clear and accurate information in plain English about their policies and how they affect consumers use.
22	D2-Family security	As consumers shift from a narrowband to broadband world we also must confront new questions about how to protect consumer privacy.
23	A0-Wealth	Providers hope to capitalize on a treasure trove of information about their customers interests and habits.
24	C1-Honest	But it is far from clear what consumers are told about these monitoring practices and what protections are in place to safeguard their interests.
25	B1-A varied life	Given the highly personal uses of the Internet from managing bills and investments seeking medical information exploring religious beliefs or conducting a job search -- this trend should give all consumers pause
26	D2-Family security	Given the highly personal uses of the Internet from managing bills and investments seeking medical information exploring religious beliefs or conducting a job search -- this trend should give all consumers pause

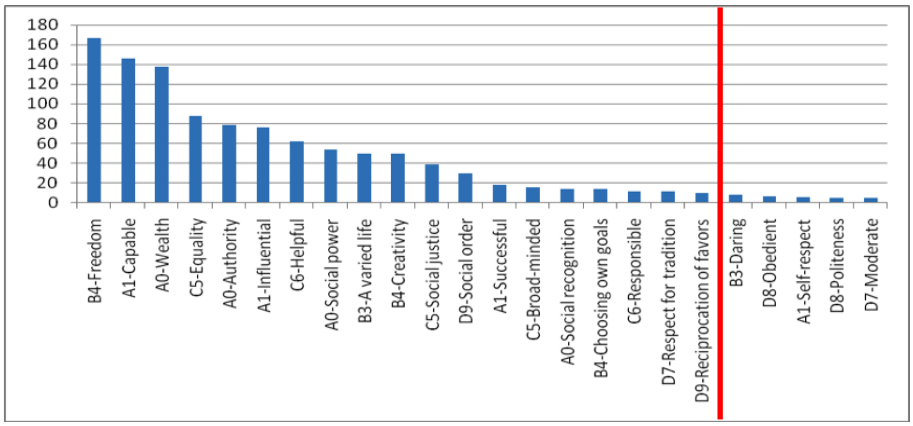
2nd-level Class Distribution

Values & IT

1st Level	2nd Level	Sentences
Self-enhancement (A)	Power (A0)	288
	Achievement (A1)	250
	Hedonism (A2)	2
Openness to change (B)	Stimulation (B3)	58
	Self-direction (B4)	234
Self-transcendence (C)	Universalism (C5)	148
	Benevolence (C6)	85
Conservation (D)	Tradition (D7)	17
	Conformity (D8)	12
	Security (D9)	48

3rd-Level Class Distribution

Values & IT



Classification Accuracy

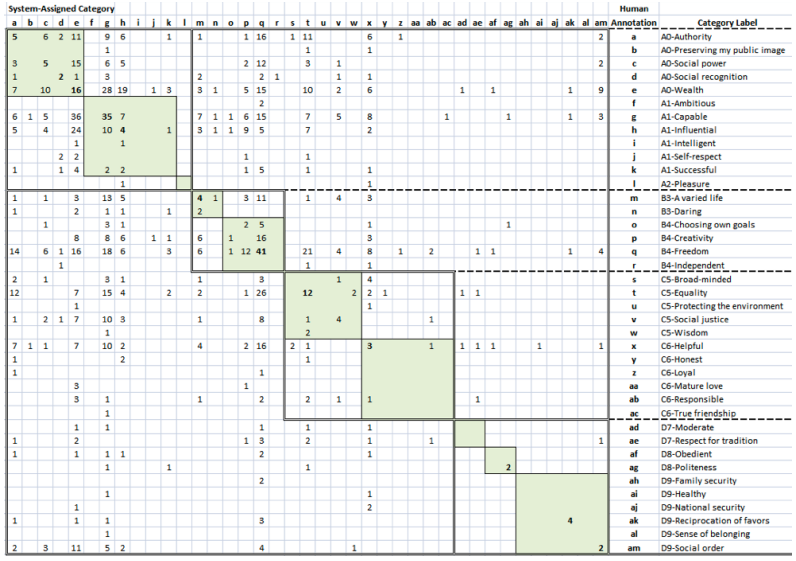
Values & IT

Overall accuracy for 3rd-level classifiers

1NN	25.88%
3NN	37.09%
5NN	38.64%
7NN	38.64%
10NN	37.09%
13NN	36.26%
15NN	35.84%

SVM Confusion Matrix

Values & IT



System-Assigned Category																																								Human Annotation	Category Label
a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	ab	ac	ad	ae	af	ag	ah	ai	aj	ak	al	am		
5		6	2	11		9	6			1		1			1	16		1	11				6	1														2		a	A0-Authority
						1													1				1																b	A0-Preserving my public image
3		5		15		6	5								2	12			3		1																2	c	A0-Social power
1			2	1		3						2				2	1			1		1																d	A0-Social recognition
7		10		16		28	19		1	3		3	1		5	15			10		2		6						1		1					1		9	e	A0-Wealth
																2																							f	A1-Ambitious
6	1	5		36		35	7					7	1	1	6	15			7		5		8					1				1					1		3	g	A1-Capable
5		4		24		10	4			1		3	1	1	9	5			7				2																h	A1-Influential
				1			1																																i	A1-Intelligent
			2	2											1				1																				j	A1-Self-respect
1			1	4		2	2								1	5			1				1																k	A1-Successful
							1																1																l	A2-Pleasure
1		1		3		13	5					4	1		3	11			1		4		3																m	B3-A varied life
1				2		1	1			1		2																												n	B3-Daring
		1				3	1								2	5							1								1								o	B4-Choosing own goals
				8		8	6		1	1		6		1		16							3																p	B4-Creativity
14		6	1	16		18	6			3		6		1	12	41			21		4		8		1		2			1	1					1		4	q	B4-Freedom
			1																1				1																r	B4-Independent
2		1				3	1					1				3				1			4																s	C5-Broad-minded
12				7		15	4			2		2			1	26			12			2	2	1				1	1										t	C5-Equality
				1																			1																u	C5-Protecting the environment
1		2	1	7		10	3					1				8			1		4							1										v	C5-Social justice
						1													2																				w	C5-Wisdom
7	1	1		7		10	2					4			2	16		2	1				3					1		1	1	1		1			1	x	C6-Helpful
1							2												1																				y	C6-Honest
1																1																							z	C6-Loyal
																																							aa	C6-Mature love
				3											1																								ab	C6-Responsible
				3		1						1				2			2		1		1							1									ac	C6-True friendship
						1																																		
				1		1										1			1				1																	ad	D7-Moderate
1				2											1	3			2				1				1											1	ae	D7-Respect for tradition
1				1		1	1									2							1																af	D8-Obedient
						1				1									1											2									ag	D8-Politeness
																2																							ah	D9-Family security
						1																	1																ai	D9-Healthy
				1																			2																aj	D9-National security
1				1		1										3																				4				ak	D9-Reciprocation of favors
						1																																		al	D9-Sense of belonging
2		3		11		5	2									4						1																2	am	D9-Social order

Ishita, Cheng, Oard, Fleischmann & Resnik, Working Paper

Future Directions

Values & IT

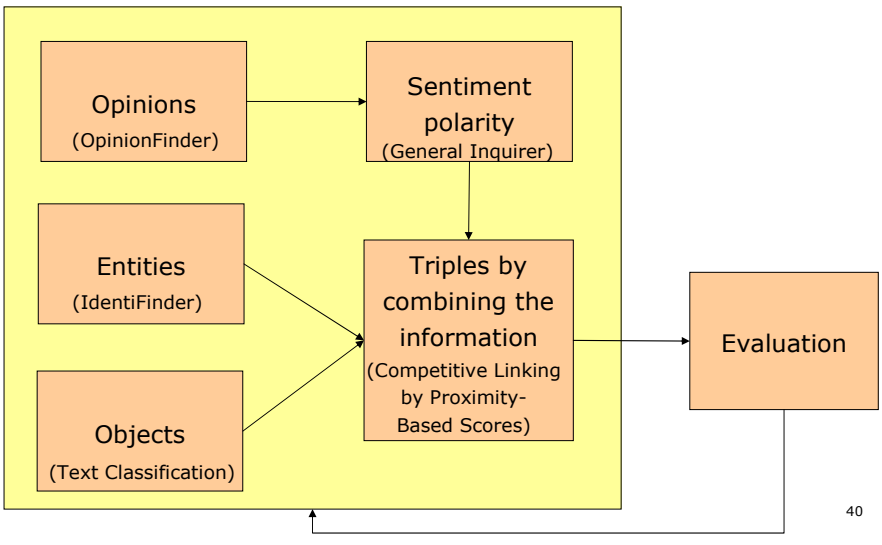
- More and larger collections
  - All congressional testimonies
  - NASA oral history archive
  - Trade press
  - Academic journals
  - Blogs
- Tighter integration
  - Keep human coder in the loop
  - Integrate SML and actor-network theory

Fleischmann, Oard, Cheng, Wang, & Ishita, Working Paper; Templeton, Fleischmann, & Oard, Working Paper

What's Next for PopIT?

- Improved accuracy
  - Hierarchical multi-label classifiers
- Additional content types
  - Academic journal articles
  - Blogs
  - Interviews
- Classification
  - Cross-domain (e.g., trade press : blogs)
  - Non-topical (e.g., sentiment)
  - Social network (e.g., opinion leaders)
- Integrated systems

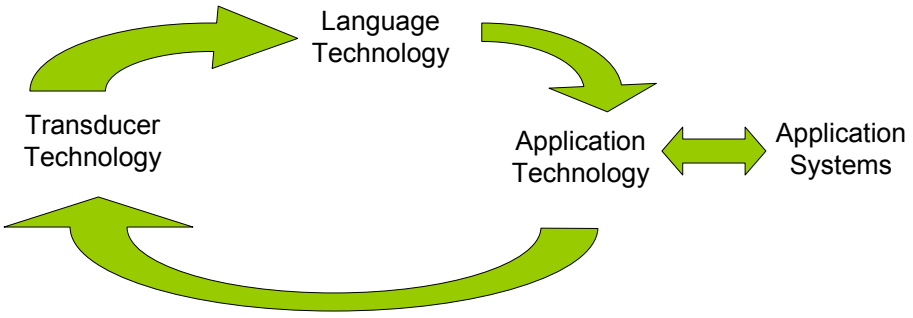
Sentiment Analysis Plans



Spanning Disciplines

- Coupled models as "boundary object"
  - Input representation
  - Transformation
  - Output representation
- Layered uncertainty
  - Meaning of the text
  - Meaning of the coding frame
  - Purpose of the coding frame

Interdisciplinary Innovation Cycle



Publications

Journal Articles

- Wang, P. "Popular Concepts beyond Organizations: Exploring New Dimensions of Information Technology Innovations," *Journal of the Association for Information Systems* (12:2), 2009, pp. 1-30.
- Wang, P. & Swanson, E. B. "Customer Relationship Management as Advertised: Exploiting and Sustaining Technological Momentum," *Information Technology and People* (21:4), 2008, pp. 323-349.

Dissertation

- Wu, Y. "Classifying Attitude by Topic Aspect for English and Chinese Document Collections," Ph.D. Dissertations. University of Maryland, College of Information Studies, 2008.

Book Chapters

- Oard, D.W. (in press). A Whirlwind Tour of Automated Language Processing for the Humanities and Social Sciences," in *Promoting Digital Scholarship: Formulating Research Challenges in the Humanities, Social Sciences, and Computation*, Council on Library and Information Resources.
- Wang, P. "Whatever Happened to BPR? The Rise, Fall, and Possible Revival of Business Process Reengineering: From the Organizing Vision Perspective," in *Business Process Transformation*, V. Grover & M.L. Markus (eds.). M.E. Sharpe, Armonk, NY, 2008a, pp. 23-40.

Publications (continued)

Conference Papers

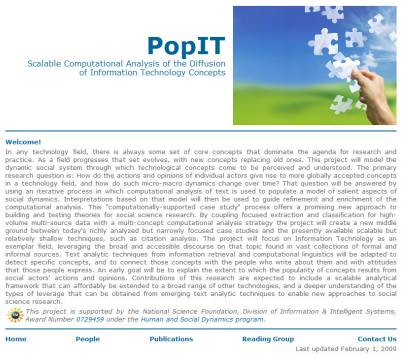
- Cheng, A-S., Fleischmann, K.R., Wang, P., & Oard, D.W. "Advancing Social Science Research by Applying Computational Linguistics," in *Proceedings of the 2008 Annual Meeting of the American Society for Information Science and Technology*. Columbus, OH, 2008.
- Elsayed, T., Lin, J., & Oard, D.W. "Pairwise document similarity in large collections with MapReduce," in *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics*, Companion Volume. Columbus, OH, 2008.
- Wang, P. "Assimilating Information Technology Innovation: The Longitudinal Effects of Institutionalization and Resource Dependence," in *Proceedings of Twenty-Ninth International Conference on Information Systems*. Paris, France, 2008b.
- Wang, P., Tsui, C.-j., Fleischmann, K. R., Oard, D. W., & Wang, L. "Understanding IT Innovations through Discourse Analysis," poster presented at the *Fourth iConference*. Chapel Hill, NC, 2009.
- Wu, Y. & Oard, D. "Bilingual topic aspect classification with a few training examples," in *Proceedings of the 31st annual international ACM SIGIR Conference on Research and Development in Information Retrieval*. Singapore, 2008, pp. 203-210.

Working Papers

- Cheng, A.-S. & Fleischmann, K.R. Value perspectives in net neutrality: A content analysis of public hearings.
- Ishita, E., Cheng, A.-S., Oard, D.W., Fleischmann, K.R., & Resnik, P. Detecting and Classifying Expressions of Human Values.
- Fleischmann, K.R., Oard, D.W., Cheng, A.-S., Wang, P., & Ishita, E. Automatic Classification of Human Values: Applying Computational Thinking to Information Ethics
- Templeton, T.C., Fleischmann, K.R., & Oard, D.W. Using Automated Text Classification with Actor-Network Theory.

Where Are We?

4120D Hornbake Bldg.
University of Maryland
College Park



http://terpconnect.umd.edu/~pwang/PopIT/

Acknowledgement and Contact

- This work is supported by the National Science Foundation under Grant IIS-0729459.

- Ping Wang pwang@umd.edu
- Ken Fleischmann kfleisch@umd.edu
- Doug Oard oard@umd.edu

College of Information Studies
University of Maryland
Room 4105 Hornbake Bldg, South Wing
College Park, MD 20742-4325

COLLEGE OF INFORMATION STUDIES

Thank You from PopIT Team