Carnegie Mellon University

Discovering a Natural Language Semantics for Privacy

Travis Breaux

Science of Security, Quarterly Meeting

North Carolina State University, February 2, 2017

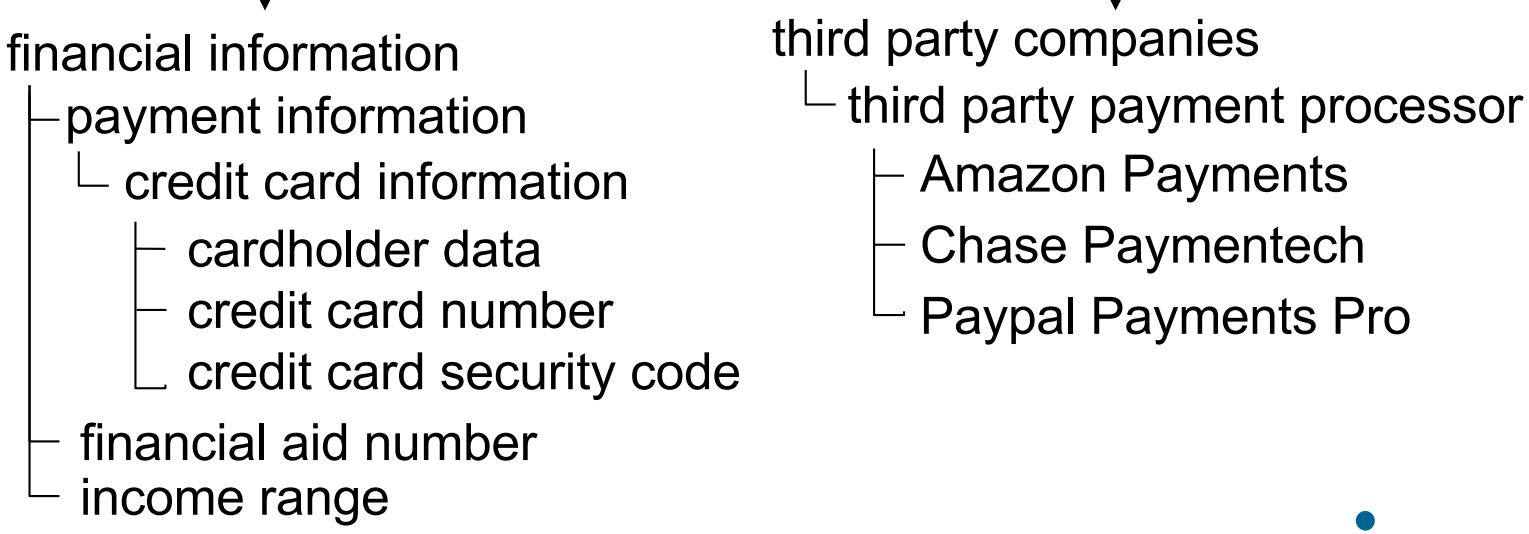
institute for SOFTWARE RESEARCH

How the text ***reads***

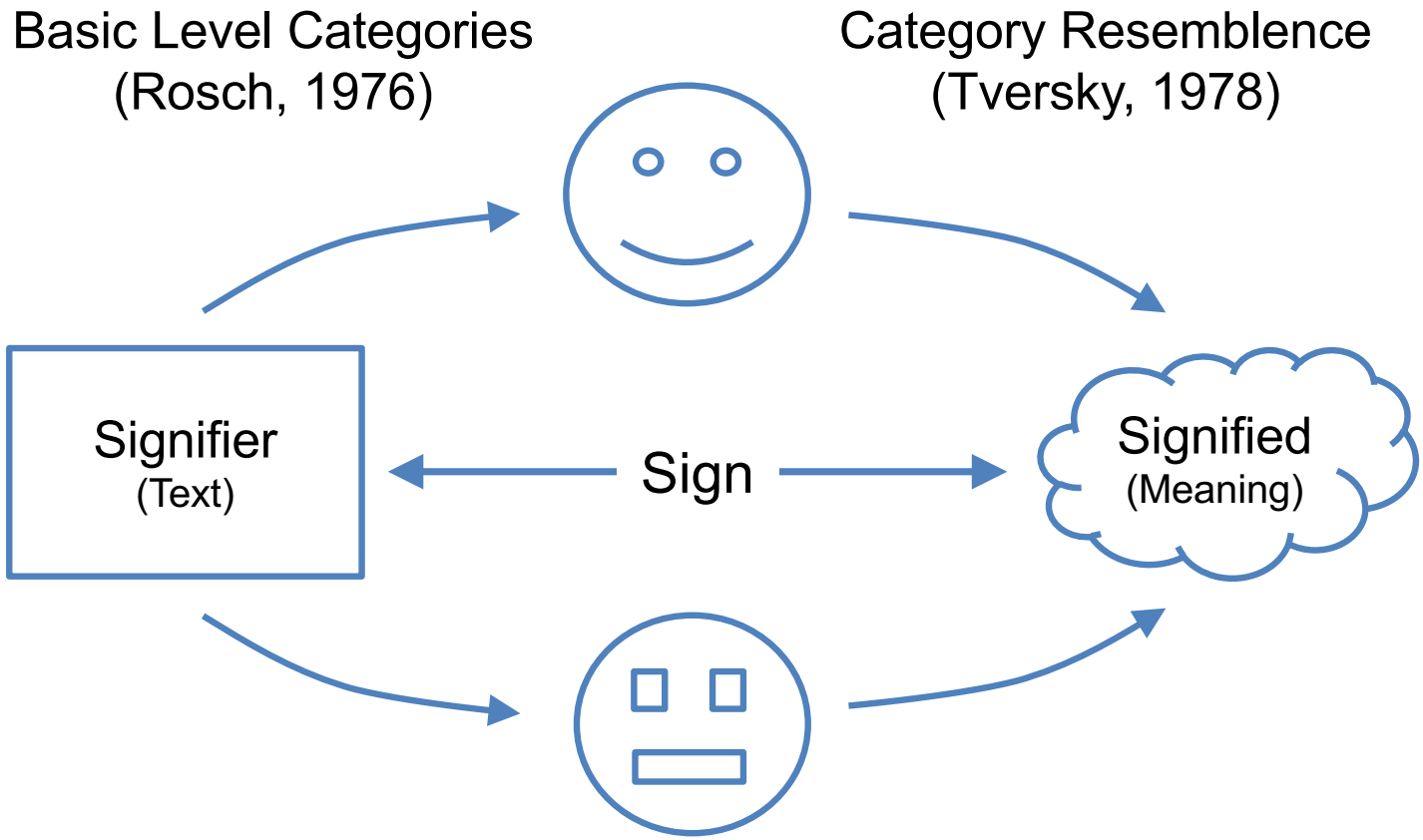
We will provide your information to third party companies to perform services on our behalf, including payment processing, data analysis, e-mail delivery, hosting services, customer service and to assist us in our marketing efforts.

What the text ***means***

We will provide your information to third party companies to perform services on our behalf, including payment processing, data analysis, e-mail delivery, hosting services, customer service and to assist us in our marketing efforts.



What do we mean by semantics?



Ferdinand de Saussure, *Course in General Linguistics*, 1916

What kinds of information exist?

What kinds of information exist?

Contact Information

- Address book
- Billing address
- Contact list
- Email header information
- Home address
- Instant messaging address
- Postal address
- Primary email address
- SMS address
- Subscriber information
- Telemarketing lists

Financial Information

- Credit card security codes
- Credit score
- Debit card PINs
- Debt
- Financial aid number
- Monthly bill
- Payment information
- Purchase history
- Recent orders

Extracted from 5 domains (gaming, health, news, shopping, telecom)

What kinds of information exist?

Location Information

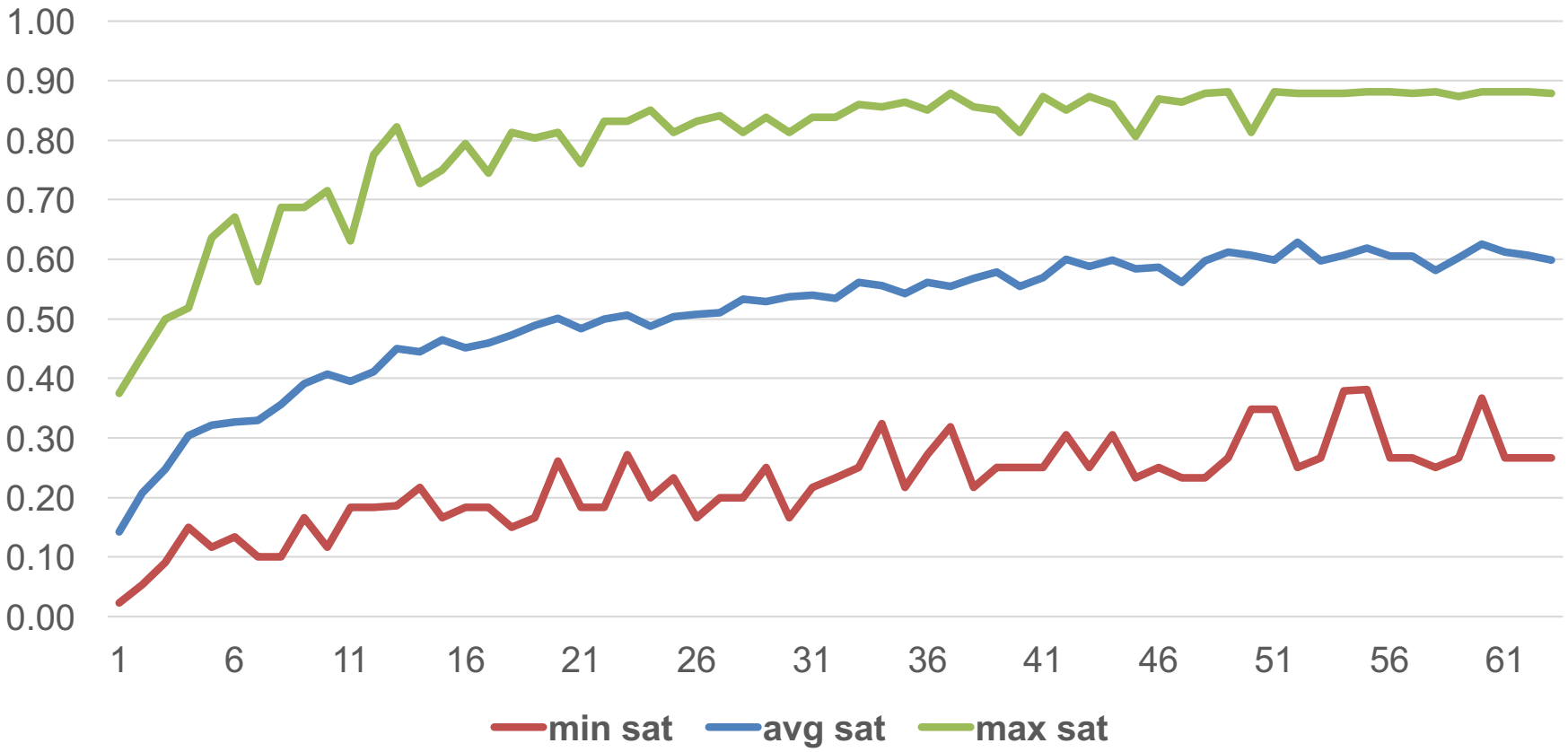
- Cell tower ID
- Foot-traffic data
- GPS geo-location
- GPS location
- IP address
- IP-based geolocation data
- Physical geographic area
- Real-time geographic location information
- ZIP code
- ZIP + 4

Health Information

- Blood glucose
- Disease conditions
- DNA
- Fitness-related activities
- Eye color
- Health status
- Mood icon
- Prescription
- Pharmacist records
- Pulse rate
- Saliva sample

Extracted from 5 domains (gaming, health, news, shopping, telecom)

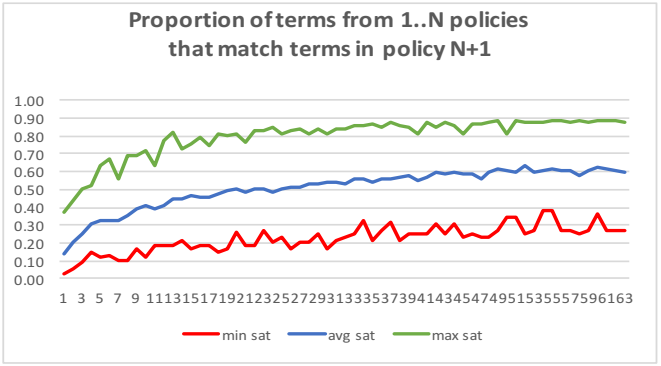
Proportion of information type phrases from 1..N policies that match phrases in policy N+1



Observed over 100 runs with 65 policies in pseudo-random order
3322 unique types, average 51 new types introduced by each new policy

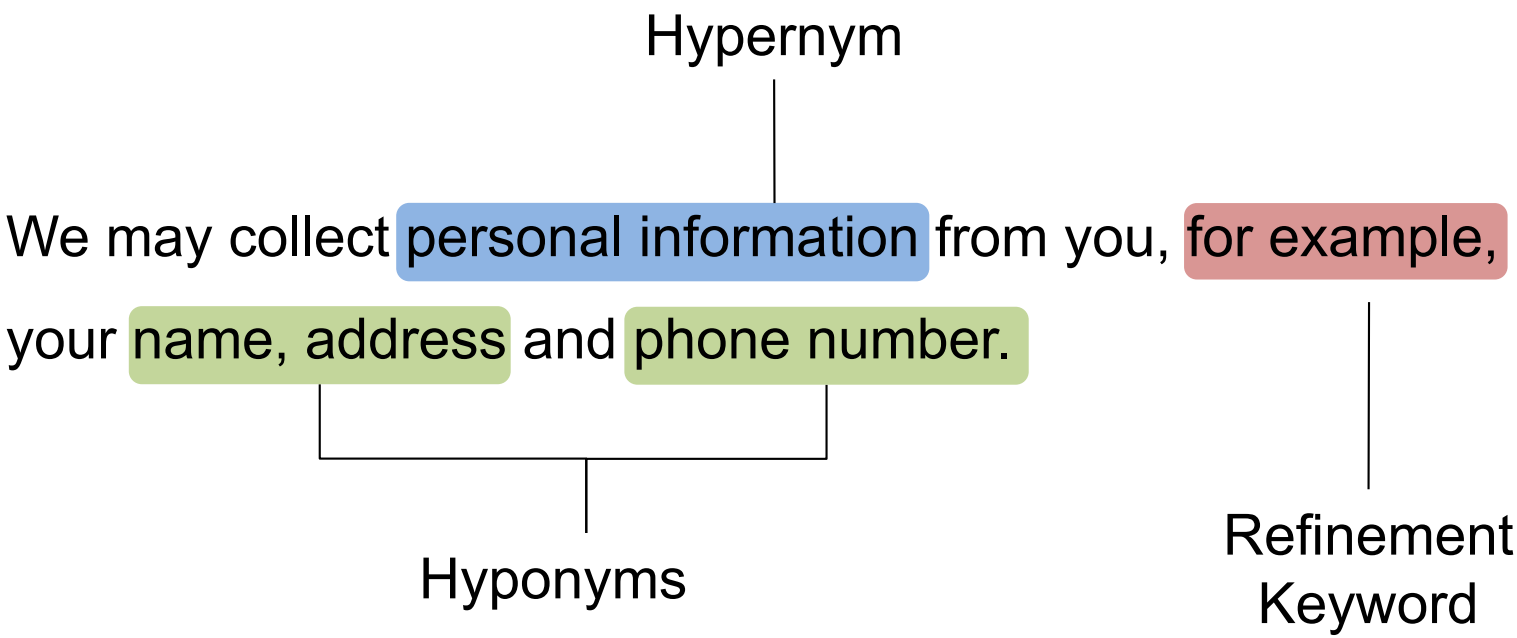
Economics of Ground Truth

- Lexicon built from 65 privacy policies contains 3322 unique information type phrases
- This yields >5.52 million pairwise comparisons to discern relatedness



- For one person, this is 766 days of continuous 24-hour comparison
- With crowdsourcing, we can pay 15 people per comparison, $0.02 each, for a total $1.65M
- The comparison matrix is a sparse matrix

Hyponymy and Hearst Patterns



Excerpt from Barnes and Noble Policy, May 7, 2013.

M. A. Hearst, "Automatic acquisition of hyponyms from large text corpora," *14th Conf. Computational Linguistics*, v. 2, 1992, pp. 539-545.

Example Tregex Pattern

(NP (PRP We))

(VP (MD may)

(VP (VB collect)

(NP (JJ personal) (NN information))

(PP (IN from)

(NP (PRP you)))

(, ,)

(PP (IN for)

(NP

(NP (NN example))

(, ,)

(NP (PRP$ your) (NN name) (, ,) (NN address)

(CC and)

(NN phone) (NN number))))))

This noun phrase (NP) is assigned to the variable "hypernym"

This prepositional phrase describes the keywords that indicate the hyponymy relation

This noun phrase (NP) is assigned to the variable "hyponym"

* The A $ B means "A is a sibling of B" and the A < B means "A immediately dominates B"

Matching Tregex Pattern*

(NP=hypernym $ (IN < for) < (NP< (NN < example)) < NP=hyponym)

Hearst Pattern Study

Shopping	Telecommunications	Social Networking
Barnes and Noble	AT&T	Facebook
Costco	Charter Comm.	Kik
Lowes	Comcast	LinkedIn
Overstock	Time Warner	Snapchat
Walmart	Verizon	WhatsApp

- Above 15 policies yield 1300 unique information type phrases based on two or more crowd worker annotations

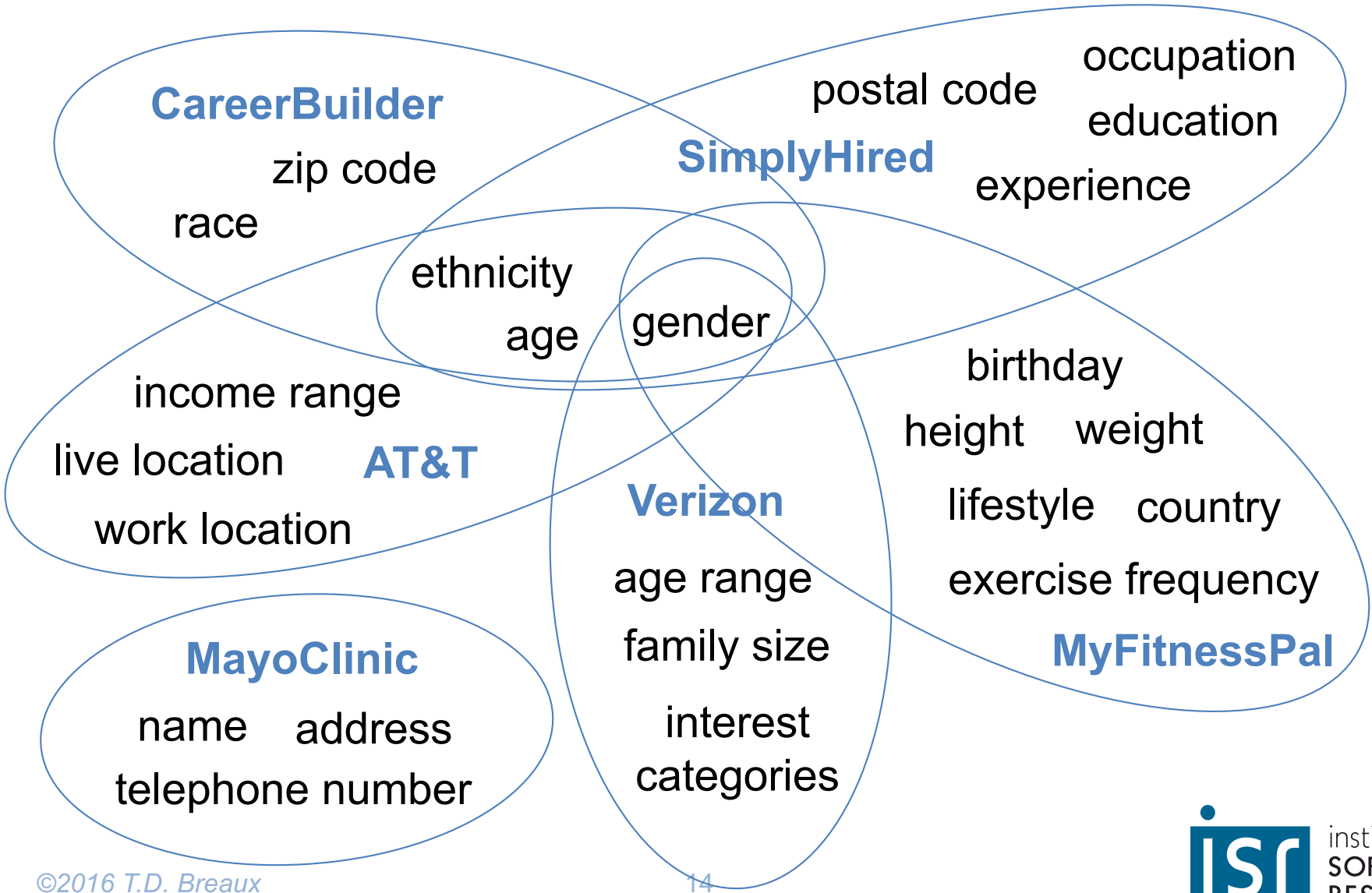
Hearst Pattern Results

Category	Refinement Keywords	Totals
HKO	such as, such, including, for example, include, includes, concerning, is, e.g., like, i.e., of your, contain, relating to, that relates to, generally not including, consists of, concerning, either, (), :, -	195
OKH	(and | or | any | as well as any | certain) other, constitute, as, and any other, is known as, classifies as	34
HO		2
KHO	following types of	1

H: Hypernym, K: Keyword, O: Hyponym

Annotations: Precision = 0.73, Recall = 0.58
Category Pairs: Precision = 0.83, Recall = 0.52
Lexicon Coverage: 23% of 1300 information type phrases

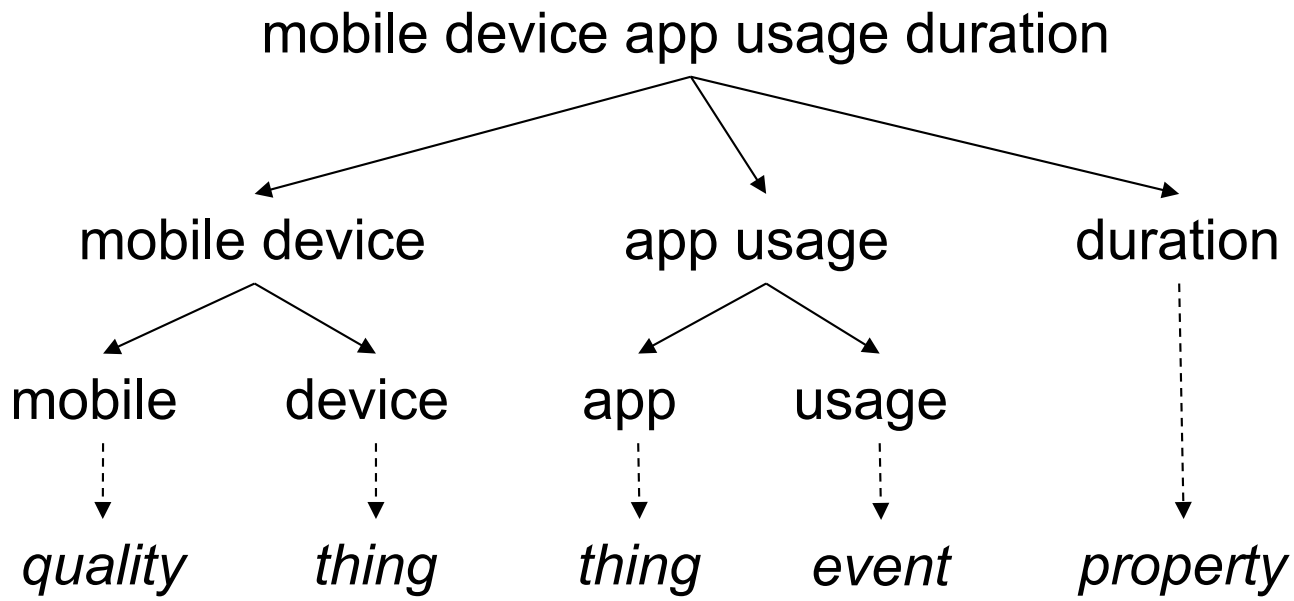
Demographic Information



Multiple ways to describe data

- device-related information
- device-related data
- device-specific information
- device's hardware
- device udid
- device mac address
- hardware device id
- mobile device ids
- mobile mac addresses
- mobile device phone number
- mobile device operating system
- mobile device technical information
- mobile unique device id
- nearby device information
- operating system
- telephone number
- unique identifier

Inferring lexical variants



Can we encode a small amount of semantic information to logically infer word relatedness?

Shallow typing, sematic rules

Step 1: Apply flat typology to phrase:

Q T T E P

mobile device app usage duration

Typology Key:

Q: Quality
T: Thing
P: Property
G: Agent
E: Event

Step 2: Automatically apply rules to typed phrase to generate variants

Definition	Example
QT has-part QTX	mobile device has-part mobile device information
QTX is-kind-of QX	mobile device information is-kind-of mobile information
QTX is-kind-of TX	mobile device information is-kind-of device information
QT has-part T	mobile device has-part app
TE is-kind-of EX	app usage is-kind-of usage information
E has-part P	usage has-part duration

Surveying ontology preferences

1. **browser** : **web browser type** ☐ click to swap word order
   - ○ is a part of
   - ○ is a kind of
   - ○ is equivalent to
   - ○ is unrelated to
   - ○ unsure or unclear

2. **contact** : **contact list** ☐ click to swap word order
   - ○ is a part of
   - ○ is a kind of
   - ○ is equivalent to
   - ○ is unrelated to
   - ○ unsure or unclear

3. **screen content** : **user content** ☐ click to swap word order
   - ○ is a part of
   - ○ is a kind of
   - ○ is equivalent to
   - ○ is unrelated to
   - ○ unsure or unclear

Sample survey results

	P	W	H	O	E	U	X
browser : web browser type	3	11	3	3	9	0	1
contact : contact list	24	2	0	0	4	0	0
screen content : user content	6	2	4	6	4	6	2
mobile device: unique device id	3	19	1	2	2	2	1

P: Part-of
W: Whole-of
H: Hypernym-of (superclass)
O: Hyponym-of (subclass)

E: Equivalent-to
U: Unrelated
X: Unsure or unclear

Lexicon containing 351 unique information type phrases

Sample survey results

	P	W	H	O	E	U	X
browser : web browser type	3	11	3	3	9	0	1
contact : contact list	24	2	0	0	4	0	0
screen content : user content	6	2	4	6	4	6	2
mobile device: unique device id	3	19	1	2	2	2	1

P: Part-of
W: Whole-of
H: Hypernym-of (superclass)
O: Hyponym-of (subclass)

E: Equivalent-to
U: Unrelated
X: Unsure or unclear

Lexicon containing 351 unique information type phrases

Results: Precision=0.964, Recall=0.543
Among 639 false negatives, most require augmented semantics
Reduced paired comparisons by 7,719 or 12% of 62,853

Conclusion and Future Work

- Hearst patterns:
  - Yields general categories (demographic, contact, personal)
  - Difficult to encode, about 17% false positives
- Variant rules:
  - Yield lexical hypernyms and meronyms
  - Linear shallow typing, less than 5% false positives

Future Work:

- Improve search for non-lexical hypernyms and meronyms
- Correlate information type and data purpose (semantic roles)
- Score information types by privacy risk (w/ context)

Questions?

Relevant Publications:

- Jaspreet Bhatia, Morgan C. Evans, Sudarshan Wadkar and Travis D. Breaux, "Automated Extraction of Regulated Information Types using Hyponymy Relations," *IEEE Workshop on Artificial Intelligence and Requirements Engineering,* 2016.
- Mitra Bokaei Hosseini, Sudarshan Wadkar, Travis D. Breaux, Jianwei Niu, "Lexical Similarity of Information Type Hypernyms, Meronyms and Synonyms in Privacy Policies," *AAAI Fall Symposium on Language Technologies and Privacy*, 2016.

This research was funded by NSF Frontier Award #1330596, NSF CAREER Award #1453139 and NSA Award #141333.