More of a discussion on ***web privacy***

Ralf Lämmel

Software Languages Team, University of Koblenz-Landau

Joint work with Ekaterina Pek

Presentation at EMN, 28 Feb 2012

© 2010-2012, Software Languages Team, University of Koblenz-Landau

TOC

- ***What's*** web privacy?
- The language-based approach of ***P3P***
- Web privacy -- ***all fine***?
- Research ***challenges***
- An ***empirical*** study on P3P usage

© 2010-2012, Software Languages Team, University of Koblenz-Landau

What's web privacy?

© 2010-2012, Software Languages Team, University of Koblenz-Landau

Excerpt of a privacy policy

[COMPANY NAME] is the sole owner [*or describe other entities, e.g., parent, affiliates, who have an ownership interest]* of the information collected on [NAME OF SITE]. [COMPANY NAME] collects personally identifiable information from our users at several different points on our Web site.

Orders

If you purchase a product or service from us, we request certain personally identifiable information from you on our order form. You must provide contact information (such as name, email, and shipping address) and financial information (such as credit card number, expiration date).

We use this information for billing purposes and to fill your orders. If we have trouble processing an order, we will use this information to contact you.

...

Source: www.**truste**.org/docs/Model_**Privacy_Policy**_Disclosures.doc

The TRUSTe seal for web privacy

debnroo.com

debnroo
real deals. real fast.

Save 5% Now!
click here to save 5%
on every order

TRUSTe
VERIFIED

Connect with us! Follow us!
Customer Service Returns Center Order Status

Home Garden Pets Kitchen Outdoors

debnroo@gmail.com
970-416-6300

Items: None $0.00 Cart Checkout

Search

Debnroo.com

FAQs About Us Contact Us

Categories

- Garden
  - Environmental Products
  - Garden Decor
  - Garden Tools & Gadgets
  - Growing Medium
  - Indoor Gardening
  - Landscaping Supplies
  - Pest Control
  - Pots & Planters
  - Small Animal Supplies
- Home
  - Apparel/Clothing
  - Appliances
  - Electronics
  - Furniture and Decor
  - Garage
  - Gift Giving Ideas
  - Green and Environmental
  - Health And Beauty
  - Kitchen
  - Outdoors
  - Safety and Security
  - Seasonal
  - Toys
  - Wholesale Lots
- Pets
  - Aquarium
  - Birds
  - Cats
  - Chicken

Best Sellers

Cat Beds for Comfort and Warmth!

Sage

- Heated Beds
- Cup Beds
- Hooded Beds
- Mats
- Perches

Gift Giving Ideas

Small Appliances

Cat Beds

Dog Beds

Debnroo Deals (Click for All)

Why buy from debnroo?

debnroo

Excerpt con'd

Social Security Numbers

We require your social security number to provide [describe the service you provide here.] When you enter your social security number on our registration [order form, application form], we encrypt it using secure socket layer technology (SSL).

TRUSTe (an organization that grants seals for web privacy) recommends that web sites collect a social security number ONLY when it is required for performing a service the site provides.

Source: www.**truste**.org/docs/Model_**Privacy_Policy_**Disclosures.doc

Some “textual” policies

http://www.**facebook**.com/privacy

http://www.**facebook**.com/about/privacy/

http://www.**google**.com/policies/privacy/

http://www.**ibm**.com/privacy/us/en/

http://www.**microsoft**.com/privacy

https://**twitter**.com/privacy

...

URL rewritten to my settings

Effective per 1 March

General info, **not** a policy

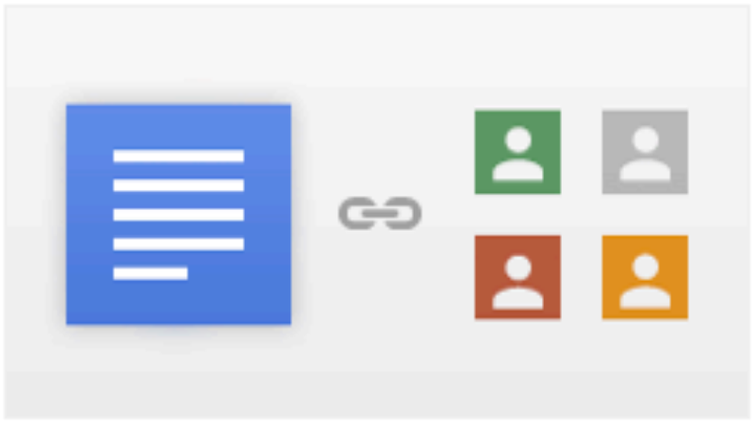
© 2010-2012, Software Languages Team, University of Koblenz-Landau

One policy, one Google experience

We're getting rid of over 60 different privacy policies across Google and replacing them with one that's a lot shorter and easier to read. Our new policy covers multiple products and features, reflecting our desire to create one beautifully simple and intuitive experience across Google.

This stuff matters, so please read our updated Google Privacy Policy and Terms of Service when you have a few minutes. These changes will take effect on March 1, 2012.

OK, got it | Read policies now



Easy to work across Google

Our new policy reflects our desire to create a simple product experi… that does what you need, when yo… Whether you're reading an em… reminds you to schedule a fa… together or finding a favorite … you want to share, we want t… can move across Gmail, Calendar, Search, YouTube, or whatever your life calls for with ease.

Tailored for you

If you're signed into Google, we can do things like suggest search queries — or …

Easy to share and collaborate

When you post or create a document online, you often want others to see and contribute. By remembering the contact … e people you want to … make it easy for you to … oogle product or service … icks and errors.

If you are gmail user, for example, then you remember this Google screen.

Source: http://www.google.com/intl/en/policies/privacy/preview/

Sketch of Google's new privacy policy

- **The Privacy Policy explains:**
  - what information Google collects,
  - why that information is collected,
  - how Google use that information, and
  - what choices Google offers -- including
    - how to access and update information.

© 2010-2012, Software Languages Team, University of Koblenz-Landau

Sketch of Google's new privacy policy

- **Information Google collects:**
  - Information you give Google
  - Information Google gets from your use of Google services
    - Device information
    - Log information
    - Location information
    - Unique application numbers
    - Local storage
    - Cookies and anonymous identifiers

© 2010-2012, Software Languages Team, University of Koblenz-Landau

The language-based approach of P3P

© 2010-2012, Software Languages Team, University of Koblenz-Landau

Platform for Privacy Preferences

Initiative

P3P 1.0: A New Standard in Online Privacy

"The World Wide Web Consortium, the group that designs standards for the Web, is creating a new way [P3P] for Web sites to transmit the site's privacy policy automatically, and allow users to signal only the information they are willing to share."
-- The New York Times 2/22/2000

http://www.w3.org/P3P/brochure.html

last revised $Date: 2006/05/12 15:20:18 $ by $Author: rigo $

Gmail - Inbox (7,029) - rlae | FAQs on privacy practices, s | sample truste privacy statem | dict.leo.org - Results for 've | P3P - Google Search

www.google.de/search?ix=seb&q=P3P&um=1&ie=UTF-8&hl=en&tbm=isch&source=og&sa...LT5OOB4TDswaz7OGFDw&biw=724&...

+Ralf Search Images Maps YouTube Gmail Documents Calendar Translate M

"BTW"

Google P3P

rlaemm... 0 + Share

We're changing our privacy policy and terms. This stuff matters.

Learn more | Dismiss

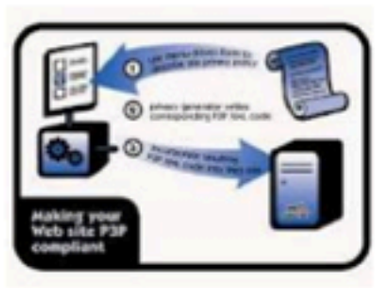
Search About 534,000 results (0.37 seconds) SafeS

Everything
Images
Maps
Videos
News
Shopping
Blogs
More

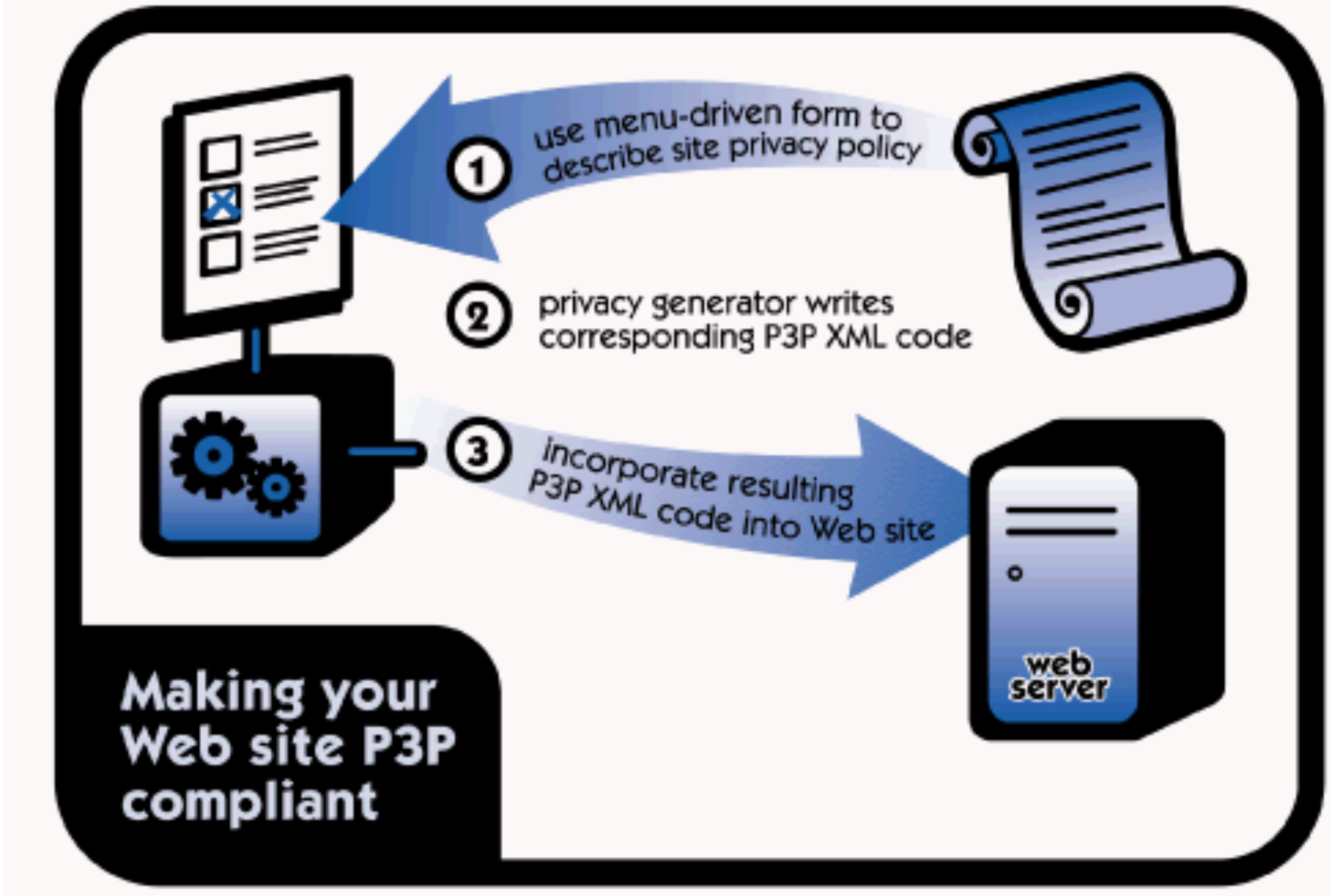
All results
By subject

Any size
Large
Medium
Icon
Larger than...
Exactly...

Related searches: p3p persona persona 3 portable persona 3 psp

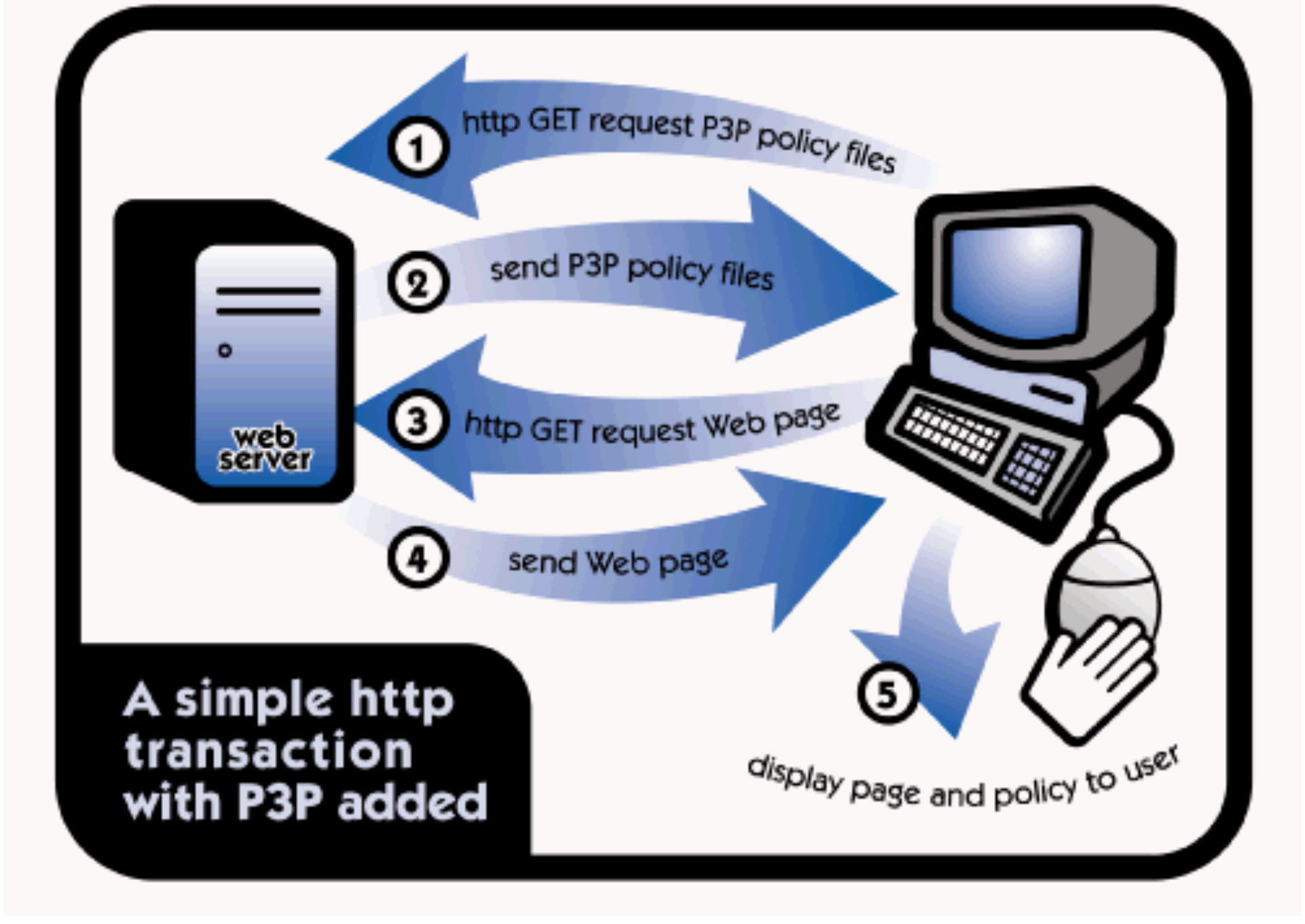


Model_Privacy_Policy_....doc Model_Privacy_Policy_....doc Show All



http://www.w3.org/P3P/brochure.html

last revised $Date: 2006/05/12 15:20:18 $ by $Author: rigo $



http://www.w3.org/P3P/brochure.html

last revised $Date: 2006/05/12 15:20:18 $ by $Author: rigo $

A quick intro to P3P

© 2010-2012, Software Languages Team, University of Koblenz-Landau

The “full privacy” policy

```
<POLICY>
  <STATEMENT>
    <NON-IDENTIFIABLE/>
  </STATEMENT>
</POLICY>
```

© 2010-2012, Software Languages Team, University of Koblenz-Landau

Full privacy with Tor?

What is Tor?

Tor is free software and an open network that helps you defend against a form of network surveillance that threatens personal freedom and privacy, confidential business activities and relationships, and state security known as traffic analysis

Learn more about Tor »

Why Anonymity Matters

Tor protects you by bouncing your communications around a distributed network of relays run by volunteers all around the world: it prevents somebody watching your Internet connection from learning what sites you visit, and it prevents the sites you visit from learning your physical location. Tor works with many of your existing applications, including web browsers, instant messaging clients, remote login, and other applications based on the TCP protocol.

The “logging only” policy

```
<POLICY>
  <STATEMENT>
    <PURPOSE><admin/><current/><develop/></PURPOSE>
    <RECIPIENT><ours/></RECIPIENT>
    <RETENTION><indefinitely/></RETENTION>
    <DATA-GROUP>
      <DATA ref="#dynamic.clickstream"/>
      <DATA ref="#dynamic.http"/>
    </DATA-GROUP>
  </STATEMENT>
</POLICY>
```

© 2010-2012, Software Languages Team, University of Koblenz-Landau

A variation on “logging only” with less exposure

```
<POLICY>
  <STATEMENT>
    <PURPOSE><admin/><current/><develop/></PURPOSE>
    <RECIPIENT><ours/></RECIPIENT>
    <RETENTION><legal-requirement/></RETENTION>
    <DATA-GROUP>
      <DATA ref="#dynamic.clickstream"/>
    </DATA-GROUP>
  </STATEMENT>
</POLICY>
```

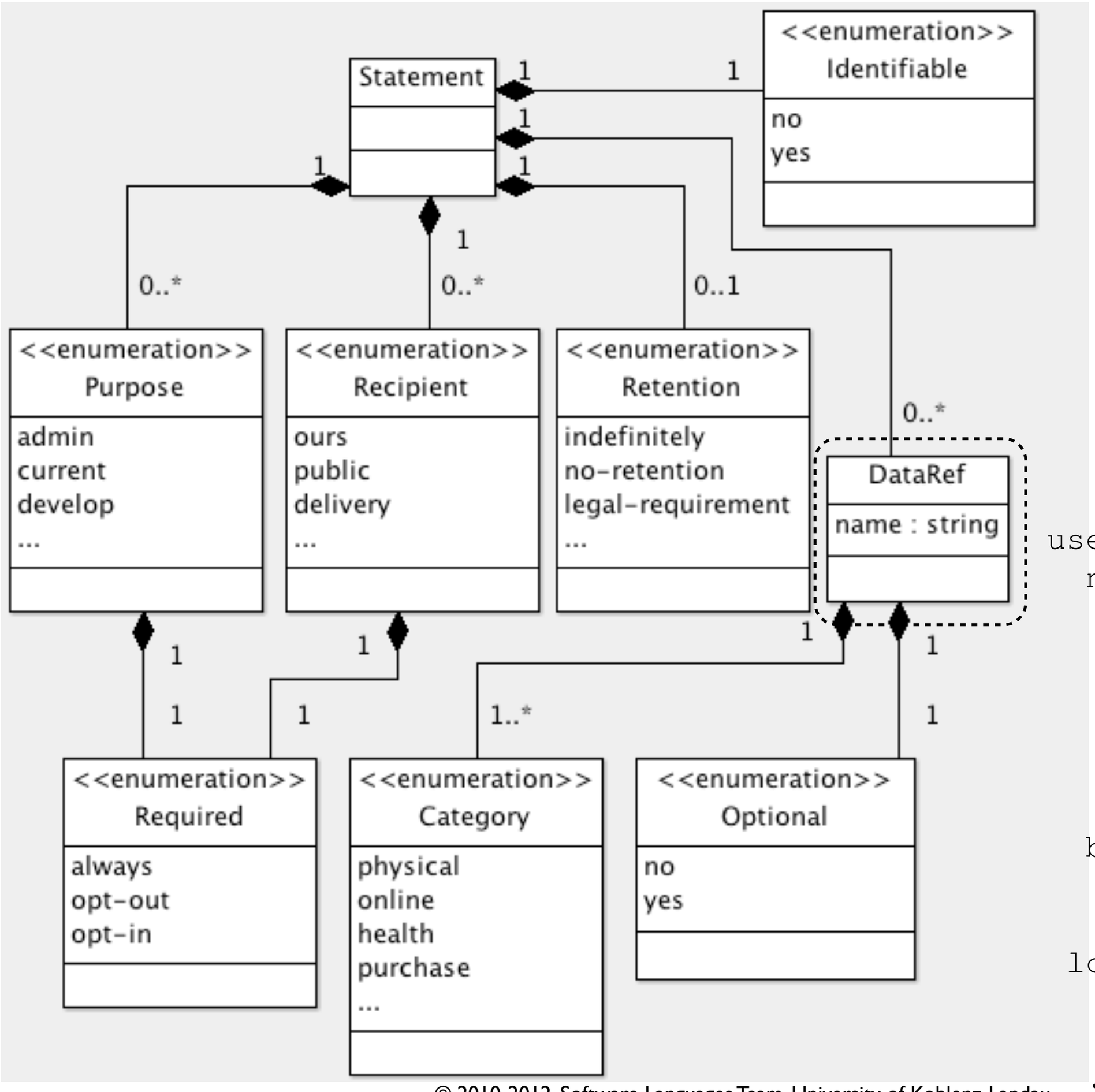
© 2010-2012, Software Languages Team, University of Koblenz-Landau

Use of “wildcards” and categories

```
<DATA ref="#dynamic.miscdata">
  <CATEGORIES>
    <political/>
  </CATEGORIES>
</DATA>
```

© 2010-2012, Software Languages Team, University of Koblenz-Landau

Abstract syntax




```
user
   name
     prefix
     given
     middle
     family
     suffix
     nickname
  bdate
     ymd.year
     ...
login
     id
     password
...
```

© 2010-2012, Software Languages Team, University of Koblenz-Landau

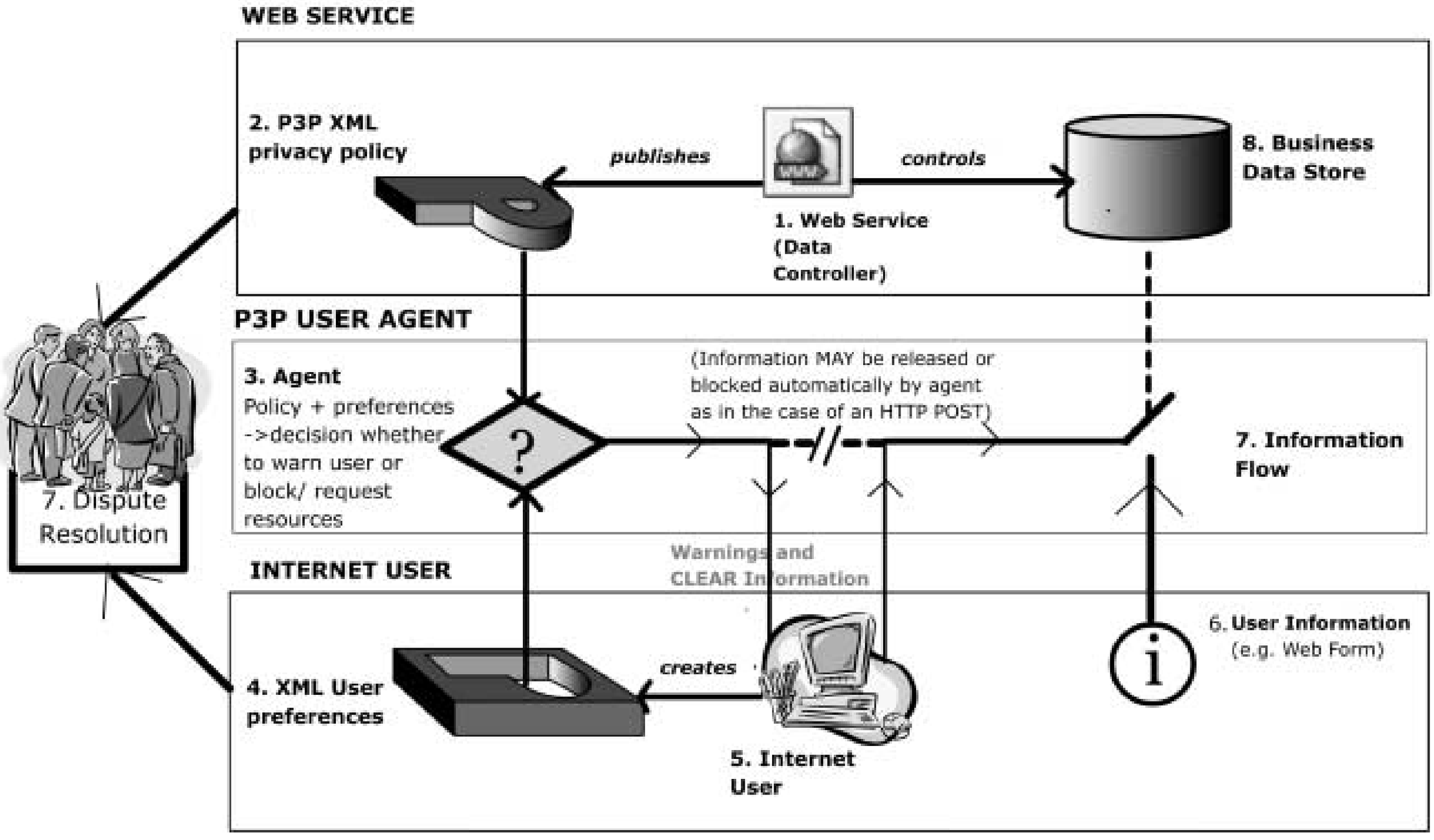
In need of a normal form

- Styles:
  - Use ***one*** statement with ***multiple*** data refs
  - Use ***multiple*** statements with ***one*** data ref
  - ...
- Normal form

r-purpose(data,purpose,required)
r-recipient(data,recipient,required)
r-retention(data,retention)
r-data(data,optional,identifiable)
r-category(data,category)

© 2010-2012, Software Languages Team, University of Koblenz-Landau

The P3P process



Giles Hogben, Tom Jackson, and Marc Wilikens: *A Fully Compliant Research Implementation of the P3P Standard for Privacy Protection: Experiences and Recommendations*, Proceedings of ESORICS 2002

Participants, Supporters, Developers

The following companies and organizations have been active participants in developing P3P.

- America Online
- AT&T
- Center for Democracy & Technology
- Citigroup
- Crystaliz
- Direct Marketing Association
- Electronic Network Consortium
- Geotrust
- Gesellschaft für Mathematik und Datenverarbeitung (GMD)
- Hewlett Packard
- IBM
- IDcide
- International Security, Trust, and Privacy Alliance
- Internet Alliance
- Jotter Technologies Inc.
- Microsoft
- NCR
- NEC
- Netscape
- Nokia
- Ontario Office of the Information and Privacy Commissioner
- Phone.com, Inc.
- Privacy Commission of Schleswig-Holstein, Germany
- TRUSTe

Web privacy -- all fine?

© 2010-2012, Software Languages Team, University of Koblenz-Landau

Google Bypassing User Privacy Settings

Published Monday, February 20, 2012 7:31 PM

152 comments

When the IE team heard that Google had bypassed user privacy settings on Safari, we asked ourselves a simple question: is Google circumventing the privacy preferences of Internet Explorer users too? We've discovered the answer is yes: Google is employing similar methods to get around the default privacy protections in IE and track IE users with cookies. Below we spell out in more detail what we've discovered, as well as recommendations to IE users on how to protect their privacy from Google with the use of IE9's Tracking Protection feature. We've also contacted Google and asked them to commit to honoring P3P privacy settings for users of all browsers.

We've found that Google bypasses the P3P Privacy Protection feature in IE. The result is similar to the recent reports of Google's circumvention of privacy protections in Apple's Safari Web browser, even though the actual bypass mechanism Google uses is different.

Internet Explorer 9 has an additional privacy feature called Tracking Protection which is not susceptible to this type of bypass. Microsoft recommends that customers who want to protect themselves from Google's bypass of P3P Privacy Protection use Internet Explorer 9 and click here to add a Tracking Protection List. Customers can find additional lists and information on this page.

http://blogs.msdn.com/b/ie/archive/2012/02/20/google-bypassing-user-privacy-settings.aspx

Attorneys general have 'strong concerns' about Google privacy rules

By **Doug Gross**, CNN
February 22, 2012 -- Updated 2318 GMT (0718 HKT) | Filed under: **Web**

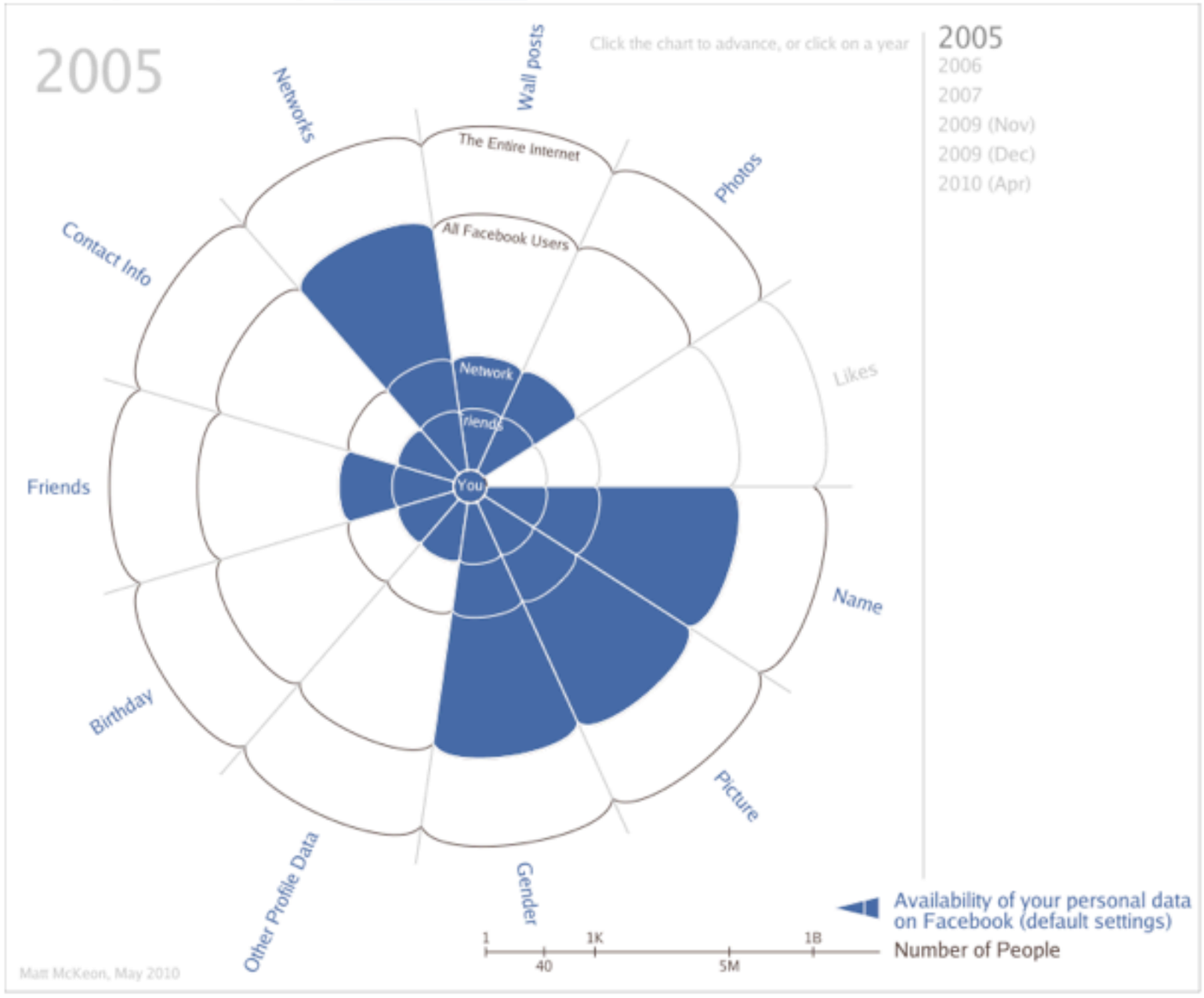
Google says its new privacy policy is meant to be clearer to users and doesn't change privacy controls.

http://edition.cnn.com/2012/02/22/tech/web/google-privacy-attorneys-general/index.html

The Evolution of Privacy on Facebook

Changes in default profile settings over time

Trouble seeing the vis? Try switching to an image-based version.



http://mattmckeon.com/facebook-privacy/

Sony Hacked Again; 25 Million Entertainment Users' Info at Risk

By Jason Schreier May 2, 2011 | 7:11 pm | Categories: Online Gaming

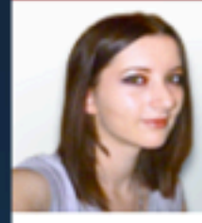
Follow @jasonschreier

658 Tweet 2 +1 247 Share

Like Send 2,589 people like this. Be the first of your friends.

Sony Online Entertainment games like *DC Universe Online* were taken down Monday, after a security breach that exposed 24.6 million Sony users' accounts and some credit card numbers.
Image courtesy Sony Online Entertainment

http://www.wired.com/gamelife/2011/05/sony-online-entertainment-hack/



Facebook does not erase user-deleted content

By Zack Whittaker | April 28, 2010, 7:50am PDT

Summary: *When users upload content, pictures and video to Facebook, it is not erased from their servers even when the user 'deletes' it. Video and experiment.*

128 Comments
51
Tweet
642
Like
7
0
+1
more +

Nearly a year on since Cambridge University researchers discovered that Facebook, along with other major social networks, doesn't erase server-side copies of your uploaded data, the world's biggest social network is still guilty of such a sin.

Since then, the site has nearly doubled in size. It's now has the population of the third biggest country in the world, with tens of thousands of servers holding your data, which as soon as it is uploaded, belongs to them to do as and what they wish with it.

Four US Senators are jumping on the privacy bandwagon concerned over users' privacy. As Sam Diaz points out, profile data by third-party developers (ie. anybody who can write an "application" for the site:

> "...used to be limited to 24 hours but that was recently lifted by Facebook. At the f8 conference last week, the executives said this was more of a technicality and suggested that it wasn't as big of a deal as it might sound."

MOBILE WORLD CONGRESS 2012
business MOBILE.FR
Du 27 février au 1er mars
part à Barcelone
Suivre l'évènement
business MOBILE.FR

The best of ZDNet, delivered

ZDNet Newsletters
Get the best of ZDNet delivered straight to your inbox
Your Email
ZDNet's White Paper Membership Newsletter: Stay current with site news and updates from White Papers
ZDNet's Must-Read News Alerts: Breaking IT news as it happens
Subscribe Today
Be a fan on Facebook
Follow us on Twitter

Status: P3P Work suspended

After a successful Last Call, the P3P Working Group decided to publish the P3P 1.1 Specification as a Working Group Note to give P3P 1.1 a provisionally final state.

13 November 2006

The P3P Specification Working Group took this step as there was insufficient support from current Browser implementers for the implementation of P3P 1.1. The P3P 1.1 Working Group Notecontains all changes from the P3P 1.1 Last Call. The Group thinks that P3P 1.1 is now ready for implementation. It is not excluded that W3C will push P3P 1.1 until Recommendation if there is sufficient support for implementation.

On the other hand, P3P keeps being the basis of a number of research directions in the area of privacy world wide. One might cite the PRIME Project as well as the Policy aware Web. Many other approaches also follow the descriptive metadata approach started by P3P. Such projects are invited to send email to <rigo@w3.org> to be listed here.

http://www.w3.org/P3P/

Lorrie Faith Cranor (CMU, Developer P3P) on the reasons for lack of P3P adoption

Arguably, the largest barrier to P3P adoption has not been problems with the P3P vocabulary or difficulties with the technical mechanisms, but rather lack of incentives to adopt. As Dyson observed in 1997, "Industry disclosure schemes often founder without strong government/public pressure. Otherwise, companies are simply too busy to adopt them, and customers don't factor the information disclosed into their buying habits."[10] By the time the P3P specification was released in 2002, government pressure had subsided and industry had largely lost interest in P3P.

http://www.ftc.gov/os/comments/privacyreportframework/00453-58003.pdf

[Emphasis added by this speaker]

CENTER FOR DEMOCRACY & TECHNOLOGY

Looking Back at P3P: Lessons for the Future

by Ari Schwartz

November 11, 2009

- Building and Over-Building
- Caught Up in the Politics of Privacy: Criticized by some ...
  - ... privacy advocates as an industry subterfuge
  - ... industry advocates as a substitute for legislation
  - ... in industry for providing consumers too much transparency
- Web sites build to the implementation, not the specification.

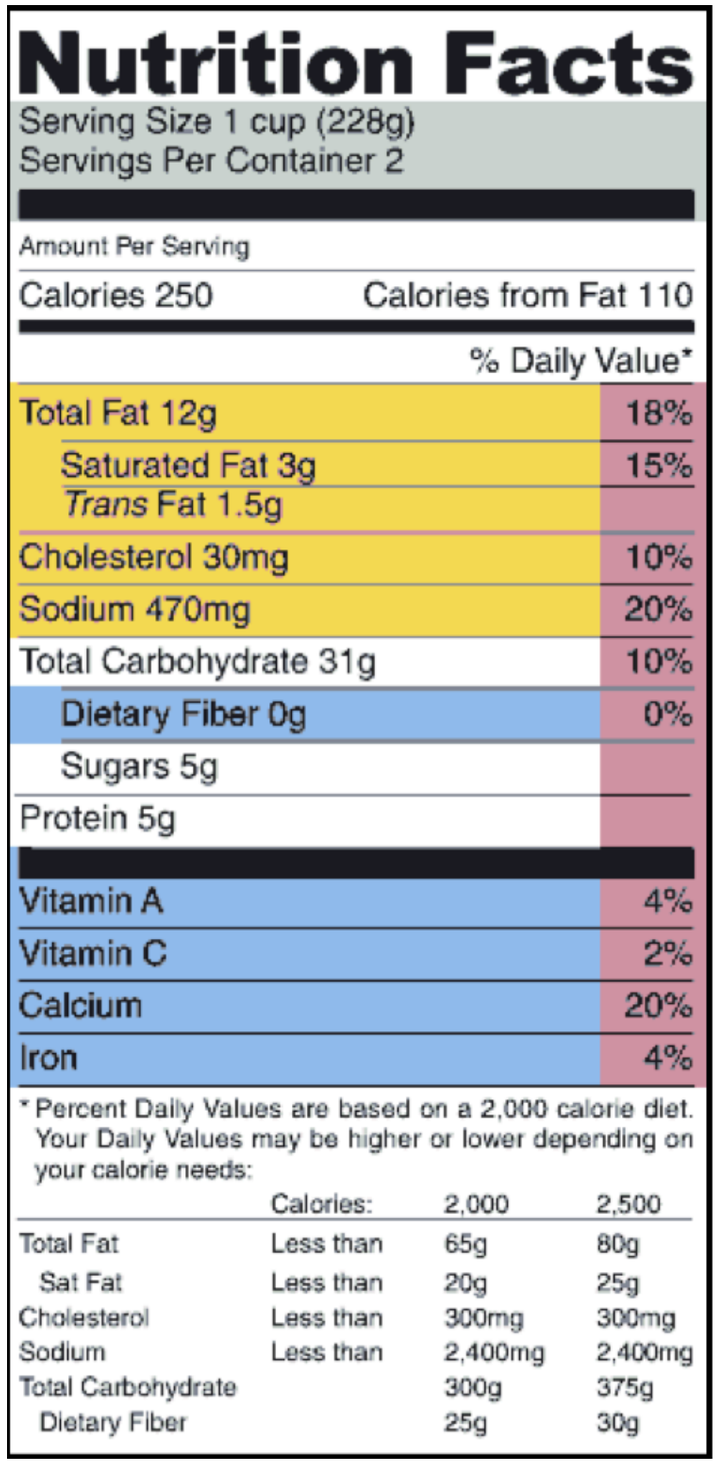
https://www.cdt.org/paper/looking-back-p3p-lessons-future

Research challenges

© 2010-2012, Software Languages Team, University of Koblenz-Landau

Research challenges

- Integration of privacy and access control
- Understandability of privacy policies
- Effective control of privacy
- Privacy for users or for legislators
- Privacy as specified or as implemented
- Learn from the past -- scientifically



Nutrition Facts

Serving Size 1 cup (228g)
Servings Per Container 2

Amount Per Serving	
Calories 250	Calories from Fat 110
	% Daily Value*
Total Fat 12g	18%
Saturated Fat 3g	15%
Trans Fat 1.5g	
Cholesterol 30mg	10%
Sodium 470mg	20%
Total Carbohydrate 31g	10%
Dietary Fiber 0g	0%
Sugars 5g	
Protein 5g	
Vitamin A	4%
Vitamin C	2%
Calcium	20%
Iron	4%

* Percent Daily Values are based on a 2,000 calorie diet. Your Daily Values may be higher or lower depending on your calorie needs:

	Calories:	2,000	2,500
Total Fat	Less than	65g	80g
Sat Fat	Less than	20g	25g
Cholesterol	Less than	300mg	300mg
Sodium	Less than	2,400mg	2,400mg
Total Carbohydrate		300g	375g
Dietary Fiber		25g	30g

A "Nutrition Label" for Privacy

Patrick Gage Kelley,[*] Joanna Bresee,[*] Lorrie Faith Cranor,[*] Robert W. Reeder[**]

[*] Carnegie Mellon University [**] Microsoft

Symposium On Usable Privacy and Security (SOUPS) 2009, July 15-17, 2009, Mountain View, CA, USA.

A "Nutrition Label" for Privacy

Patrick Gage Kelley,* Joanna Bresee,* Lorrie Faith Cranor,* Robert W. Reeder**
*Carnegie Mellon University
**Microsoft

Symposium On Usable Privacy and Security (SOUPS) 2009, July 15-17, 2009, Mountain View, CA, USA.

The Acme Policy

types of information	how we use your information: provide service & maintain site	research & development	marketing	telemarketing	profiling	who we share your information with: other companies	public forums
contact information	!	!	OUT	OUT		IN	
cookies	!	!	OUT	OUT		IN	
demographic information							
financial information							
health information							
preferences	!	!	OUT	OUT		IN	!
purchasing information	!	!	OUT	OUT		IN	
social security number & govt ID	!						
your activity on this site	!	!	OUT	OUT		IN	!
your location							

understanding this privacy policy

- ! — we **will** use your information in this way
- (blank) — we **will not** collect or we **will not** use your information in this way
- OUT — we **will** use your information in this way unless you opt-out
- IN — we **will not** use your information in this way unless you opt-in

Some related work on "enforcement"

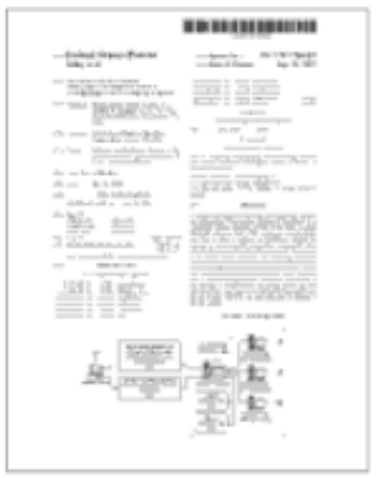
- P. Ashley: *Enforcement of a P3P Privacy Policy*, in: Proceedings of AISM 2004.
- Katia Hayati and Martin Abadi: *Language-Based Enforcement of Privacy Policies*, in: Proceedings of PET 2004.
- Yuh-Jong Hu, Hong-Yi Guo, and Guang-De Lin: Semantic Enforcement of Privacy Protection Policies via the Combination of Ontologies and Rules, in: Proceeedings of SUTC 2008
- Daniel J. Weitzner, Jim Hendler, Tim Berners-Lee, and Dan Connolly: *Creating a Policy-Aware Web: Discretionary, Rule-based Access for the World Wide Web*, in: Web and Information Security, 2006

United States Patent
Ashley et al.

(10) **Patent No.: US 7,797,726 B2**
(45) **Date of Patent: Sep. 14, 2010**

Method and system for implementing privacy policy enforcement with a privacy ...

Paul Anthony Ashley et al



› **Overview**
Abstract
Drawings
Description
Claims

+1 1

Go

"+1" by this speaker

A method is presented for enforcing a privacy policy concerning management of personally identifiable information in a centralized manner through a privacy proxy agent. A proxy intercepts a message from a first system to a second system, e.g., from a server to a client, and determines whether the message is associated with an operation on personally identifiable information; if not, then the proxy sends the message to the second system, but if so, then the proxy determines whether the operation on the personally identifiable information is compliant with a privacy policy and with user preference information with respect to the privacy policy for a user who is associated the personally identifiable information. If the message is compliant with the privacy policy and user preference data, then the proxy sends the first message to the second system; otherwise, an error indication is returned to the first system.

Inventors: Paul Anthony Ashley, Sridhar R. Muppidi, Mark Vandenwauver
Original Assignee: International Business Machines Corporation
Primary Examiner: Gilberto Barron, Jr.
Secondary Examiner: Venkat Perungavoor
Attorneys: Jeffrey S. LaBaw, David H. Judson
Current U.S. Classification: 726/1; 726/30

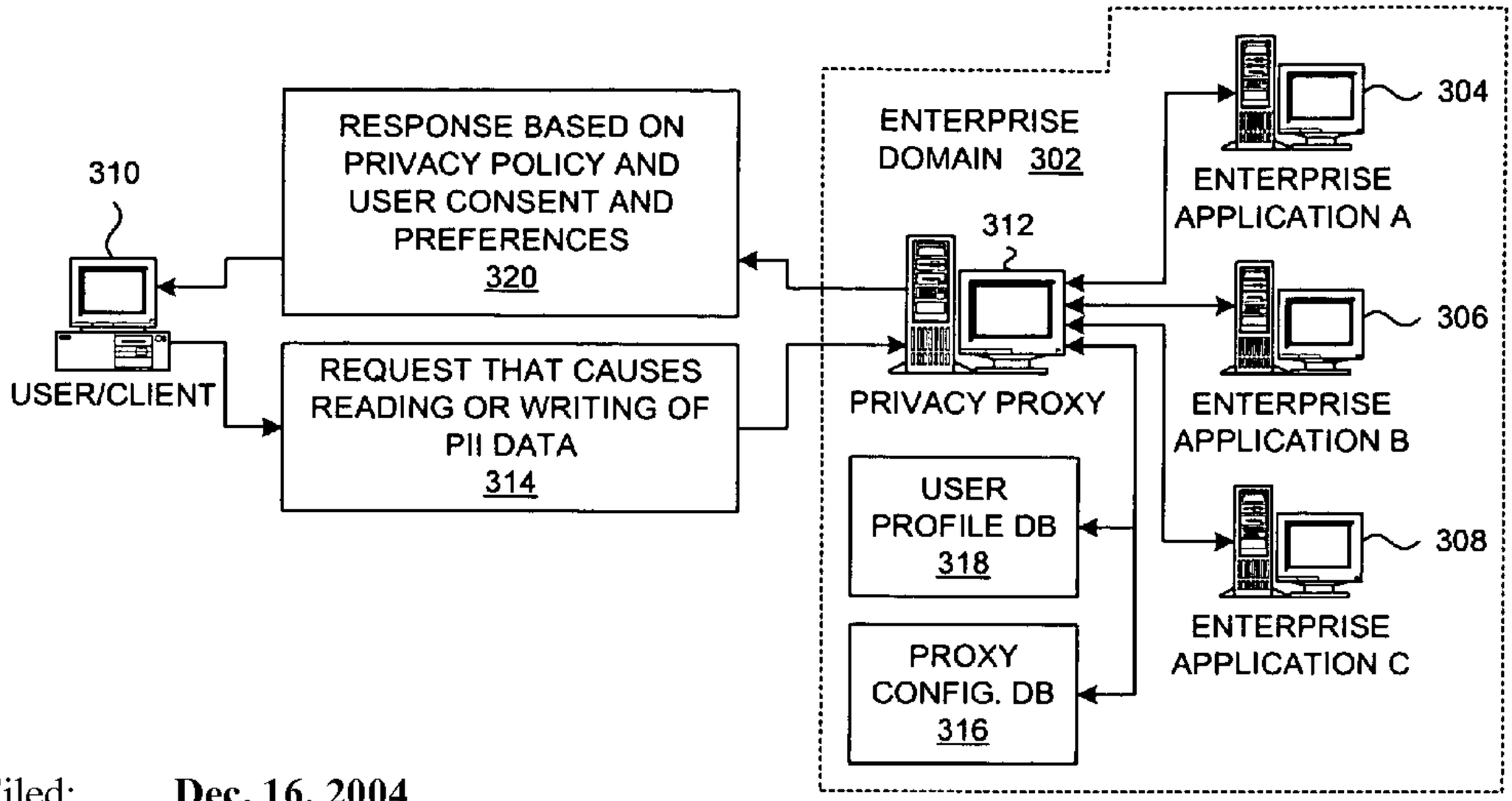
Source: http://www.google.com/patents/US7797726

United States Patent

Ashley et al.

(10) **Patent No.: US 7,797,726 B2**

(45) **Date of Patent: Sep. 14, 2010**



(22) Filed: **Dec. 16, 2004**

(65) **Prior Publication Data**

US 2006/0136985 A1 Jun. 22, 2006

An empirical study on P3P usage

© 2010-2012, Software Languages Team, University of Koblenz-Landau

Research questions

- What part of the vocabulary is used?
- Is the language correctly used?
- What is a significant policy?
- What are common policies?
- What language extensions circulate?

© 2010-2012, Software Languages Team, University of Koblenz-Landau

Our empirical work in comparison

http://softlang.uni-koblenz.de/p3p/

Property	**[18]**	**[19]**	**[20]**	**[21]**
Time stamp	12/2006?	11/2005	12/2006	11/2006
Size of corpus	3846?	1482?	3846?	3282
Number of sources	11	6	5	1
Syntax error rates	○	○	●	○
Boundary breakdown	–	●	–	●
Website evolution	–	●	●	–
P3P language coverage	–	–	○	–
Privacy classification	●	–	○	–
Legislation	–	–	–	●
HRPP interpretation	–	–	●	–

© 2010-2012, Software Languages Team, University of Koblenz-Landau

A P3P corpus

© 2010-2012, Software Languages Team, University of Koblenz-Landau

http://www.dmoz.org/

dmoz open directory project

In partnership with Aol Search.

about d | link | editor login

Web sites in scope

Arts
Movies, Television, Music...

Business
Jobs, Real Estate, Investing...

Computers
Internet, Software, Hardware...

Games
Video Games, RPGs, Gambling...

Health
Fitness, Medicine, Alternative...

Home
Family, Consumers, Cooking...

Kids and Teens
Arts, School Time, Teen Life...

News
Media, Newspapers, Weather...

Recreation
Travel, Food, Outdoors, Humor...

Reference
Maps, Education, Libraries...

Regional
US, Canada, UK, Europe...

Science
Biology, Psychology, Physics...

Shopping
Clothing, Food, Gifts...

Society
People, Religion, Issues...

Sports
Baseball, Soccer, Basketball...

World
Català, Dansk, Deutsch, Español, Français, Italiano, 日本語, Nederlands, Polski, Русский, Svenska...

Become an Editor Help build the largest human-edited directory of the web

Copyright © 2012 Netscape

Results of URL scraping

# Website URLs	4,009,337
# Distinct website URLs	3,860,709
# Distinct domains	2,957,657

© 2010-2012, Software Languages Team, University of Koblenz-Landau

Results of policy location and downloading

	# Files	# Distinct domains
Located policy reference files	50,776	43,195
Referenced policy files	13,536	8,736
Downloaded policies	8,768	7,746

© 2010-2012, Software Languages Team, University of Koblenz-Landau

Schema-based validation

Criterion	# Files	# Policies	# Distinct domains
All downloads	8,768	—	7,746
XML	7,899	—	7,554
Well-formed XML	7,675	—	7,371
Schema-validated XML	**5,905**	**6,182**	5,673

© 2010-2012, Software Languages Team, University of Koblenz-Landau

Leveraged analyses

- Analysis of vocabulary
- Analysis of constraints
- Analysis of metrics
- Analysis of cloning
- Analysis of extensions

© 2010-2012, Software Languages Team, University of Koblenz-Landau

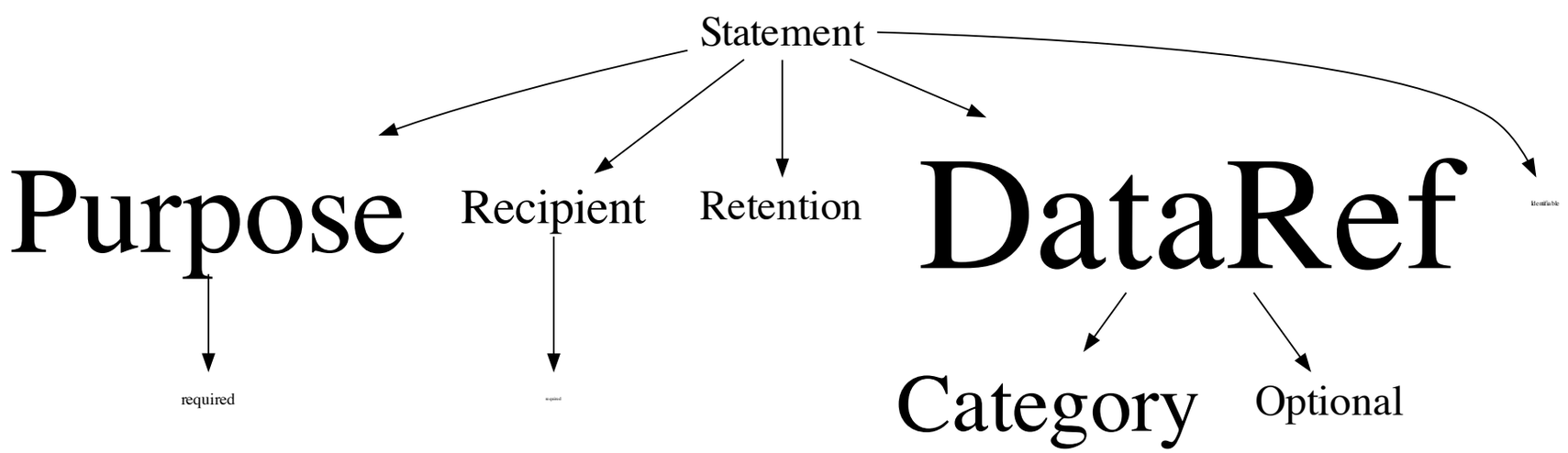
Leveraged analyses

- **Analysis of vocabulary**
- Analysis of constraints
- Analysis of metrics
- Analysis of cloning
- Analysis of extensions

© 2010-2012, Software Languages Team, University of Koblenz-Landau

Frequency of use for P3P's syntactical domains

DataRef	40979
Purpose	31254
Category	20896
Recipient	12496
Optional	10823
Retention	10623
Statement	10623
Required (purpose)	4171
Identifiable	1718
Required (recipient)	1155



© 2010-2012, Software Languages Team, University of Koblenz-Landau

Recipients

© 2010-2012, Software Languages Team, University of Koblenz-Landau

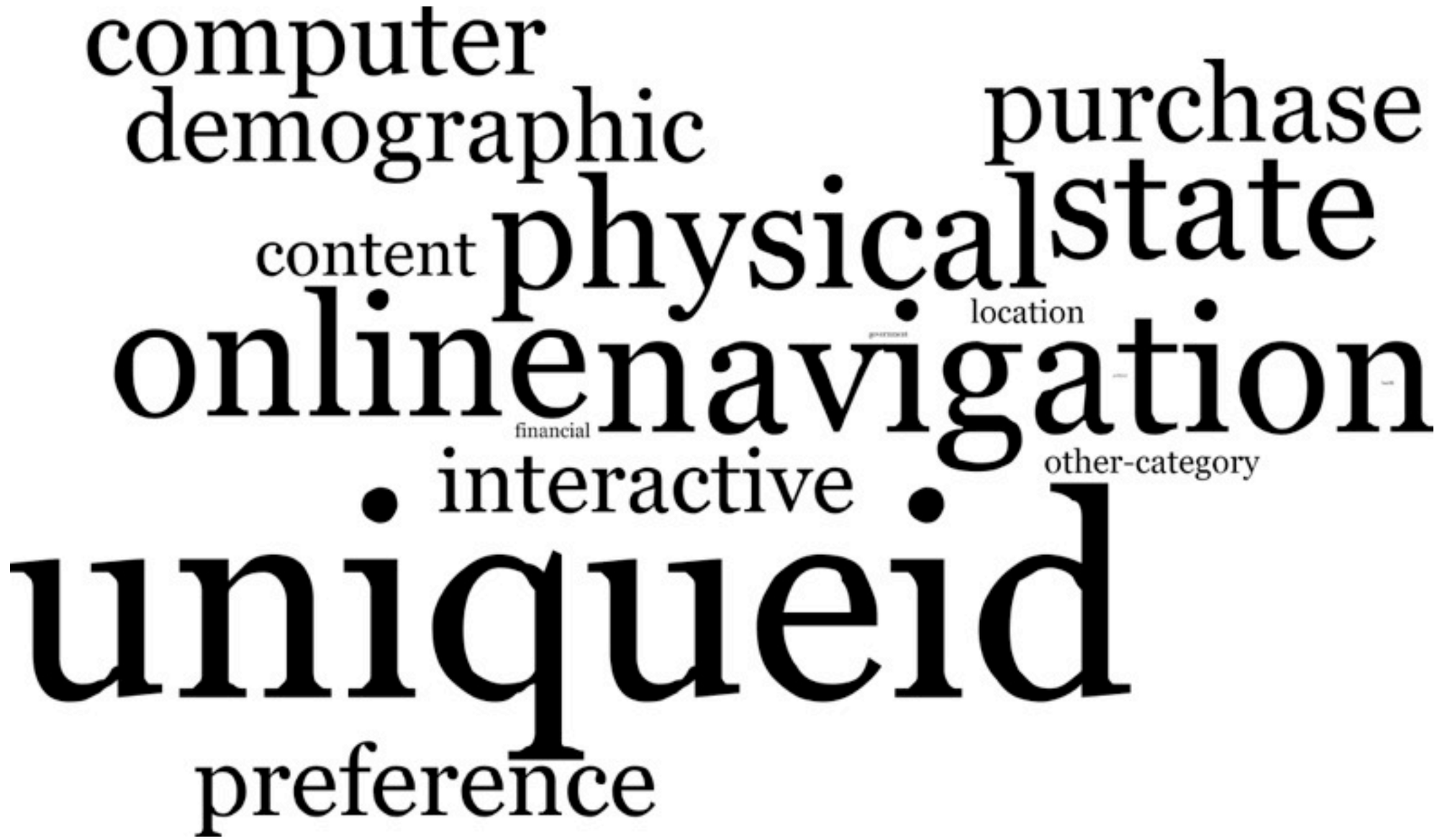
Purposes

© 2010-2012, Software Languages Team, University of Koblenz-Landau

Retention

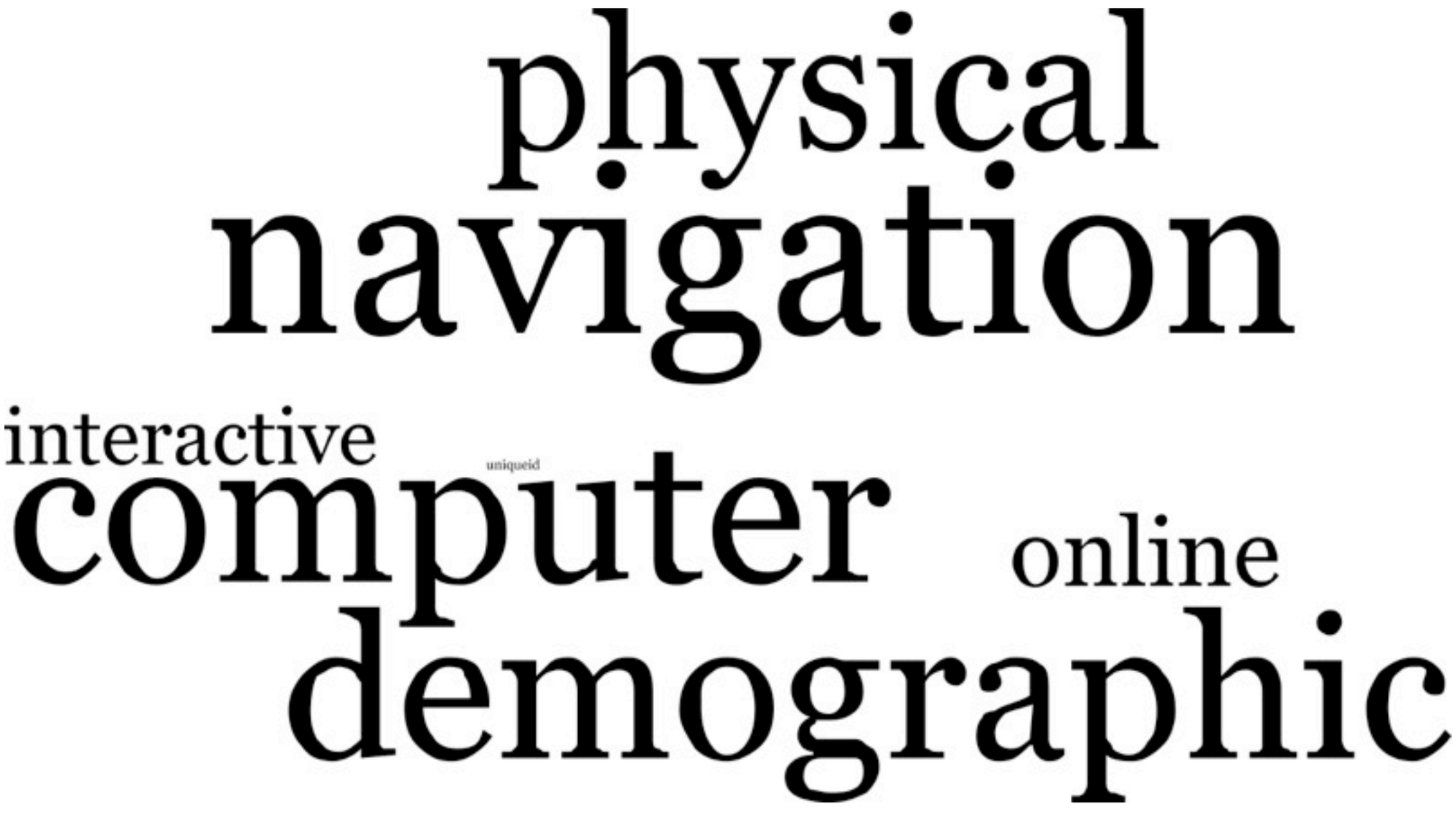
© 2010-2012, Software Languages Team, University of Koblenz-Landau

Explicit categories



© 2010-2012, Software Languages Team, University of Koblenz-Landau

Implicit categories



© 2010-2012, Software Languages Team, University of Koblenz-Landau

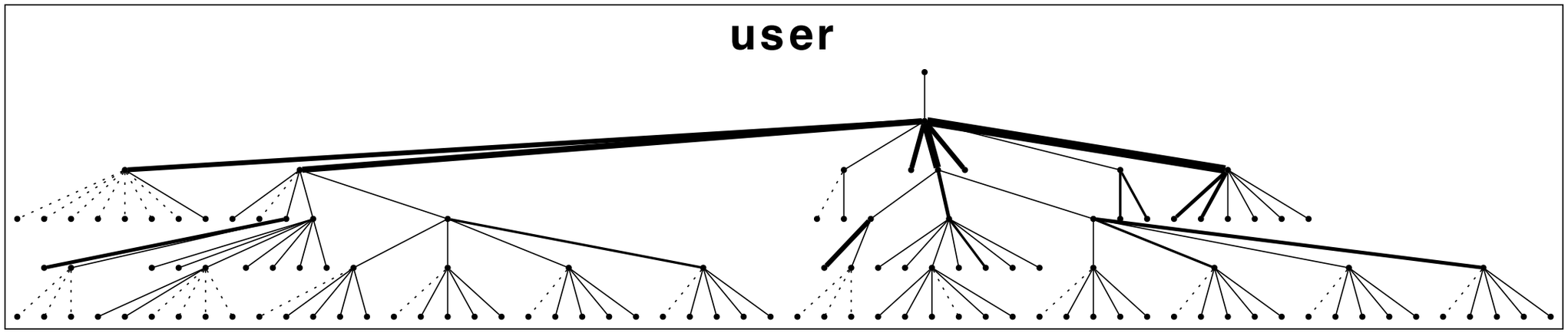
Data refs

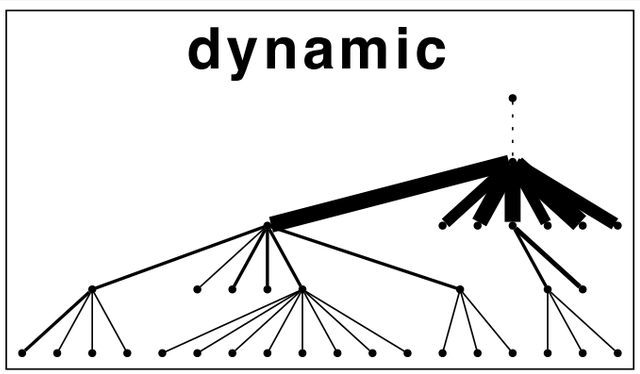
dynamic.cookies	3351
dynamic.clickstream	3204
dynamic.http	3058
dynamic.miscdata	2753
dynamic.searchtext	1816
user.name	995
dynamic.clientevents	726
dynamic.interactionrecord	579
user.home-info	510
user.home-info.online.email	476
user.name.given	418
user.name.family	418
user.business-info.online.email	387
user.business-info	364
thirdparty.name	361

© 2010-2012, Software Languages Team, University of Koblenz-Landau

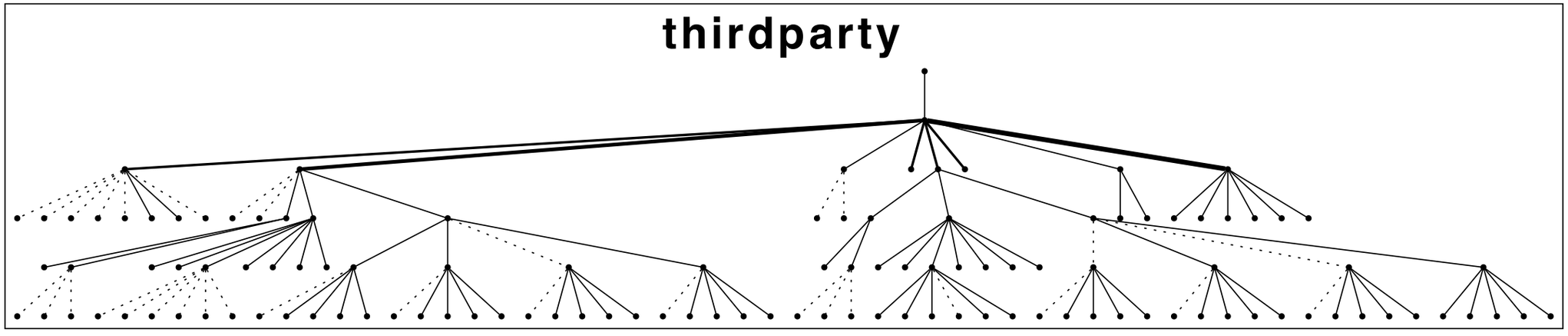
Data refs

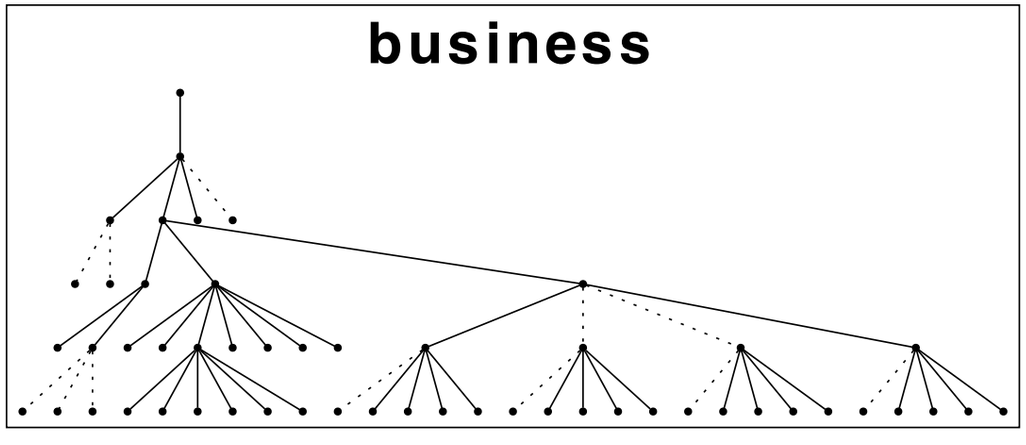
© 2010-2012, Software Languages Team, University of Koblenz-Landau





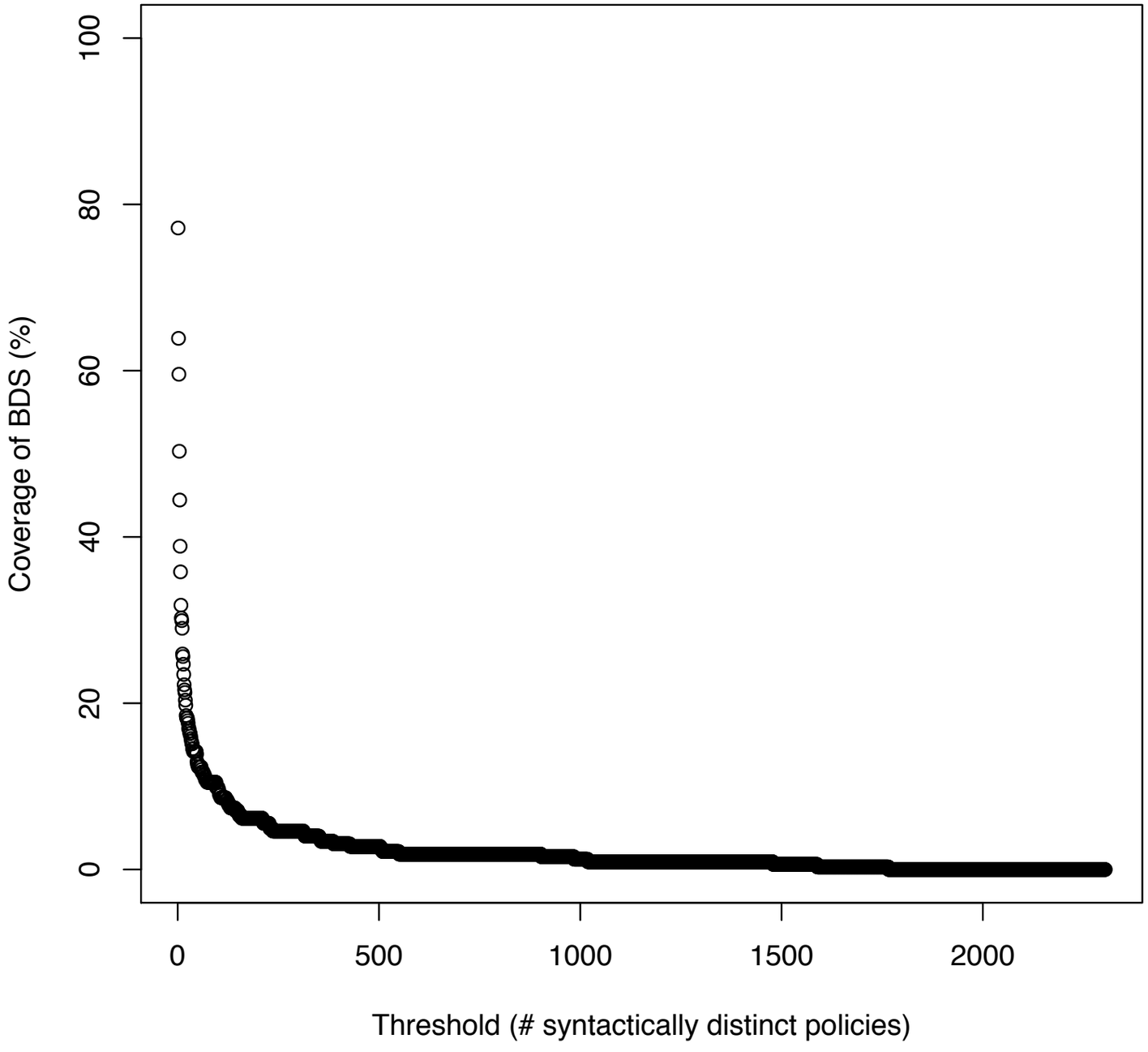
Data refs





© 2010-2012, Software Languages Team, University of Koblenz-Landau

Coverage of BDS with a threshold on the number of syntactically distinct policies



Threshold	BDS coverage
1	77.16 %
2	63.88 %
3	59.56 %
4	50.30 %
5	44.44 %
6	38.88 %
7	35.80 %
8	31.79 %
9	30.24 %
10	29.93 %

© 2010-2012, Software Languages Team, University of Koblenz-Landau

Leveraged analyses

- Analysis of vocabulary
- **Analysis of constraints**
- Analysis of metrics
- Analysis of cloning
- Analysis of extensions

© 2010-2012, Software Languages Team, University of Koblenz-Landau

Key-constraint violations

Relation	# Policies	# Syn. distinct policies
r-data	448	248
r-retention	139	104
r-purpose	50	33
r-recipient	7	7
Any of them	**544** (of 6,182)	**312** (of 2,304)

r-purpose(data,purpose,required)
r-recipient(data,recipient,required)
r-retention(data,retention)
r-data(data,optional,identifiable)
r-category(data,category)

© 2010-2012, Software Languages Team, University of Koblenz-Landau

Data-schema constraint violations

Issue	# Policies	# Syn. distinct policies
BDS legacy	580	358
Unresolvable data reference	115	85
Missing CDS	11	7
CDS legacy	203	157
Any of them	**848** (of 6,182)	**579** (of 2,304)

© 2010-2012, Software Languages Team, University of Koblenz-Landau

Coherence constraints

For example:

'public recipient ⇒ indefinite retention'

Source	Constraint	# Policies	# Syn. distinct policies
[30]	'ours ...'	77	45
[30]	'public ...'	40	24
[30]	'historical ...'	6	6
[1]	'contact'	71	47
[1]	'telemarketing'	353	107
[1]	'individual-analysis'	451	94
[1]	'individual-decision'	432	82
	Any of them	**898** (of 6,182)	**264** (of 2,304)

© 2010-2012, Software Languages Team, University of Koblenz-Landau

Summary of violations

Constraint category	**# Policies**	**# Syn. distinct policies**
Key	544	312
Hierarchy	17	15
Data-schema	848	579
Coherence	898	264
Any of them	**2,193** (of 6,182)	**1,083** (of 2,304)

Important disclaimer: some details such as the choice of key constraints are subjective.

© 2010-2012, Software Languages Team, University of Koblenz-Landau

Leveraged analyses

- Analysis of vocabulary
- Analysis of constraints
- **Analysis of metrics**
- Analysis of cloning
- Analysis of extensions

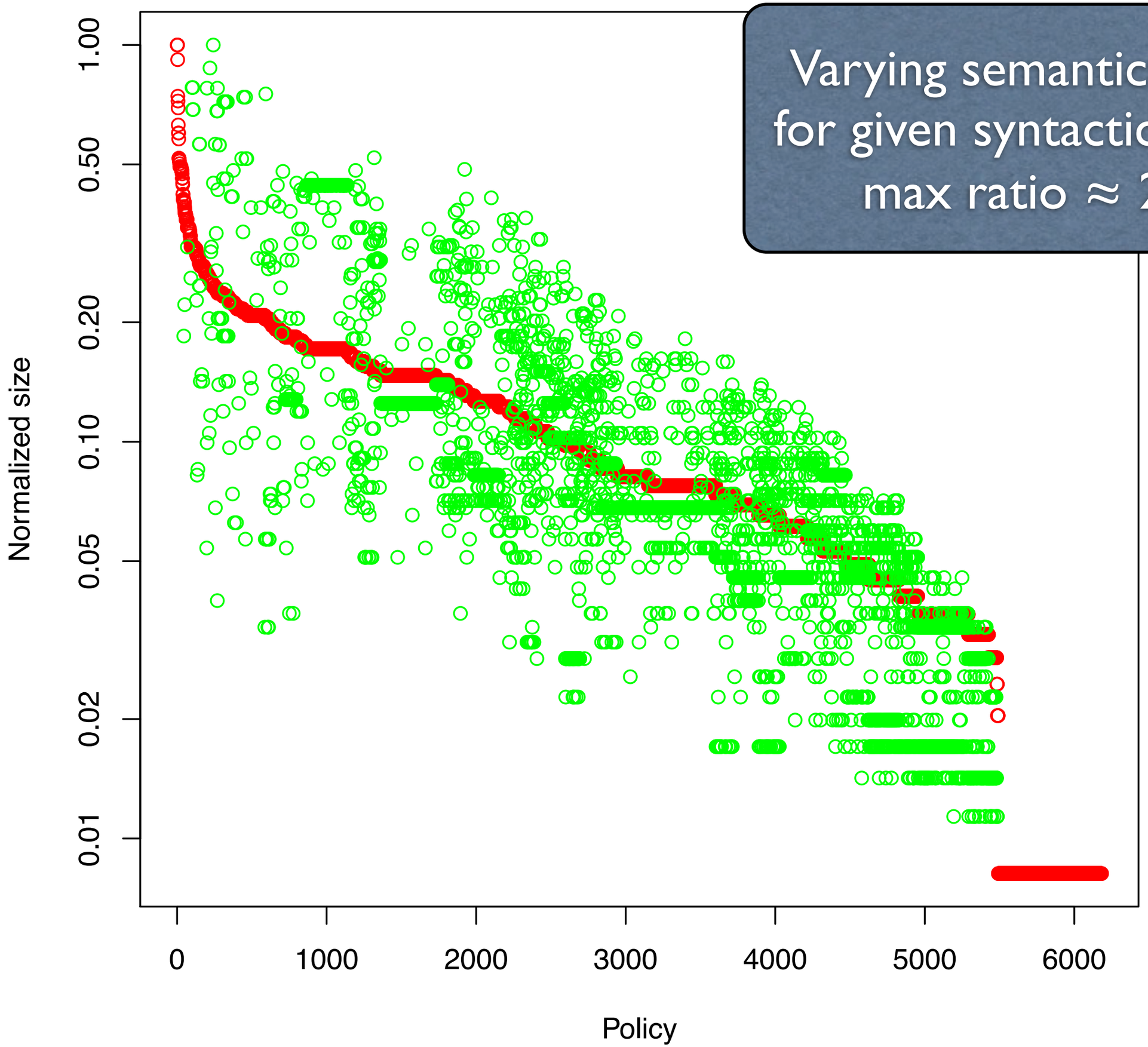
© 2010-2012, Software Languages Team, University of Koblenz-Landau

Metrics

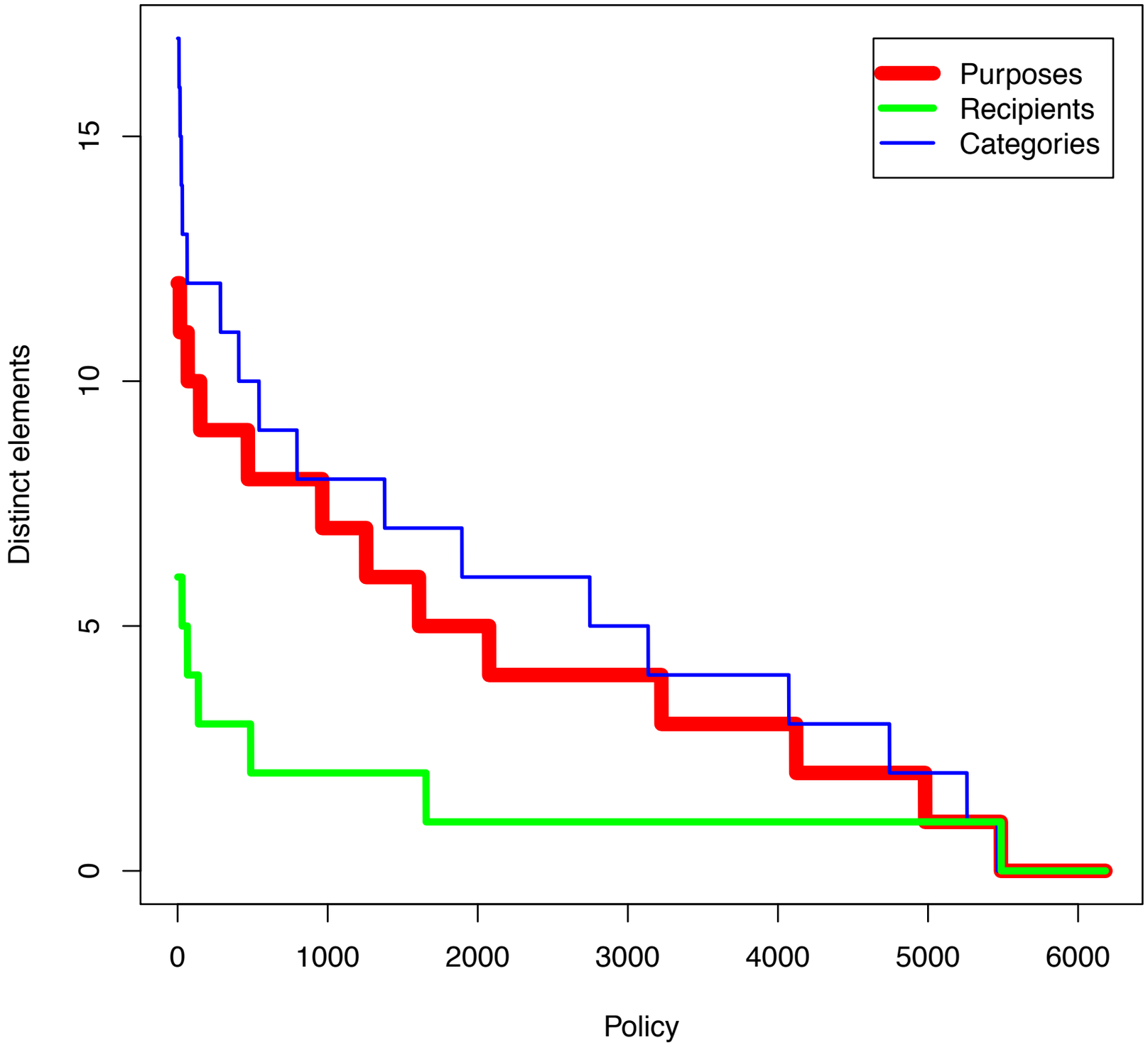
- Syntactical size
- Semantical size
- Vocabulary size
- Data size

© 2010-2012, Software Languages Team, University of Koblenz-Landau

Varying semantical size for given syntactical size; max ratio ≈ 22.



© 2010-2012, Software Languages Team, University of Koblenz-Landau



© 2010-2012, Software Languages Team, University of Koblenz-Landau

Determine significant policies

- Short-list top-10 for all 4 metrics.
- Consider policies in more than 1 short list.
- De-prioritize syntactical size.
- Also, include top position on vocabulary size.

© 2010-2012, Software Languages Team, University of Koblenz-Landau

The sites with the most significant policies

- http://www.astrodata.ch

astrology

- http://www.50cent.com

rapping

- http://kerntrophies.com/

trophies

© 2010-2012, Software Languages Team, University of Koblenz-Landau

The most significant policies

- http://www.astrodata.ch/w3c/policy-general.xml
- http://privacypolicy.umusic.com/w3c/p3p.xml
- http://kerntrophies.com/w3c/p3p.xml

"404"
per 28 Feb 2012

© 2010-2012, Software Languages Team, University of Koblenz-Landau

Leveraged analyses

- Analysis of vocabulary
- Analysis of constraints
- Analysis of metrics
- **Analysis of cloning**
- Analysis of extensions

© 2010-2012, Software Languages Team, University of Koblenz-Landau

Clone detection

- No parametrized clones
- No fragment clones
- Classification of clones
  - ***Textual*** clones
  - ***Syntactical*** clones
  - ***Semantical*** clones

© 2010-2012, Software Languages Team, University of Koblenz-Landau

Type	# Clone groups	# Added clone groups	# Added cloned policies	# Enlarged clone groups	# Merged clone groups	% Clones
txt	296	296	1,941	–	–	31.40
syn	496	280	1,857	46	114	30.04
sem	351	20	69	0	32	1.12
Total	–	–	3,867	–	–	62.55

We are left with 1,385 semantically distinct policies of 4,869 semantically valid policies of a total of 6,182 policies in the corpus.

© 2010-2012, Software Languages Team, University of Koblenz-Landau

Top-10 textual clone groups

	Entity	**Card.**	**Avg. dist.**	$SYN(\cdot)$	$SEM(\cdot)$	$\boldsymbol{VOCA}(\cdot)$	$\boldsymbol{DATA}(\cdot)$
1	John Hensley	175	1.000	42	156	0.650 (9; 3; 12)	188
2	CybrHost	139	1.000	2	0	0.000 (0; 0; 0)	0
3	PhotoBiz	137	1.000	2	0	0.000 (0; 0; 0)	0
4	Boatventures	96	1.000	15	16	0.163 (2; 1; 3)	3
5	Real Estate	69	1.000	42	–	0.363 (7; 1; 6)	–
6	Hilton Hotels	68	0.779	9	6	0.123 (2; 1; 1)	1
7	NASA	54	0.000	13	29	0.180 (4; 1; 1)	30
8	Rezidor SAS	51	0.000	12	16	0.163 (2; 1; 3)	29
9	Bravenet	46	0.000	17	16	0.123 (2; 1; 1)	29
10	Wetpaint	38	0.237	19	24	0.180 (4; 1; 1)	30
	Average	–	0.602	17.30	26.30	–	–

© 2010-2012, Software Languages Team, University of Koblenz-Landau

Top-10 syntactical clone groups

	Sample entity	**Card.**	**Avg. dist.**	$SYN(\cdot)$	$SEM(\cdot)$	$\boldsymbol{VOCA}(\cdot)$	$\boldsymbol{DATA}(\cdot)$
1	Accountancy	331	0.000	36	44	0.367 (8; 2; 2)	33
2	CybrHost	284	0.109	2	0	0.000 (0; 0; 0)	0
3	IBM tracking	218	0.124	19	24	0.180 (4; 1; 1)	30
4	Johnston Press	116	0.586	20	24	0.220 (4; 1; 3)	30
5	IBM logging	102	0.098	9	12	0.137 (3; 1; 0)	28
6	IBM purchase	67	0.104	51	–	0.363 (5; 2; 6)	–
7	Beach Suites	64	0.125	31	29	0.353 (6; 2; 4)	30
8	1066 Pools Ltd	30	0.000	24	10	0.377 (6; 1; 8)	1
9	Art of War	24	0.083	12	17	0.117 (1; 1; 2)	30
10	WebSolutions	20	0.000	31	–	0.283 (7; 1; 2)	–
	Average	–	0.123	23.50	16.00	–	–

© 2010-2012, Software Languages Team, University of Koblenz-Landau

Leveraged analyses

- Analysis of vocabulary
- Analysis of constraints
- Analysis of metrics
- Analysis of cloning
- **Analysis of extensions**

© 2010-2012, Software Languages Team, University of Koblenz-Landau

Example of an extension

```
<RETENTION>
  <legal-requirement/>
  <EXTENSION optional="no">
    <use-duration>
      <one-year/>
    </use-duration>
    <retention-basis>
      <internal-company-regulation> [Text] </internal-company-regulation>
    </retention-basis>
  </EXTENSION>
</RETENTION>
```

© 2010-2012, Software Languages Team, University of Koblenz-Landau

retention : (***retension-duration*** | ***retention-basis*** | ***use-duration***)*

credit-privacy-law	=	(**one-year**)*
e-trade-law	=	(**five-year**)*
internal-company-regulation	≐	ε
other-basis	≐	ε
other-duration	≐	ε
retension-duration	=	(***other-duration***)*
retention-basis	≐	(***credit-privacy-law*** | ***e-trade-law*** | ***internal-company-regulation*** | ...)*
use-duration	=	(***other-duration*** | **five-year** | **instance** | **one-month** | **one-year** | ...)*

© 2010-2012, Software Languages Team, University of Koblenz-Landau

statement : (***collection-method*** | ***destruction-method*** | group-info | statement-group)*

collection-method	=	(***other-method*** | **delivery** | **document** | **purnish** | **qnaboard** | **subscription** | ...)*
destruction-method	=	(***other-method*** | **format** | **shatter**)*
other-method	≐	ε

purpose : (***ppurpose*** | ***purpose*** | **age** | **cert** | **complaint** | **content** | **payback** | **proxy-agreement** | **statement**)*

ppurpose	≐	(**ACCOUNT** | **BROWSING** | **delivery** | **FEEDBACK** | **finmgt** | **government** | **login** | **marketing** | ...)*
purpose	=	(**FEEDBACK** | **government**)*

recipient : (***delivery*** | ***jurisdiction*** | ***recipient-description*** | ***recipient-duration*** | ***recipient-name*** | ***same*** | **delivery**)*

delivery	≐	ε
jurisdiction	=	(***long-description***)*
long-description	≐	ε
other-duration	≐	ε
other-purpose	≐	ε
recipient-description	=	(***other-purpose*** | **admin** | **age** | **agency** | **cert** | **complaint** | **content** | **delivery** | ...)*
recipient-duration	=	(***other-duration*** | **five-year** | **instance** | **one-year** | **six-month**)*
recipient-name	≐	ε
same	≐	ε

retention : (***retension-duration*** | ***retention-basis*** | ***use-duration***)*

credit-privacy-law	=	(**one-year**)*
e-trade-law	=	(**five-year**)*
internal-company-regulation	≐	ε
other-basis	≐	ε
other-duration	≐	ε
retension-duration	=	(***other-duration***)*
retention-basis	≐	(***credit-privacy-law*** | ***e-trade-law*** | ***internal-company-regulation*** | ...)*
use-duration	=	(***other-duration*** | **five-year** | **instance** | **one-month** | **one-year** | ...)*

data-group : (data-group)*

categories	=	(**COMPUTER** | **UNIQUEID**)*
category	=	(**COMPUTER** | demographic | navigation | purchase | state | **UNIQUEID**)*
data-group	=	(datatype)*
datatype	=	(dynamic | user)*
dynamic	=	(miscdata)*
home-info	=	(telecom)*
miscdata	=	(category)*
telecom	=	(**TELEPHONE**)*
user	=	(home-info)*

© 2010-2012, Software Languages Team, University of Koblenz-Landau

Results of the study

- Chosen approach to validation is too weak.
- There are common dominating policies.
- Interesting policies may be located through metrics.
- The study is non-conclusive on the need for extensions.
- The study promotes explorative, multi-dimensional analysis.

© 2010-2012, Software Languages Team, University of Koblenz-Landau

Thank you!

Let's discuss some more.

© 2010-2012, Software Languages Team, University of Koblenz-Landau