

19 September 2017

The Best Privacy Defense is a Good Privacy Offense

Obfuscating a Search Engine User's Profile

Jörg Simon Wicker

The University of Auckland

Stefan Kramer

Johannes Gutenberg University Mainz



Privacy Defence

- Privacy on the internet is an important and unsolved issue
- Privacy preserving data mining addresses this from a service provider perspective
- From a user's perspective, what can the user do to ensure the protection of his or her data?



Search Engines

Google





what to visit in macedonia

All Images Maps News Videos More Settings Tools

About 40,100,000 results (0.94 seconds)

Visit Macedonia - Small Group & Tailormade Tours
www.traveltheunknown.com/Macedonia/Tours
 Discover The Charms Of Macedonia - Highly Rated Tours - Trusted Tour Agency
 Financially Insured - Highly Recommended - Stunning Destinations - Fly From Anywhere - Group Tours
 Services: Group Tours, Adventure Holidays, Private Tours, Tailormade Tours
[The Balkan Odyssey](#) [Tour The Balkans](#)
[Balkan Explorer](#) [Balkans Archaeology](#)

Macedonia (FYROM) / Top sights

			
Lake Ohrid Lakes, monasteries, and churches	Matka Canyon Lakeside hotel with a wine cellar	Church of St. John at Kaneo Grand historic church with lake views	Samuel's Fortress, Ohrid History

[More Macedonia \(FYROM\) sights](#)

10 Places You Must Visit in Macedonia - David's Been Here
davidsevenhere.com/2015/12/30/10-places-you-must-visit-in-macedonia/
 Dec 30, 2015 - Other must do's in Ohrid include seeing the fresco of Saint Sofia (Sveti Sofia) Church

Search Engines

Google


what to visit in macedonia

All Images Maps News Videos More


About 40,100,000 results (0.94 seconds)

Visit Macedonia - Small Group & Tailormade Tours
Ad www.traveltheunknown.com/Macedonia/Tours
Discover The Charms Of Macedonia - Highly Rated Tours - Trusted Tour Agency
Financially Insured - Highly Recommended - Stunning Destinations - Fly From Any
Services: Group Tours, Adventure Holidays, Private Tours, Tailormade Tours
The Balkan Odyssey Balkan Explorer Tour The Balkans Balkans Archaeology


Macedonia (FYROM) / Top sights



Lake Ohrid
Lakes, monasteries,
and churches



Matka Canyon
Lakeside hotel with a
wine cellar



Church of St. John
at Kaneo
Grand historic church
with lake views

More Macedonia (FYROM) sights

b

ohrid macedonia

Web Images Videos News My saves

969,000 RESULTS Date Language Region

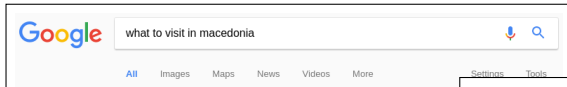
Car Rental in Ohrid, Macedonia - Best Car Rental Prices
Ad www.Rentalcars.com/Macedonia/Ohrid
Search, Compare and Save Using the World's Biggest Online Car Hire Service.
Brands: Toyota, Ford, Chevrolet, Chrysler, Jeep, Hyundai, Mercedes
"Travel Standard of Excellence - WebAward 2016" – WebAward 2016

Ohrid - Wikipedia
<https://en.wikipedia.org/wiki/Ohrid>

Overview Contents Name History Demographics

Ohrid is a city in the Republic of Macedonia and the seat of Ohrid Municipality. It is the largest city on Lake Ohrid and the eighth-largest city in the country, with over 42,000 inhabitants as of 2002. Ohrid is notable for once having had 365 churches, one for each day of the year, and has been referred to as a "Jerusalem". The city is rich in picturesque houses and monuments, and tourism is predominant. It is located southwe...

Search Engines



About 40,100,000 results (0.94 seconds)

Visit Macedonia - Small Group & Tailormade Tours

www.traveltheunknown.com/Macedonia/Tours
Discover The Charms Of Macedonia - Highly Rated Tours - Trusted Tour Agency Financially Insured - Highly Recommended - Stunning Destinations - Fly From Any Services: Group Tours, Adventure Holidays, Private Tours, Tailormade Tours

The Balkan Odyssey
Balkan Explorer

Tour The Balkans
Balkans Archaeology

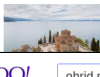
Macedonia (FYROM) / Top sights



Lake Ohrid
Lakes, monasteries,
and churches



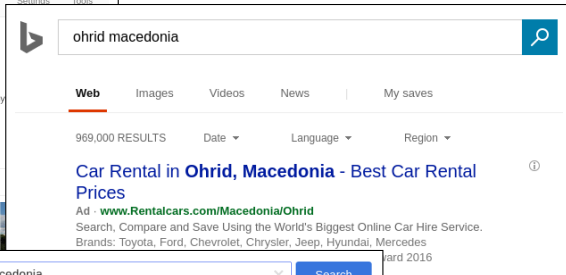
Matka C
Lakeside
wine cellar



More Macedonia (FYROM)

10 Places You Must Visit

davidbeenhere.com/2015/12/30/
Dec 20, 2015 - Other must do's in



ohrid macedonia

Web

Images

Videos

News

My saves

969,000 RESULTS

Date

Language

Region

Car Rental in Ohrid, Macedonia - Best Car Rental Prices

Ad - www.Rentalcars.com/Macedonia/Ohrid

Search, Compare and Save Using the World's Biggest Online Car Hire Service.

Brands: Toyota, Ford, Chevrolet, Chrysler, Jeep, Hyundai, Mercedes

ward 2016

YAHOO!
NEW ZEALAND

ohrid macedonia

Search

Web

Images

News

Answers

Recipes

Sports

Finance

Ads related to: ohrid macedonia

Ohrid Macedonia - Top 10 Hotels (with Prices) - tripadvisor.co.nz

www.tripadvisor.co.nz/

Save money & book with TripAdvisor® (the world's largest travel website).

Millions of hotel reviews - Easy price comparison - Candid traveller photos

380 Hotels in Ohrid - Lowest price guarantee - booking.com

www.booking.com/Ohrid-Hotels

Book your Hotel in Ohrid online

ory

Demographics

Ohrid Municipality. It is the country, with over 42,000 365 churches, one for each The city is rich in picturesque cated southwe...

Search Engines

Google

what to visit in macedonia

All Images Maps News Videos More

About 40,100,000 results (0.94 seconds)

Visit Macedonia - Small Group & Tailormade Tours
www.traveltheunknown.com/Macedonia/Tours
 Discover The Charms Of Macedonia - Highly Rated Tours - Trusted Tour Agency
 Financially Insured - Highly Recommended - Stunning Destinations - Fly From Any
 Services: Group Tours, Adventure Holidays, Private Tours, Tailormade Tours
 The Balkan Odyssey Tour The Balkans
 Balkan Explorer Balkans Archaeology

Macedonia (FYROM) / Top sights

Lake Ohrid
Lakes, monasteries,
and churches

Matka C
Lakeside
wine cellar

More Macedonia (FYROM)

10 Places You Must Visit
davidsevenhere.com/2015/12/30/
 Dec 20, 2015 - Other must do's for

b

ohrid macedonia

Web Images Videos News My saves

969,000 RESULTS Date Language Region

Car Rental in Ohrid, Macedonia - Best Car Rental Prices
 Ad - www.Rentalcars.com/Macedonia/Ohrid
 Search, Compare and Save Using the World's Biggest Online Car Hire Service.
 Brands: Toyota, Ford, Chevrolet, Chrysler, Jeep, Hyundai, Mercedes

ohrid macedonia

Search

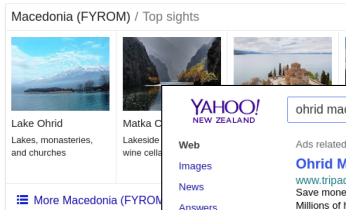
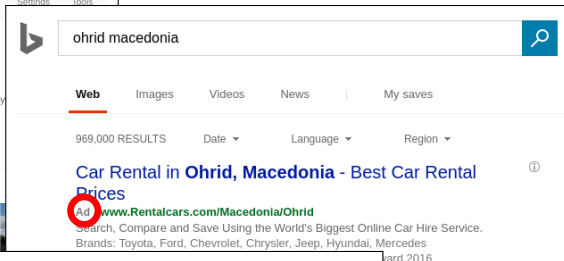
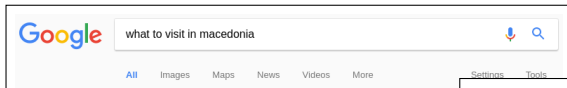
Ads related to: ohrid macedonia

Ohrid Macedonia - Top 10 Hotels (with Prices) - tripadvisor.co.nz
www.tripadvisor.co.nz/
 Save money & book with TripAdvisor® (the world's largest travel website).
 Millions of hotel reviews - Easy price comparison - Candid traveller photos

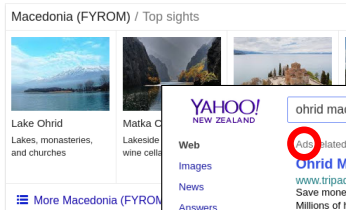
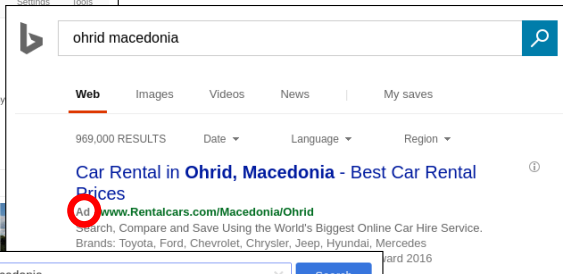
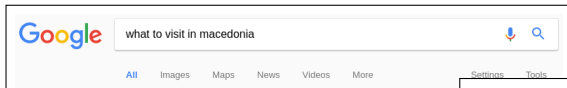
380 Hotels in Ohrid - Lowest price guarantee - booking.com
www.booking.com/Ohrid-Hotels
 Book your Hotel in Ohrid online

Ohrid Municipality. It is the
country, with over 42,000
365 churches, one for each
The city is rich in picturesque
cated southwe...

Search Engines

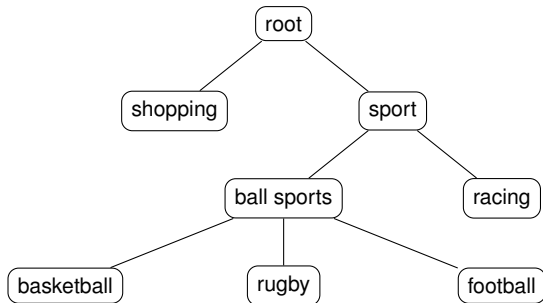


Search Engines



Personalised Advertisement

- Which ad is displayed depends on
 1. The submitted query
 2. **The user profile**
- Ads are assigned to categories
- Users are assigned to categories



Privacy Offense

- To implement a method to defend privacy, we need

Privacy Offense

- To implement a method to defend privacy, we need
 - A way to measure the privacy, i.e. objective function

Privacy Offense – Goal

- Objective function:

$$\sigma(\kappa_i, P) = \sum_{p_j \in P, \kappa_j \in T} p_j d_T(\kappa_i, \kappa_j)$$

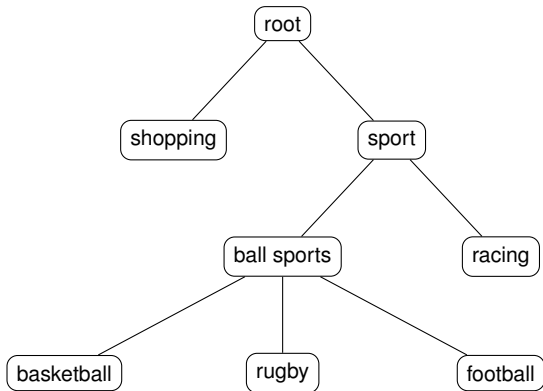
- User interest category κ , distribution of probabilities P , category tree T , tree distance d_T
- Score σ is the weighted distance between user interest category and current category the user is assigned to

Privacy Offense

- To implement a method to defend privacy, we need
 - A way to measure the privacy, i.e. objective function
 - Method to use the feedback (ads)

Category Prediction of an Ad

- Search engines provide example queries for each category
- Use sample queries of category tree as input and train independent classifiers – one for each category
- Classifiers can be applied to queries, as well as any other text
- Predictions on ads work very well due to similar structure of the text



Privacy Offense

- To implement a method to defend privacy, we need
 - A way to measure the privacy, i.e. objective function
 - Method to use the feedback (ads)
 - A set of actions

Actions

Definition of actions to choose one category κ in the set of categories K using category tree T based on reference category κ_{ref}

Actions

Definition of actions to choose one category κ in the set of categories K using category tree T based on reference category κ_{ref}

Random: $a_{random}(T, \kappa_{ref}) = random_select(\kappa_r \in K)$

Actions

Definition of actions to choose one category κ in the set of categories K using category tree T based on reference category κ_{ref}

Random: $a_{random}(T, \kappa_{ref}) = random_select(\kappa_r \in K)$

Same: $a_{same}(T, \kappa_{ref}) = \kappa_{ref}$

Actions

Definition of actions to choose one category κ in the set of categories K using category tree T based on reference category κ_{ref}

Random: $a_{random}(T, \kappa_{ref}) = random_select(\kappa_r \in K)$

Same: $a_{same}(T, \kappa_{ref}) = \kappa_{ref}$

Sibling: $a_{sibling}(T, \kappa_{ref}) = sibling(\kappa_{ref} \in K)$

Actions

Definition of actions to choose one category κ in the set of categories K using category tree T based on reference category κ_{ref}

Random: $a_{random}(T, \kappa_{ref}) = random_select(\kappa_r \in K)$

Same: $a_{same}(T, \kappa_{ref}) = \kappa_{ref}$

Sibling: $a_{sibling}(T, \kappa_{ref}) = sibling(\kappa_{ref} \in K)$

Most general: $a_{general}(T, \kappa_{ref}) = max_parent(\kappa_{ref} \in K)$

Actions

Definition of actions to choose one category κ in the set of categories K using category tree T based on reference category κ_{ref}

Random: $a_{random}(T, \kappa_{ref}) = random_select(\kappa_r \in K)$

Same: $a_{same}(T, \kappa_{ref}) = \kappa_{ref}$

Sibling: $a_{sibling}(T, \kappa_{ref}) = sibling(\kappa_{ref} \in K)$

Most general: $a_{general}(T, \kappa_{ref}) = max_parent(\kappa_{ref} \in K)$

Most specialized of sibling: $a_{specialized}(T, \kappa_{ref}) = lowest_child(all_siblings(\kappa_{ref} \in K))$

Actions

Definition of actions to choose one category κ in the set of categories K using category tree T based on reference category κ_{ref}

Random: $a_{random}(T, \kappa_{ref}) = random_select(\kappa_r \in K)$

Same: $a_{same}(T, \kappa_{ref}) = \kappa_{ref}$

Sibling: $a_{sibling}(T, \kappa_{ref}) = sibling(\kappa_{ref} \in K)$

Most general: $a_{general}(T, \kappa_{ref}) = max_parent(\kappa_{ref} \in K)$

Most specialized of sibling: $a_{specialized}(T, \kappa_{ref}) = lowest_child(all_siblings(\kappa_{ref} \in K))$

Distance-based: $a_{dist}(T, \kappa_{ref}) = \kappa_r : \forall \kappa_t \in K, d(\kappa_r, \kappa_{ref}) \geq d(\kappa_t, \kappa_{ref})$

Actions

Definition of actions to choose one category κ in the set of categories K using category tree T based on reference category κ_{ref}

Random: $a_{random}(T, \kappa_{ref}) = random_select(\kappa_r \in K)$

Same: $a_{same}(T, \kappa_{ref}) = \kappa_{ref}$

Sibling: $a_{sibling}(T, \kappa_{ref}) = sibling(\kappa_{ref} \in K)$

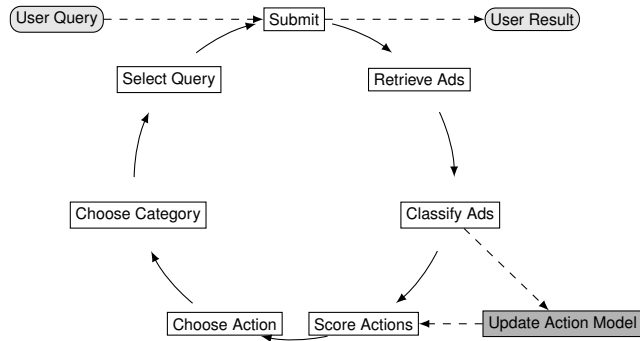
Most general: $a_{general}(T, \kappa_{ref}) = max_parent(\kappa_{ref} \in K)$

Most specialized of sibling: $a_{specialized}(T, \kappa_{ref}) = lowest_child(all_siblings(\kappa_{ref} \in K))$

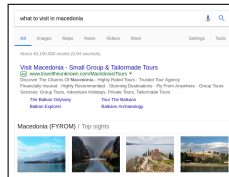
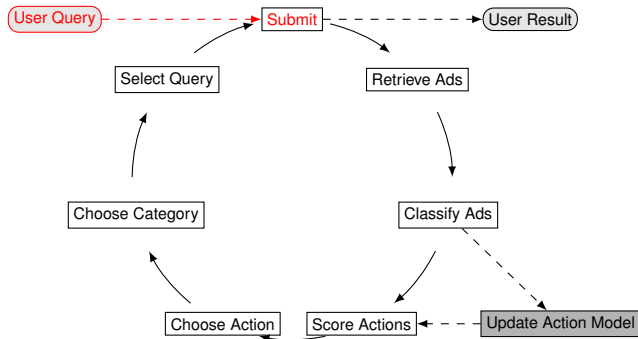
Distance-based: $a_{dist}(T, \kappa_{ref}) = \kappa_r : \forall \kappa_t \in K, d(\kappa_r, \kappa_{ref}) \geq d(\kappa_t, \kappa_{ref})$

- Jaccard Distance
- Normalized Mutual Information

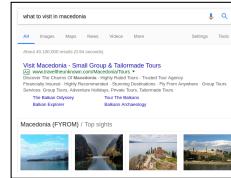
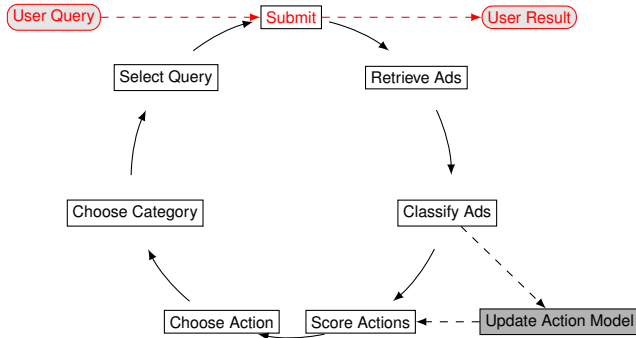
Privacy Offense – Method



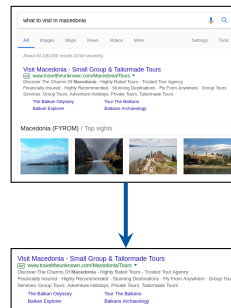
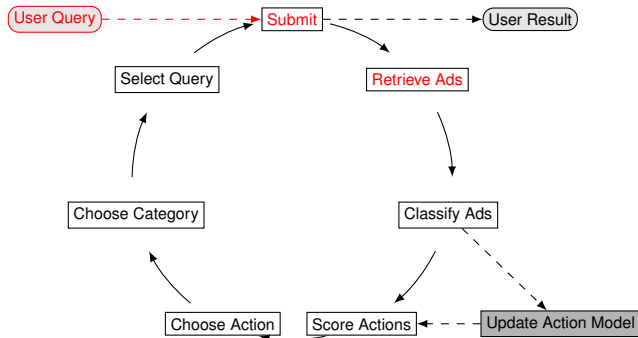
Privacy Offense – Method



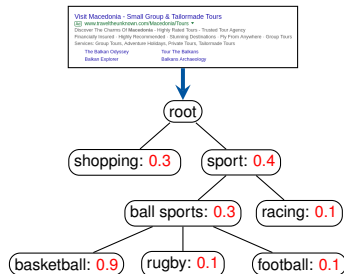
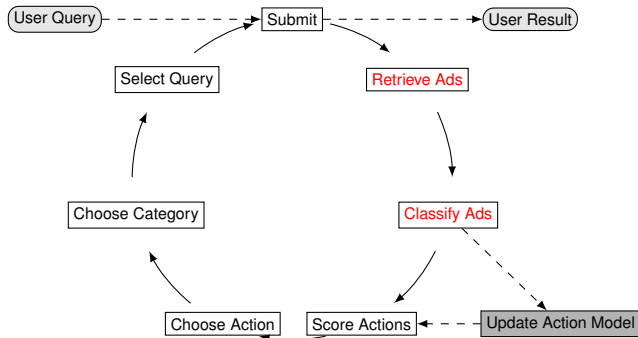
Privacy Offense – Method



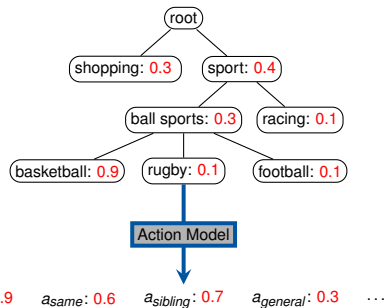
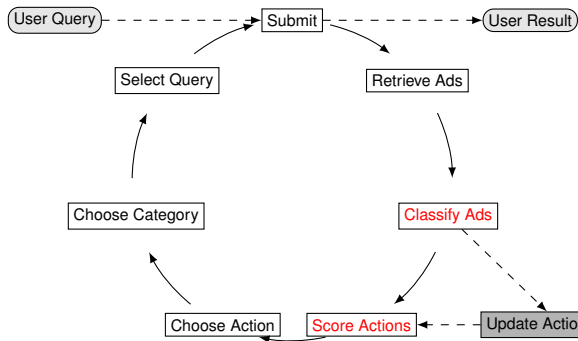
Privacy Offense – Method



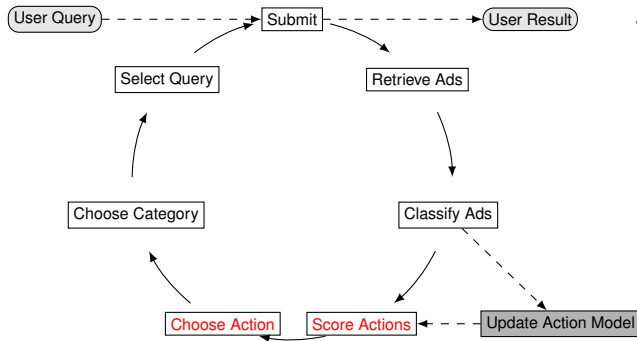
Privacy Offense – Method



Privacy Offense – Method



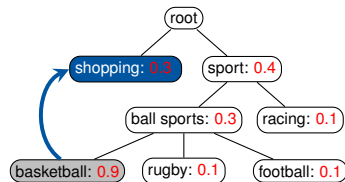
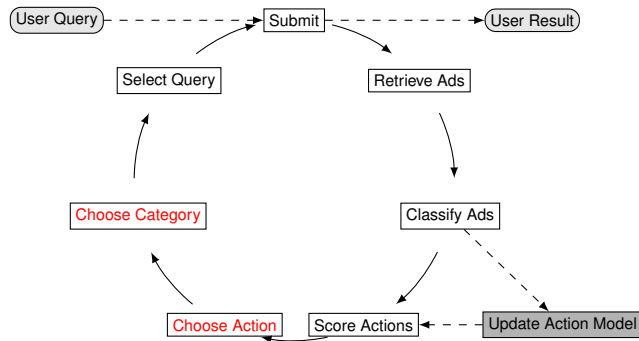
Privacy Offense – Method



$a_{random}: 0.9$ $a_{same}: 0.6$ $a_{sibling}: 0.7$ $a_{general}: 0.3$...

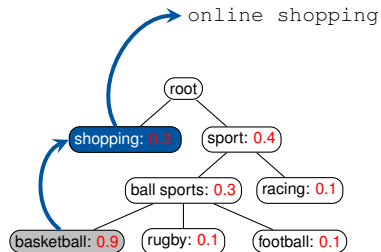
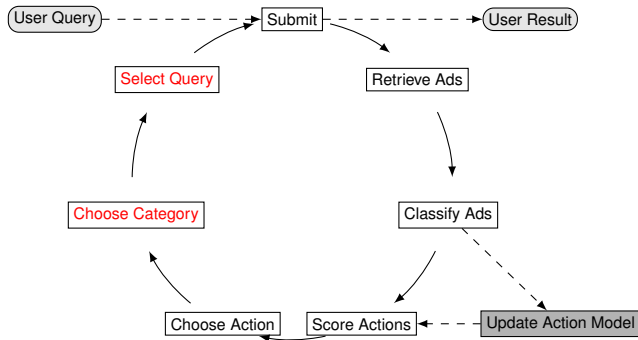
$$a_{random}(T, \kappa_{ref}) = \text{random_select}(\kappa_r \in K)$$

Privacy Offense – Method

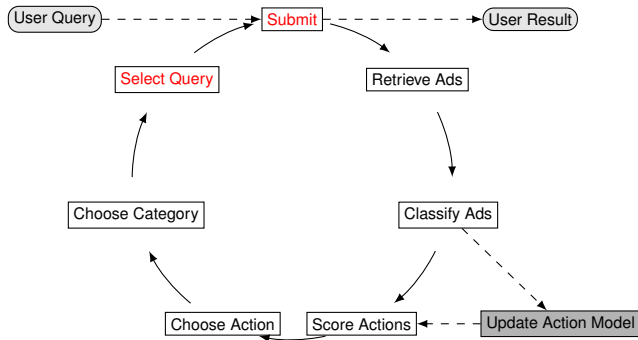


$$a_{\text{random}}(T, \kappa_{\text{ref}}) = \text{random_select}(\kappa_r \in K)$$

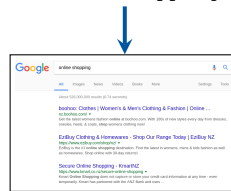
Privacy Offense – Method



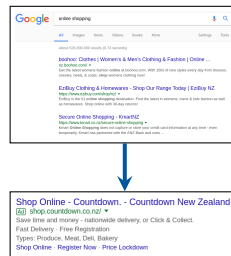
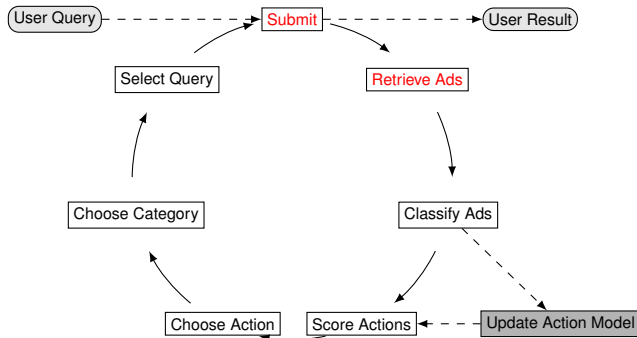
Privacy Offense – Method



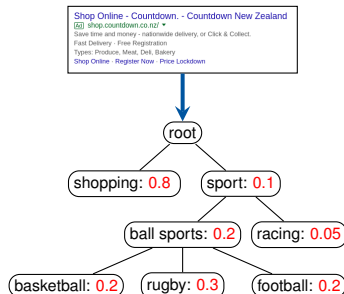
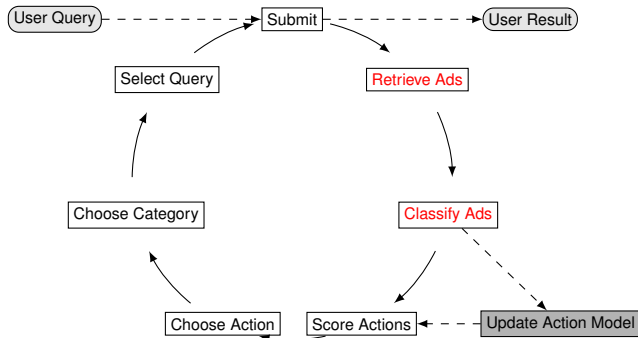
online shopping



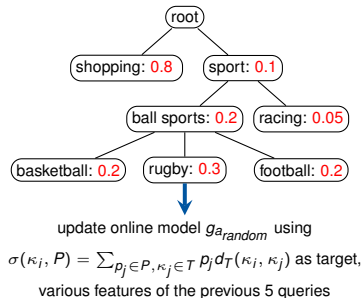
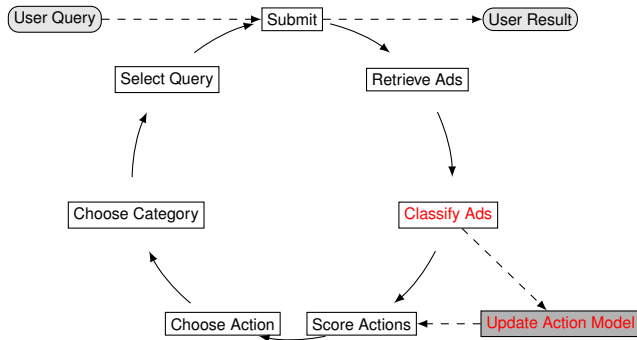
Privacy Offense – Method



Privacy Offense – Method



Privacy Offense – Method



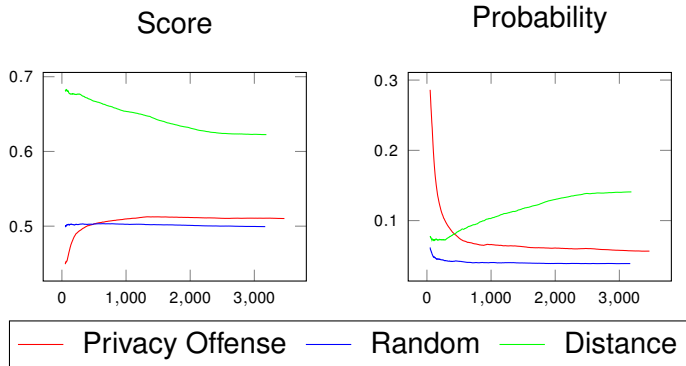
Experiments

- Users are given one interest category and either
 - Use the proposed method, or
 - Submit queries from random categories, or
 - Submit queries from the category that is the furthest away from their interest category
- All users submit in 10% of the cases random queries from their interest category

- 20 categories where used:

antiques and collectibles	bicycles and accessories
car video	computer components
cosmetic procedures	dating and personals
desktop computers	drugs and medications
erectile dysfunction	family
game systems and consoles	laptops and notebooks
make up and cosmetics	motorcycles
real estate listings	sexual enhancement
timeshares and vacation properties	toys
vitamins and supplements	weight loss

Results



All users, queries **not** in interest category

Conclusion

- Does it work?

Conclusion

- Does it work?
 - Maybe

Conclusion

- Does it work?
 - Maybe
 - But: We can trigger and see a reaction

Conclusion

- Does it work?
 - Maybe
 - But: We can trigger and see a reaction
- Simplified model

Conclusion

- Does it work?
 - Maybe
 - But: We can trigger and see a reaction
- Simplified model
 - Only one interest category

Conclusion

- Does it work?
 - Maybe
 - But: We can trigger and see a reaction
- Simplified model
 - Only one interest category
 - Discard more aspects of the search engine's model, e.g., time and date of query

Conclusion

- Does it work?
 - Maybe
 - But: We can trigger and see a reaction
- Simplified model
 - Only one interest category
 - Discard more aspects of the search engine's model, e.g., time and date of query
- Future work

Conclusion

- Does it work?
 - Maybe
 - But: We can trigger and see a reaction
- Simplified model
 - Only one interest category
 - Discard more aspects of the search engine's model, e.g., time and date of query
- Future work
 - More sophisticated model

Conclusion

- Does it work?
 - Maybe
 - But: We can trigger and see a reaction
- Simplified model
 - Only one interest category
 - Discard more aspects of the search engine's model, e.g., time and date of query
- Future work
 - More sophisticated model
 - Use more feedback than just the ads

Conclusion

- Does it work?
 - Maybe
 - But: We can trigger and see a reaction
- Simplified model
 - Only one interest category
 - Discard more aspects of the search engine's model, e.g., time and date of query
- Future work
 - More sophisticated model
 - Use more feedback than just the ads
 - Extend the use beyond search engines



SCIENCE



JOHANNES GUTENBERG
UNIVERSITÄT MAINZ

Thank you for your attention! Any questions?

<https://joerg-wicker.org>