# Détection et analyse de l'activité pédophile dans les ensembles de requêtes P2P

Raphaël Fournier-S'niehotta




Journées ResCom

29 novembre 2012

# Plan

# Plan

# Grands ensembles de requêtes

- Interaction utilisateur-moteur de recherche
- Des applications
  - classiques (amélioration de systèmes)
  - moins classiques (suivi de la grippe)
- Séquence de requêtes : $q_i = (t, u, k_1, k_2, \ldots, k_n)$
  - $t$ horodatage
  - $u$ information sur l'émetteur (adresse IP, port)
  - $(k_1, k_2, \ldots, k_n)$ suite de mots-clefs

# L'activité pédophile dans le P2P

## Problème important

- Victimes directes
- Danger pour les utilisateurs non pédophiles
- Impact sur la régulation de l'Internet

Très peu de connaissances

# Objectifs

Améliorer la connaissance
de l'activité pédophile dans le P2P

## Détection

- Élaborer une méthodologie générale
- Créer un outil de détection automatisé

## Analyse

- Dénombrer rigoureusement les requêtes
- Étudier les utilisateurs

# Problématiques

- Collecte de données adaptées

  taille, dynamicité, protocoles peu documentés

- Outil de détection automatique

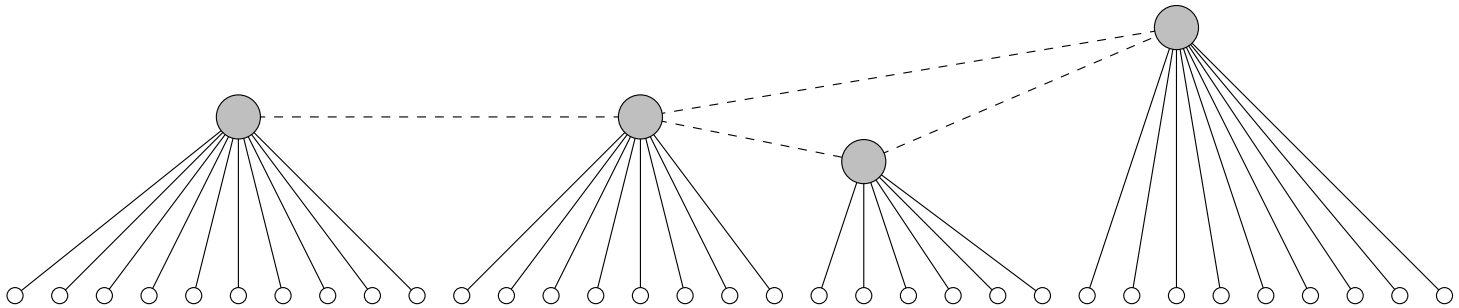
  activité cachée, langues multiples

- Inférence statistique rigoureuse

  faible quantité de requêtes pédophiles

# Données

- eDonkey (*eMule*, *MLDonkey*, *Shareaza*)



| | Durée | Nb. requêtes | Nb. IP |
|---|---|---|---|
| 2007 | 10 semaines | 107 226 021 | 23 892 531 |
| 09-12 | 147 semaines | 1 290 377 956 | 82 264 897 |
| 2009 | 28 semaines | 205 228 820 | 24 413 195 |

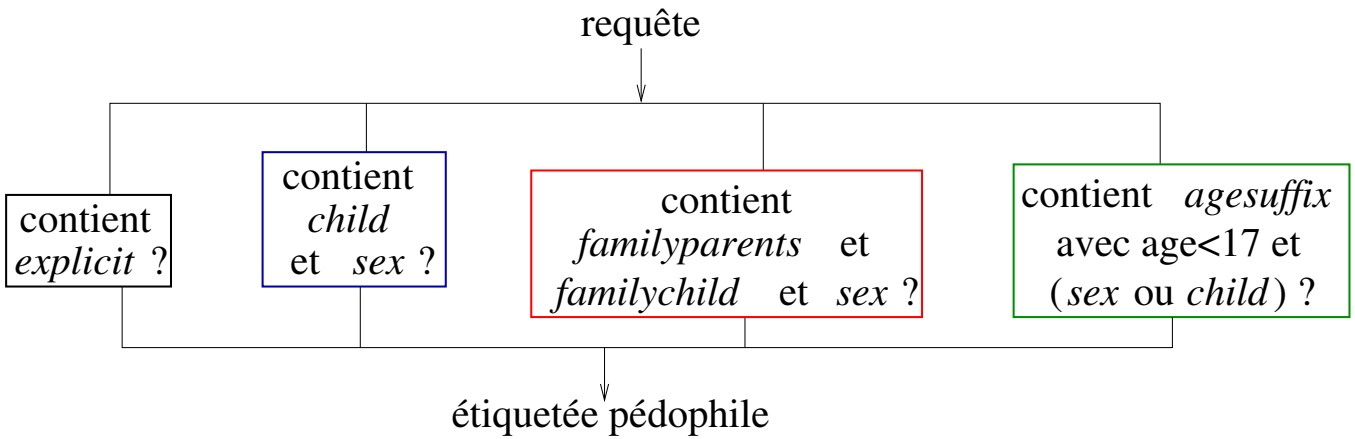
- Normalisation et anonymisation des données brutes

F. Aidouni, M. Latapy, and C.Magnien. Ten weeks in the life of an edonkey server. *Proceedings of HotP2P'09*, 2009.

# Plan

# Conception de l'outil

- 4 types de requêtes pédophiles




raygold little girl

porno infantil

incest mom son video

12yo fuck video

# Évaluation de la qualité

## Faux positifs

*“sexy daddy destinys child”*
contient “sexy”, “daddy” et “child” $\longrightarrow$ étiquetée pédophile
probablement une recherche liée à la musique

## Faux négatifs

*“pjk 12yo”* $\longrightarrow$ étiquetée non pédophile
contient un marqueur pédophile non connu

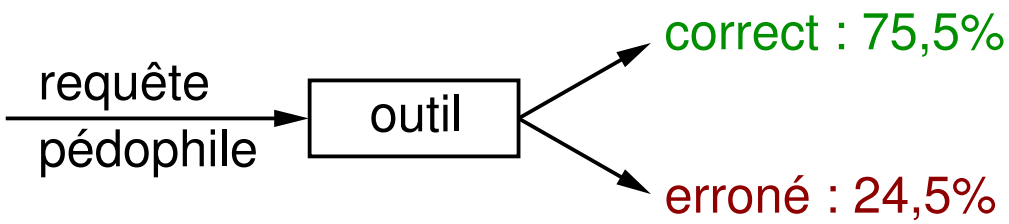
Comment estimer ces taux de faux positifs et faux négatifs ?

# Validation – Sondage

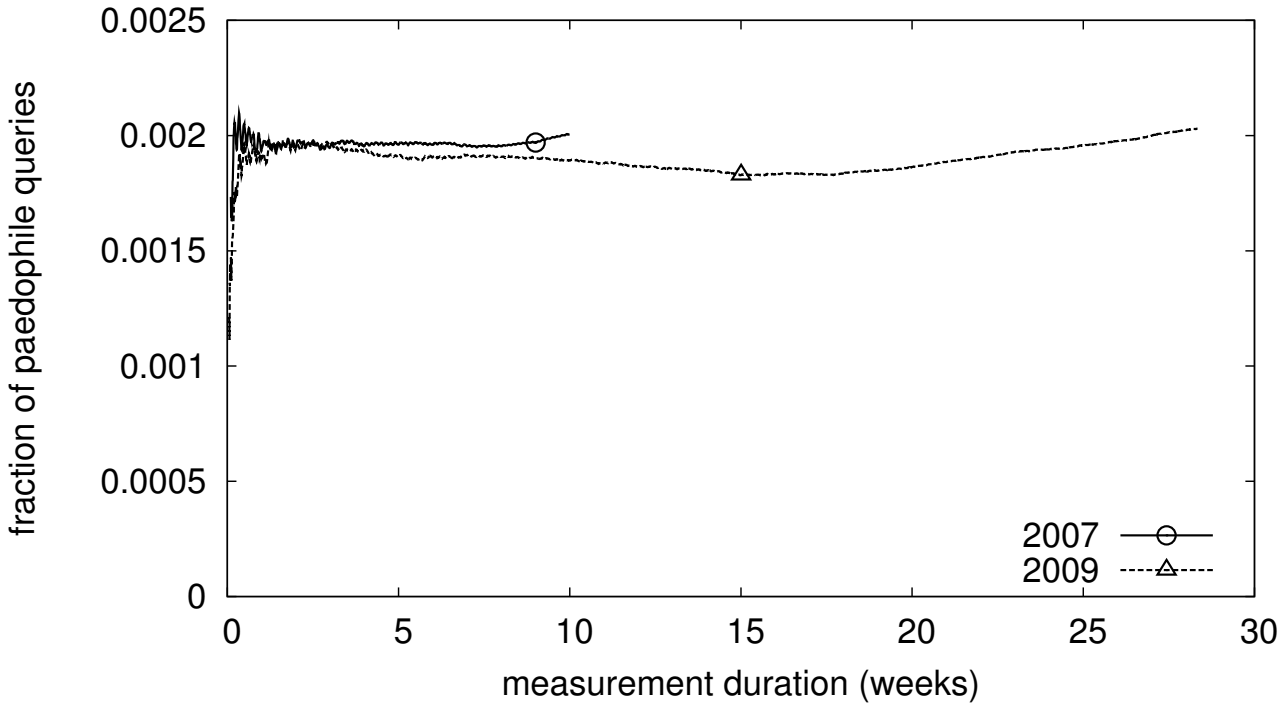
- 21 experts volontaires (Europol, forces de l'ordre, ONG)
- 3 000 requêtes choisies aléatoirement dont :
  - 1 000 étiquetées pédophiles
  - 1 000 étiquetées non pédophiles
  - 1 000 *voisines* (soumises dans les 2h avant ou après une requête étiquetée pédophile, par la même adresse IP)

| *pédo* | *prob. pédo* | *je ne sais pas* | *prob. pas* | *pas pédo* | total | pertinence |
|---|---|---|---|---|---|---|
| ... | ... | ... | ... | ... | ... | ... |
| 1174 | 111 | 20 | 64 | 789 | 2158 | 99.1 |
| ... | ... | ... | ... | ... | ... | ... |

# Résultats de la validation




# Fraction de requêtes pédophiles




## Résultat

- détection : légèrement au-dessus de 1,9 pour 1 000
- après correction : 2,5 requêtes pour 1 000 sont pédophiles
- 1 requête pédophile toutes les 33 secondes environ

# Plan

# Notion d'utilisateur

Hypothèse classique :
utilisateur $\sim$ adresse IP

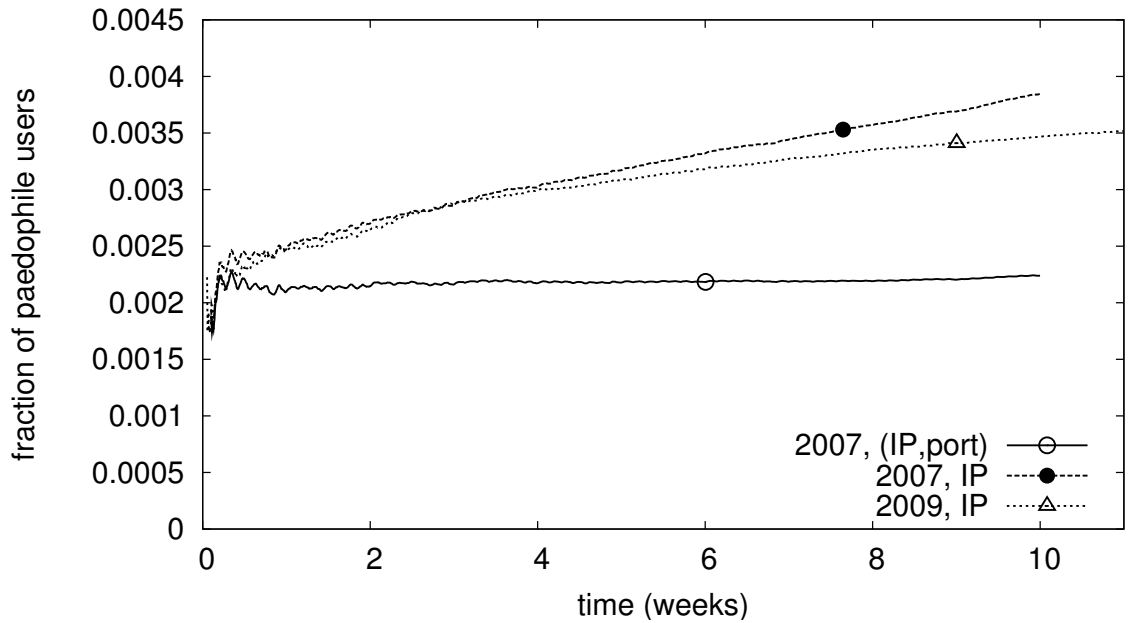
## Problèmes

- traduction d'adresse (NAT)
- renouvellement d'adresses
- plusieurs utilisateurs par ordinateur
- plusieurs ordinateurs par utilisateur

## Améliorations

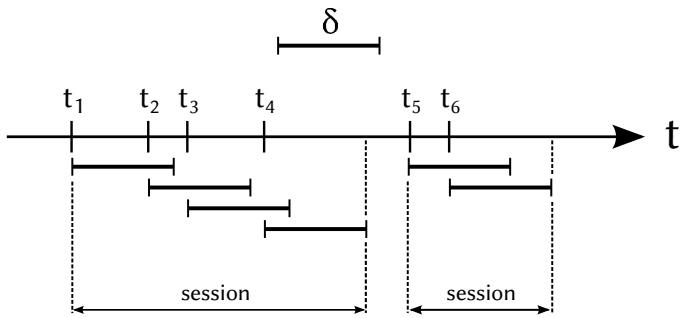
- utilisateur $\sim$ adresse IP + port de connexion
- durée de la mesure
- sessions temporelles

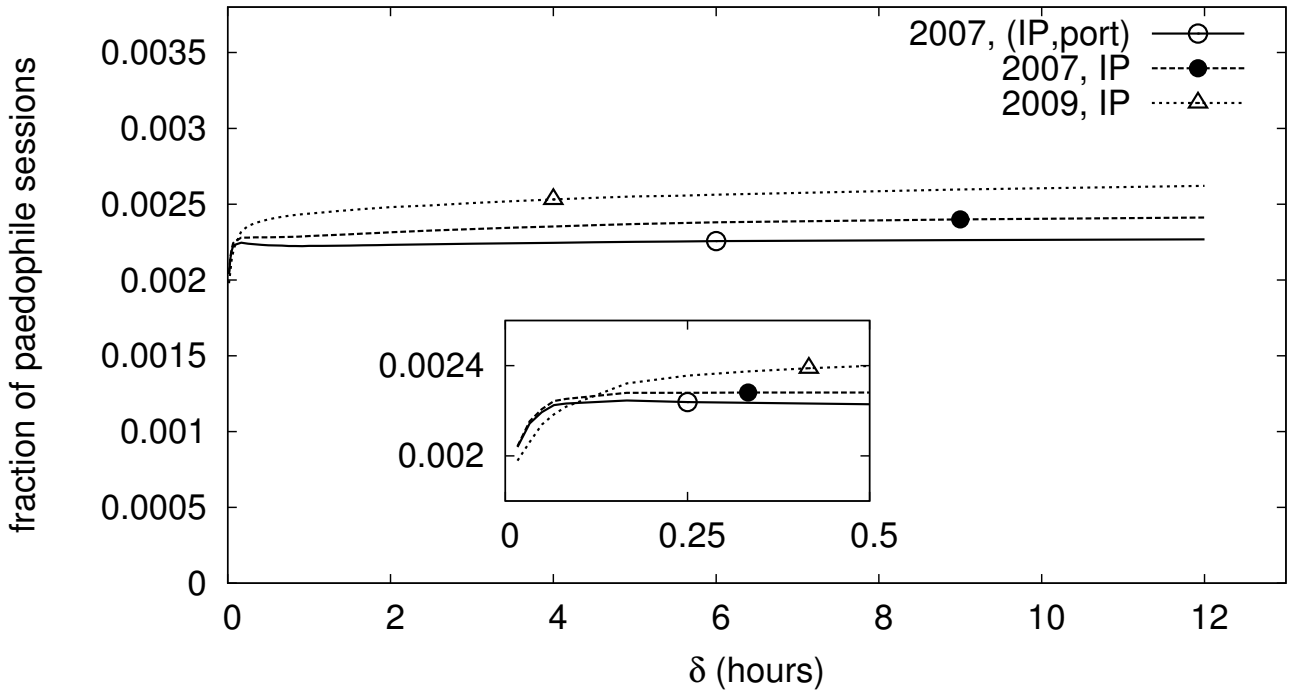
# Notion d'utilisateur : IP *vs* (IP,port)




- hypothèse : un utilisateur est pédophile s'il a fait une requête pédophile
- pollution : toutes les adresses IP vues comme pédophiles, après *un certain temps*
- convergence quand on prend l'hypothèse (IP, port)

# Notion d'utilisateur : sessions temporelles



# Fraction d'utilisateurs pédophiles

- faux positifs et négatifs sur les utilisateurs

$$p(u \in U^+ \mid u \in V(n,0)) = 1 - (1 - f'^-)^n$$

$$p(u \in U^- \mid u \in V(n,k)) = (f'^+)^k (1 - f'^-)^{n-k}$$

**Résultat**

Fraction d'utilisateurs pédophiles proche de 0,22%

# Plan

# Conclusion

Un cas de détection d'une thématique rare
dans de grands ensembles de requêtes

## Contributions

- Outil de détection de requêtes pédophiles
- Grand ensemble de requêtes pédophiles
- Estimation de la fraction de requêtes pédophiles
- Étude de la notion d'utilisateur
- Estimation de la fraction de requêtes pédophiles

## Contributions non présentées

- Dynamique temporelle de l'activité pédophile
- Comparaison avec le réseau P2P KAD

# Perspectives

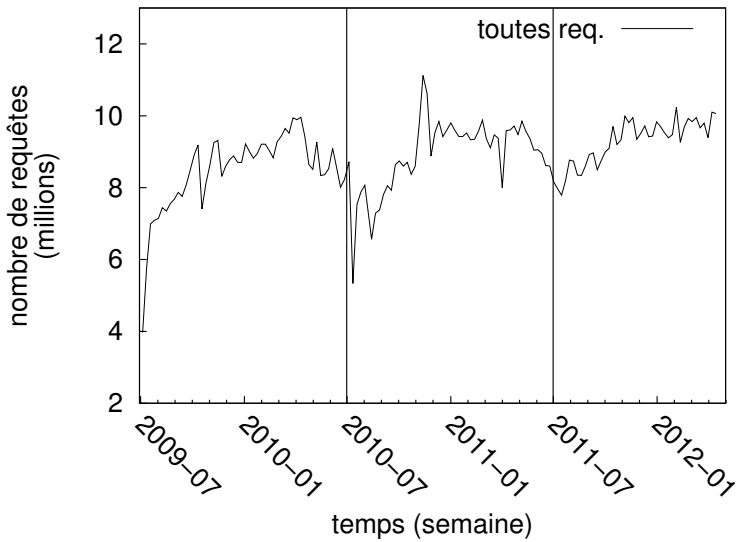
## Amélioration de l'outil de détection

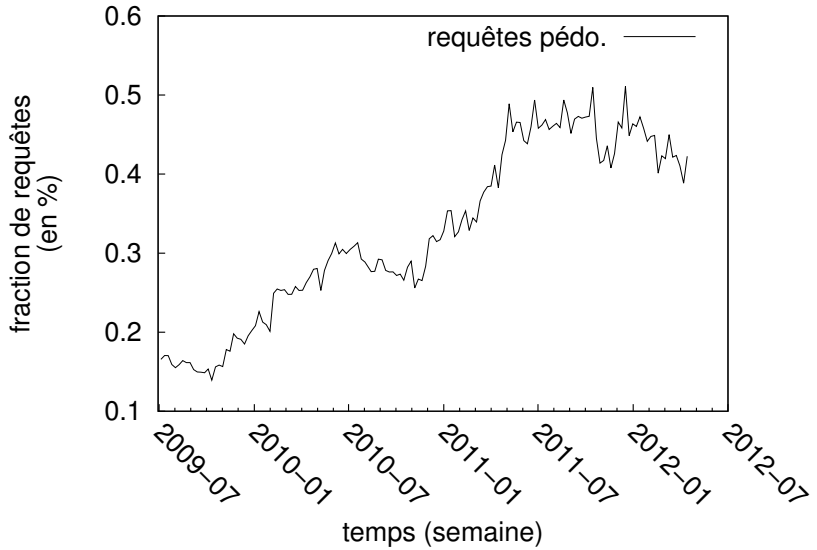
- requêtes précédente/suivante
- langues, ordre des mots, catégories
- apprentissage

## Analyse des utilisateurs

- seuil différent pour être considéré comme pédophile
- confrontation avec d'autres systèmes
- recherche de communautés
- étude détaillée des séquences de requêtes

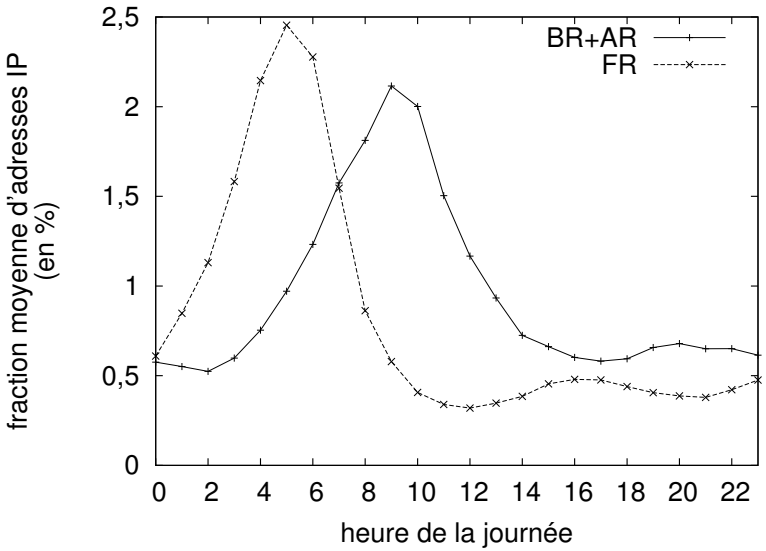
# Évolution sur une longue période

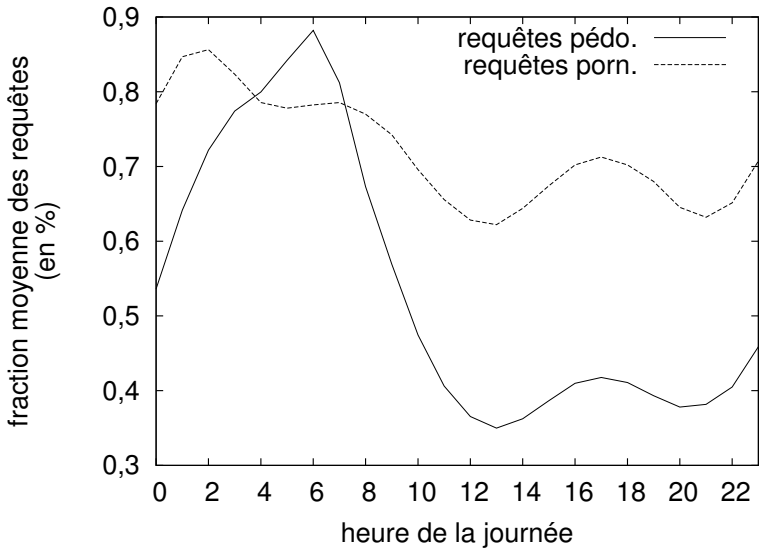



- trafic global stable sur 3 ans
- trafic pédophile en forte croissance
- augmentation du nombre d'utilisateurs pédophiles

# Dynamique journalière




- effet jour/nuit du trafic
- pic de fraction de requêtes pédophiles vers 6 heures
- différent pour les requêtes pornographiques

# KAD network

- Completely distributed protocol of clients
- No server for file indexing
- Some peers are in charge of some files and keywords

## Principle:

- Precise and targeted injection of peers into the network to control files or keywords
- Peers catch queries and control replies

## Applications:

- Which files are published for a given keyword? Which peers share them ?
- Eclipse : prevent peers from accessing content

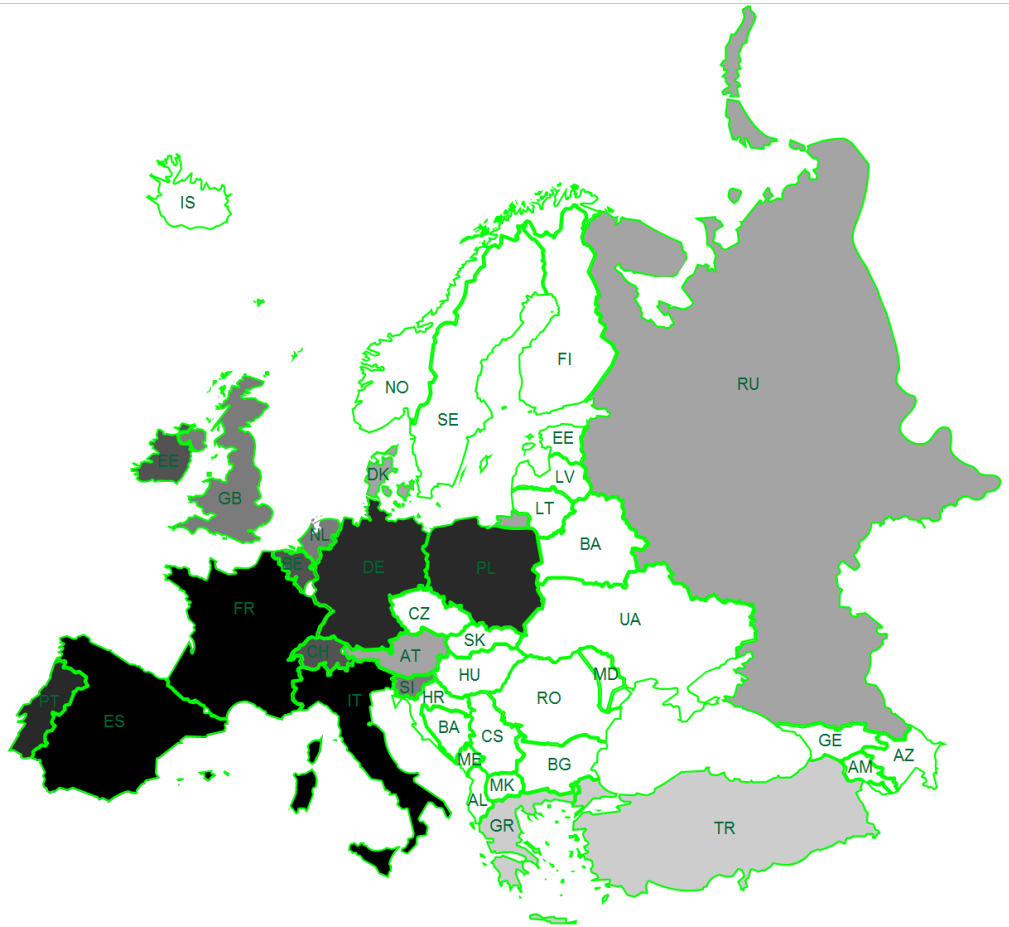
# Geo-location: statistics

| country | # queries | # paedo | ratio |
|---|---|---|---|
| IT | 19569361 | 15426 | 0.08 % |
| ES | 8881405 | 5177 | 0.06 % |
| FR | 7583815 | 8059 | 0.11 % |
| BR | 2795090 | 4849 | 0.17 % |
| IL | 2139697 | 2618 | 0.12 % |
| DE | 2093106 | 11238 | 0.54 % |
| KR | 1386799 | 336 | 0.02 % |
| US | 1053183 | 6184 | 0.59 % |
| PL | 975170 | 1178 | 0.12 % |
| AR | 810466 | 1465 | 0.18 % |
| CN | 635392 | 337 | 0.05 % |
| PT | 513327 | 434 | 0.08 % |
| IE | 511185 | 54 | 0.01 % |
| TW | 417893 | 138 | 0.03 % |
| BE | 402565 | 646 | 0.16 % |
| CH | 320054 | 1710 | 0.53 % |
| GB | 319386 | 1698 | 0.53 % |
| NL | 243646 | 1131 | 0.46 % |
| CA | 241460 | 1233 | 0.51 % |
| SI | 239572 | 167 | 0.07 % |
| MX | 210504 | 1098 | 0.52 % |
| RU | 200958 | 2712 | 1.35 % |
| AT | 184248 | 977 | 0.53 % |

Biased by:

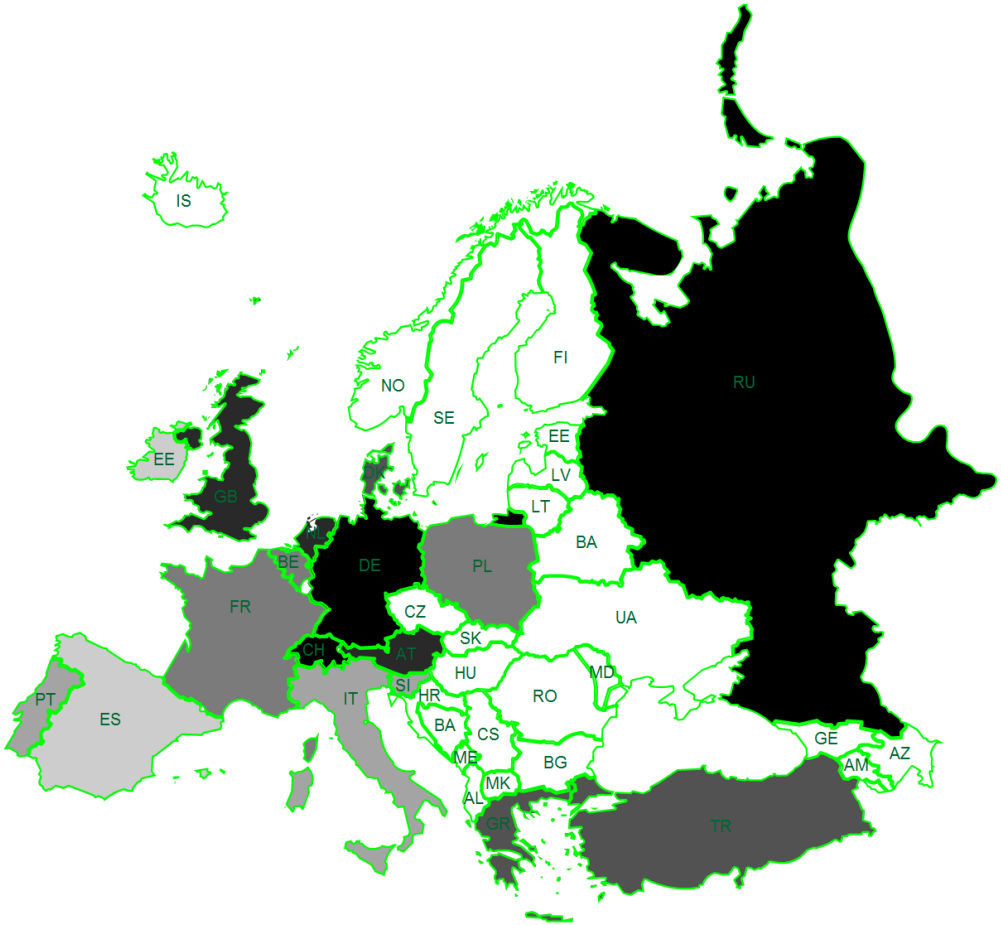
- language knowledge
- decoding problems

# Geo-location: maps



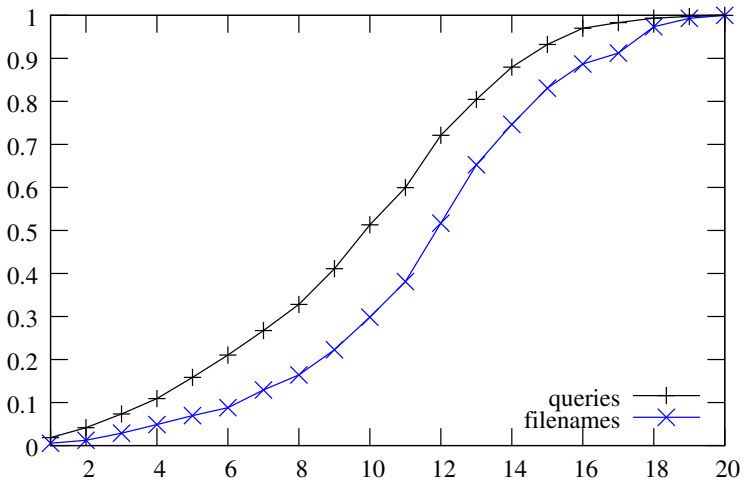

# queries

# Geo-location: maps




ratio # paedophile queries / # queries

# Ages




$x$ : ages *xyo*

$y$ : fraction of occurrences with age $\leq$ $x$

**$\leq$ 10 years old : 50% (queries) et 30% (files)**

**$\leq$ 5 years old : 15% (queries) et 7% (files)**