

Analýza velkých dát a Search retargeting

Analýza velkých dát - nové
koncepty

e FOCUS



Analýza velkých dát a Search retargeting

*e*FOCUS



Analýza veľkých dát - nové koncepty

Mgr. Martin Šeleng, PhD.,
Ústav Informatiky Slovenská Akadémia Vied



*e*FOCUS

Obsah

- Search vs Site retargeting
- Internetová reklama – ako to funguje
- Magnetic
- Problémy
- Riešenia a technológie
- Sumarizácia

Search retargeting

- Search vs Site retargeting
 - **Site retargeting** je zameraný na zdvihnutie povedomia o značke, firme, web stránke, atď. – teda snaží sa používateľov, ktorý už niekedy navštívili stránku/obchod, videli značku priviesť naspäť.
 - **Search retargeting** má za úlohu získať nových potenciálnych zákazníkov aby navštívili stránku/obchod, získali povedomie o značke.

Site retargeting



Používateľ



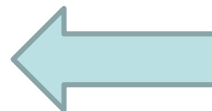
Navštívi
vašu stránku



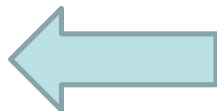
Nenakúpi
a odchádza



Pri prechádzaní
po internete
zbadá reklamu



Spomenie si



Vráti sa a nakúpi

Search retargeting



Používateľ



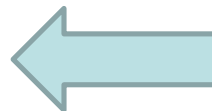
Navštíví
ľubovoľnú stránku
(môže vykonať aj
vyhľadávanie)



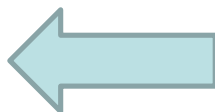
Vaša spoločnosť
získa tieto informácie



Zanalyzujete tieto
informácie (obsah stránok,
vyhľadávané dopyty)

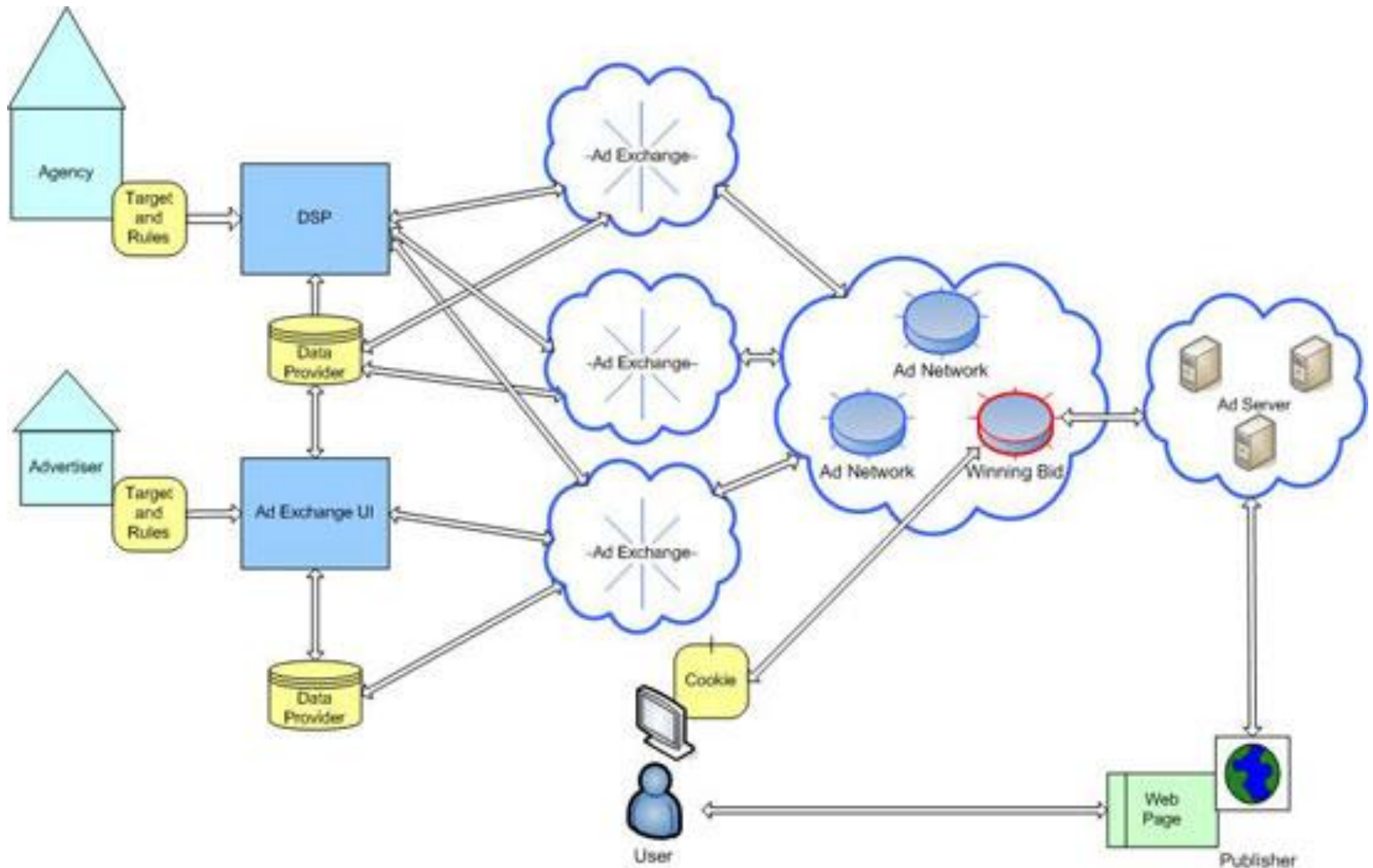


Používateľ navštíví
ľubovoľnú stránku



Zobrazíte mu reklamu na
základe získaných informácií

Internetová reklama



- Spoločnosť založená investormi: **Edison Ventures, Charles River Ventures, NYC Seed, SV Angel, Roger Ehrenberg, Founder Collective, New York City Investment Fund, Ron Conway**, sídliaca v NY, založená v roku 2008



MAGNE+IC™

The diagram illustrates the digital marketing ecosystem, showing the flow from Marketers to Publishers. The flow is: Marketers (left) -> Agencies (top left) -> Agency Trading Desks (top left) -> DSPs (top center) -> Exchanges (top center) -> Ad Networks (top right) -> Sharing Data / Social Tools (top right) -> Publishers (right). The diagram also includes various other categories like Creative Optimization, Media Planning and Attribution, DMPs and Data Aggregators, Targeted Networks / AMPs, Performance, Ad Servers, Verification / Privacy, Measurement and Analytics, Data Suppliers, Mobile, and Media Mgmt Systems and Operations. Each category contains numerous company logos.

Agencies

- Omnicom Group
- WPP
- PUBLICIS
- IPG Interpublic Group
- Havas
- AKGIS
- Merkle
- MDCX PARTNERS
- AKQA
- LBI
- ROSETTA
- epsilon
- icrossing
- Innovation
- INTERACTIVE
- camelet
- atat
- Interactive
- Reach Local
- yodle
- WebVisible

Agency Trading Desks

- Accuen The Trading Desk
- MIC
- XAVIS
- Vivaki
- CADREON
- RUN
- DIGILANT
- acoordantmedia
- IMPACT
- VARIDMEDIA

DSPs

- MediaMath
- invite media
- TURN
- DataXu
- EfficientFrontier
- TRIGGIT
- thetradedesk
- sitescout
- AdBuyer.com
- cd3

Exchanges

- doubleclick by Google
- RIGHTMEDIA
- Facebook Exchange
- ADECN
- OpenX
- adBrite
- adjug

Ad Networks

- Horizontal
- YAHOO!
- Microsoft Media Network
- Aol
- Google
- ValueClick
- Brand
- AT&T
- Adwords
- AdDynamix
- Video / Rich Media
- video
- videology
- YuMe
- BBE
- TubeMogul
- BrightRoll
- spotxchange
- adconion
- gumgum
- BREAK
- meemo
- BoarScout
- Media
- Vertical / Custom
- NetShelter
- martini media
- IGTechNetwork
- ideas people
- Targeted Networks / AMPs
- NetSeer
- peer
- CONTEXTWEB
- pulse360
- media
- degrees
- specific
- precision
- choice
- stream
- resonate
- across
- radius
- XTEND
- social
- vibe
- adaramedia
- rocketfuel
- collective
- interclick
- Performance
- epic
- MATOMY
- VANTAGE MEDIA
- VIBRANT
- adknowledge
- moxy media
- MARCHEX
- LinkShare
- neverblue
- Datran Media
- hydra
- Tatto Media
- TheUseful
- MediaTrust
- Mobile
- admob
- Jumpstart
- transpara
- pontiflex
- inmobi
- quattro
- frhythm
- Media Mgmt Systems and Operations
- DDS
- media
- telmar
- centro
- operative
- theorem
- facilite
- TRAFFIQ
- Solbright
- AdPro

Sharing Data / Social Tools

- facebook
- ShareThis
- Add This
- gigya

SSPs

- Admeld
- rubicon
- PubMatic
- RevNet
- UF
- ALTITUDE

Publisher Tools

- Yieldex
- fatTail
- ShortLift
- isocket
- BuySellAds
- BuzzFeed
- DailyMe
- scout analytics
- Mochila
- daylife
- PerfectMarket
- MAXIFIER
- METAMARKETS
- PARSELY
- Outbrain

Ad Servers

- doubleclick by Google
- atlas
- Mediapro
- mediamind
- pictela
- pointroll
- floattalking
- Oadgear
- adform

Verification / Privacy

- DoubleVerify
- Integral
- adSource
- AdJuster
- ADMP
- The Media Trust
- TRUSTe
- privacychoice
- EVIDON

Measurement and Analytics

- OMNITURE
- unica
- webtrends
- nielsen
- comScore
- MOAT
- ljento
- viz
- aprimo

Data Suppliers

- Experian
- axiom
- datalogix
- TARGUSinfo
- TruSignal
- NETEZZA
- where
- bluecava
- infogroup
- VIZ
- SymphonyIRI Group

Tag Mgmt

- TagMan
- esources
- TEALUM
- RECENT

Retargeting

- dotomi
- criteo
- Tellpart
- Simplifi
- Change
- Fetchback
- buysight
- crosspointmedia
- MAGNE+IC
- ReTargeter
- AdRolls
- myThings
- medFORGE

Creative Optimization

- iteracent
- TUMRI
- adroit
- dapper
- struq
- MIXPO
- snaps
- ipromote
- adacado
- spongecell
- CERTONA
- AdReady
- richrelevance
- ADISE

Media Planning and Attribution

- MarketShare
- COGNATE
- NETWORKED INSIGHTS
- IFACTOR
- pulsepoint
- visual
- convertro
- C3 Metrics
- ThinkVine
- COVARIO
- Korrelate

DMPs and Data Aggregators

- bluekai
- ak
- Demdex
- PARIO
- exelate
- quantcast
- briligi
- bizo
- LOTAME
- JovianDATA
- KRUX
- RedAnt
- LiveRamp
- LEGOLAS
- proximic
- PEER
- adamo

 Denotes acquired company

© LUMA Partners LLC 2013

Problémy

- Kategorizovanie dopytov/používateľov
- Predpovedanie Click-through rate (CTR)
- Predpovedanie konverzií
- Detegovanie fraudov (impresií, klikov)
- Overovanie kľúčových slov pre kampane
- Similarita kľúčových slov (kampane) – **zatiaľ neriešené**
- Optimalizovanie stratégie pri aukciách (Vickrey aukcie) – **zatiaľ neriešené**
- Klastrovanie používateľov – **zatiaľ neriešené**

Kategorizovanie dopytov

- Doc. Michal Laclavík
- **Analýza a kategorizácia veľkého množstva dopytov vyhľadávania pomocou Big Data technológií**

Predpovedanie CTR a konverzií

- Dr. Marek Ciglan
- **Problém:**
 - Z aukcie dostávame ponuky na zobrazenie reklamy
 - Ponuka obsahuje
 - identifikátor používateľa (cookie – rozšíriteľný o informácie o používateľovi ktoré máte k dispozícii),
 - informácie o stránke (url, veľkosť reklamnej plochy pre baner, umiestnenie miesta pre baner,...)
 - informácie o aukcii
 - Ponúk je rádovo niekoľko 100.000/sek
 - Ktoré ponuky kúpiť, aby sme maximalizovali počet kliknutí, alebo počet konverzií v reklamnej kampani?
 - CTR sa štandardne pohybuje okolo 0,1%!!!

Predpovedanie CTR a konverzií

- **Cieľ:** predpovedať aká je pravdepodobnosť kliknutia na reklamu/pravdepodobnosť konverzie ak danú ponuku kúpim?
- **Výzvy:**
 - Predikcia musí byť rýchla - ponúk prichádza veľa (niekoľko 100.000/sek), na každú ktorú chceme kúpiť musíme reagovať do niekoľkých milisekúnd
 - Škálovateľnosť budovania predikčného modelu - Model je budovaný z historických dát; každý deň obsahuje niekoľko 10-tok – 100-viek miliónov záznamov; model budovaný z historických dát za 30 dní

Predpovedanie CTR a konverzií

- **Čo nepomôže**

- Metódy prediktívneho modelovania vyžadujúce celú trénovaciu množinu v pamäti
- Metódy s vyššou (ako lineárnou) časovou náročnosťou

- **Čo funguje**

- Online learning – model sa buduje iteratívne, jeden záznam za čas
- Feature hashing – redukcia dimenzionality kvôli pamäťovým obmedzeniam

- **Vowpal Wabbit:**

https://github.com/JohnLangford/vowpal_wabbit/wiki

Detegovanie fraudov

- **Problém:**

- Ako vieme, že impresiu/zobrazenie reklamy, resp. kliknutie na reklamu vykonal človek a nie stroj?
- Zobrazovatelia (publishers) reklamy sú platení podľa rôznych modelov:
 - CPM (Cost Per Mille) – Inzerenti platia za každých 1000 zobrazení (**impression fraud**)
 - CPC (cost per click) – Inzerenti platia za vykonané kliknutie (**click fraud**)
 - CPA (Cost Per Action/Cost Per Acquisition) – Inzerent zaplatí až keď sa vykoná samotný nákup, resp. sa vyplní formulár, atď. (zodpovednosť prenesená na zobrazovateľov/publishers) (**click fraud**)
 - CPV (cost per view) – Presadzované spoločnosťami Google aTubeMogul najmä pre video obsah – nie len zobrazíť video, ale zaplatiť až po prezretí videa (**impression fraud**)
- 32% online reklamy je platených metódou CPM/CPV, 66% CPC/CPA, a 2% sú hybridné metódy (rok 2012)

Detegovanie fraudov

- **Cieľ:**

- Odhaliť “podozrivé” impresie a kliknutia

- **Výzvy:**

- Riešenie musí byť škálovateľné
 - Musí odhaliť opakujúce sa vzory v správaní
 - Odhaliť hodnoty nachádzajúce sa mimo štandardných hodnôt

Detegovanie fraudov

- **Riešenie:**

- Analýza impresií a kliknutí za posledných 24 hodín
- Analýza vlastností ako: doména, IP, IP3, user cookie, user agent, a ich kombinácie po dvojiciach
- Počet a objem analyzovaných dát za 24 hodín:
 - Impresie: 32 298 385
 - Klinutia: 12 985
 - Objem: 23GB komprimovaných dát
- Použité riešenia: PIG a Hadoop
- Prvá verzia založená na definovaní prahu, čo už je fraud
- Anomálie v zobrazení a v kliknutiach (príliš rýchlo po sebe)

- **Budúcnosť:**

- Impression fraud
- Učiace sa algoritmy

Overovanie kľúčových slov pre kampane

- Ing. Štefan Dlugolinský
- **Problém:**
 - Ako overiť, že množina kľúčových slov pre určitú kampaň je dobre zvolená?
 - Problém do budúcnosti: Ako generovať množinu kľúčových slov pre kampane?

Overovanie kľúčových slov pre kampane

- **Riešenie:**

- **Vstup:**

- 2 zoznamy kľúčových slov (vygenerovaný marketingovým oddelením, vygenerovaný z wikipédie na základe podobnosti konceptov)
 - Vyhľadávané dopyty (za 2 mesiace)
 - Impresie, kliknutia a konverzie (za 2 mesiace)

- **Výstup:**

- Zoznamy potenciálny používateľov na základe dopytov
 - Počet impresií, kliknutí a konverzií na základe dopytov, impresií, kliknutí a konverzií

Overovanie kľúčových slov pre kampane

Zoznam	Potenciálny používateľia	Počet dopytov
1 zoznam	101 348 651	1 248 893 051
2 zoznam	2 910 070	21 106 521
Prienik	2 554 918	19 279 070

Prienik medzi zoznamom kľúčových slov a vyhľadávanými dopytmi

Zoznam	Počet dopytov	Aktuálny používateľia	Impresie	Kliknutia	Konverzie	CTR	Conversion rate
1 zoznam	958 342 242	14 836 808	404 007 810	171 251	622 044	0.000423880	0.001539683
2 zoznam	11 440 714	752 202	21 803 930	4 770	60 915	0.000218768	0.002793762
Prienik	10 367 777	716 677	21 345 983	4 572	60 143	0.000214185	0.002817532

Prienik zoznamov kľúčových slov s vyhľadávanými dopytmi a informáciami o impresiách, kliknutiach a konverziách

Ďakujem za pozornosť



eFOCUS