# Text Classification for RTB

## Defining Targeted in Targeted Digital Advertising

# Text Classification for RTB

# Text Classification for RTB



1) User visits the website

2) Impression information is sent to the ad exchange

3) Information about users, pages, and ad spaces is sent to the bidder

4) A suitable bidding amount is determined based on the information, and bidding takes place

Publisher displaying the advertisement

Ad exchange SSP

**Advertiser A**
**Real estate company ¥70**
Condominium audience targeting

**Advertiser B**
**Game company ¥30**
Targeting men

**Advertiser C**
**Automobile company ¥100**
Retargeting

WIN!

6) The advertisement of the bidding corporation is transmitted

Auction takes place within 0.1 seconds

5) The advertiser with the highest bid acquires the right to transmit their advertisement
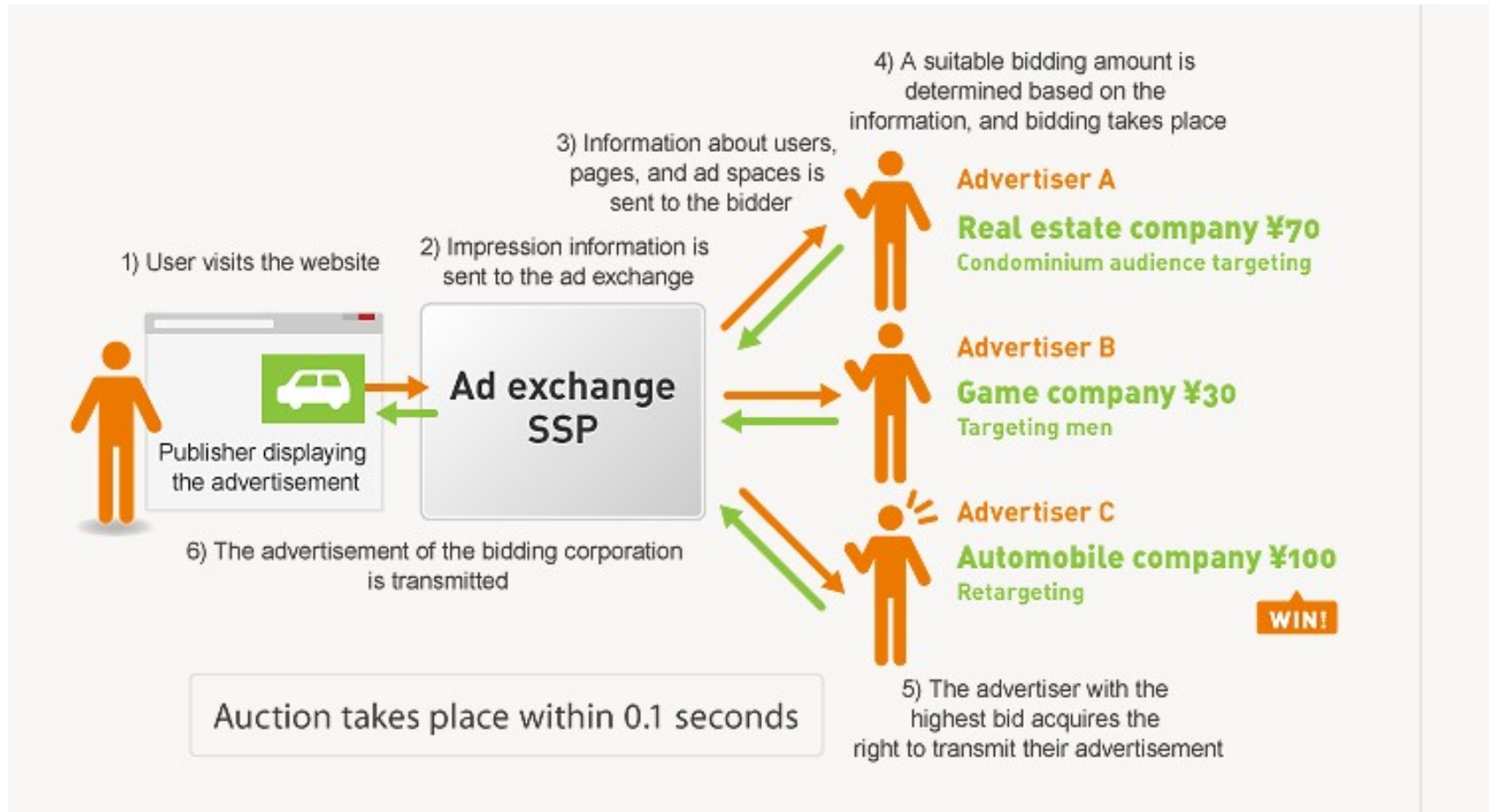
# Text Classification for RTB

- Age
- Gender
- General Interests
- Current Purchase Intent



Massive network
Who is browsing?

# Text Classification for RTB

**Buy Third Party Data**
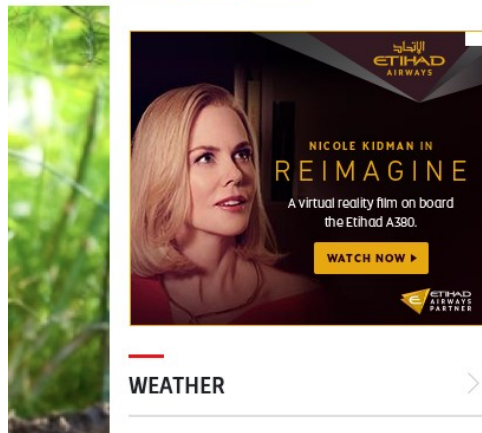




**Develop Custom Analytics**

# Text Classification for RTB

The segments we buy:

- Limited Number of Segments

- Hard to sample from a segment

- Those companies do not have info for each and every user

- It adds cost to advertisement

- How do they know? Whose data to purchase?

6

# Text Classification for RTB

**Consistency within Dataset**

- Is yesterday's male of 55 years old still male and over 55 today?

**Comparison of datasets (from different data providers)**

- Overlaps: how many new users each provider gives? How many users are in common? Do they belong to same segments?

# Text Classification for RTB

Very !!!**EXPENSIVE**!!!





Not only you pay for data, everyone is after the same user at the ad placement auction.

# Text Classification for RTB

**Advertisement is not soccer!!!**

# Text Classification: **Defining Targeted in Targeted Digital Advertising**



**USER A**

1. http://www.bgames.com/shooting-games/

2. http://www.menshealth.com/fitness/best-lower-abs-exercise

3. http://www.bbc.com/news/world-us-canada-33559853

4. http://www.paulsmith.co.uk/uk-en/shop/mens/accessories/ties

**USER B**

1. http://www.ahaparenting.com/parenting-tools/intelligent-creative-child

2. http://www.msmagazine.com
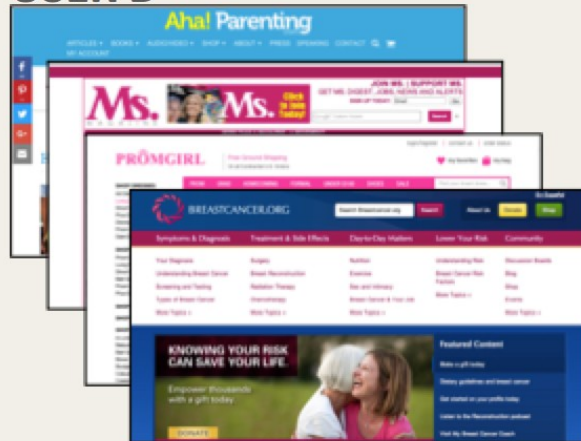
3. http://www.promgirl.com/shop/styles_and_trends/long

4. http://www.breastcancer.org

# Text Classification for RTB

# Text Classification for RTB

- The **text classification** itself is a BigData problem – quality depends on the amount of training data

- **User classification** [after discovering what they read] is definitely a BigData problem. The amount of people who browse on-line is huge. So is the number of logs we have to process!!!

# Text Classification for RTB
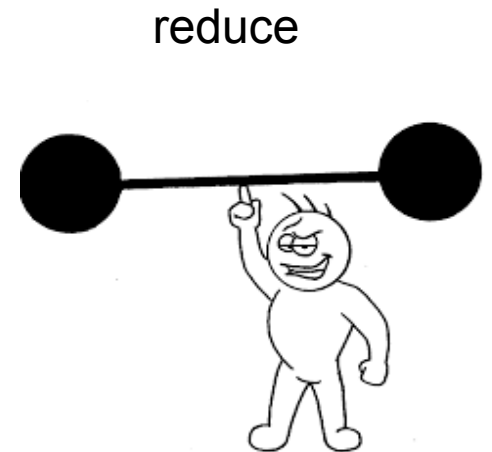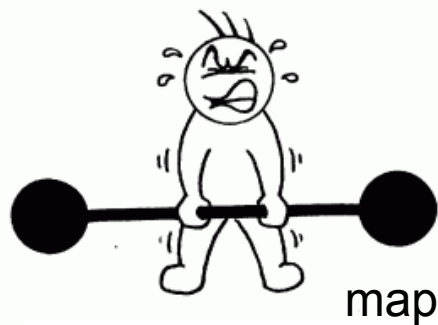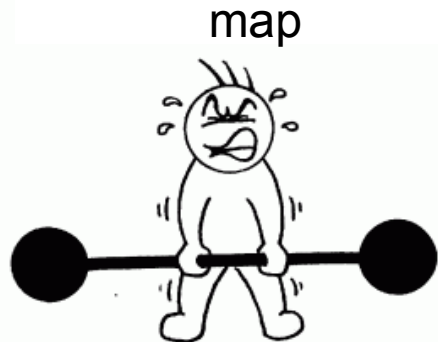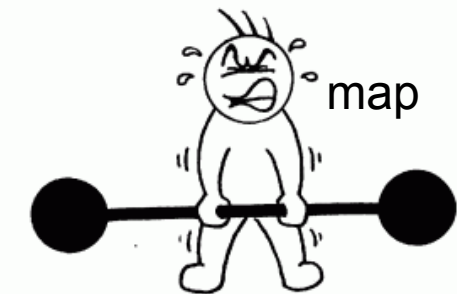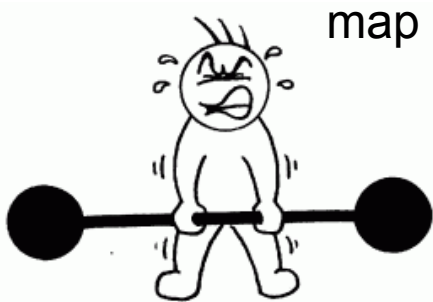
user_id, ts1, url1, ….

user_id, ts2, url2, ….

user_id, ts3, url3, ….

**Url → category**

=>

user_id, {ts1category1}, {ts2 category2}, {ts3 category3}, ...

# Text Classification for RTB

map

map

map

map

map

reduce

# Text Classification for RTB

- Store lots of data

- Look after a cluster of computers

- Support a number of services: Text Classification Service, User Labelling service, Segment Creation, Reporting

# Results

- Ability to clearly define segments
- Ability to meaningfully decrease or increase the size of a segment
- Ability to build custom segments per advertisers's request
- Ability to provide better reporting to customers
- Not spending money on third party data