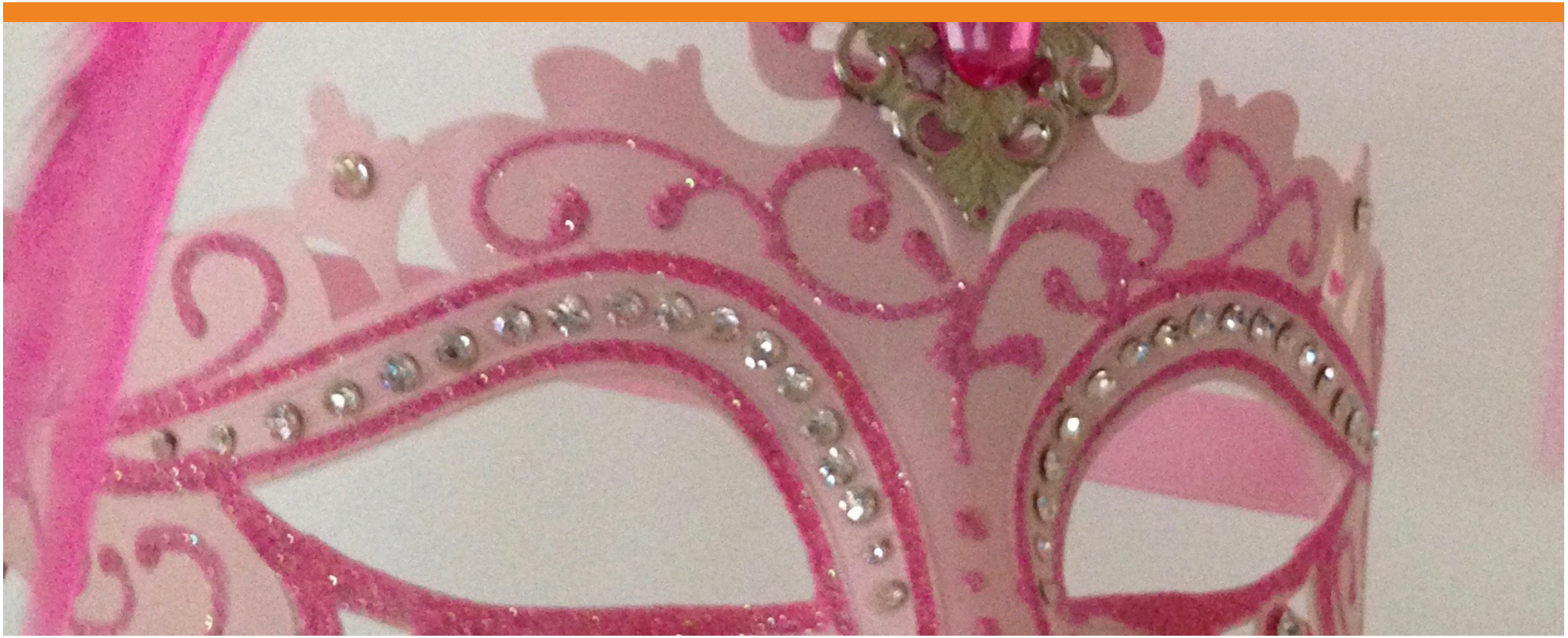


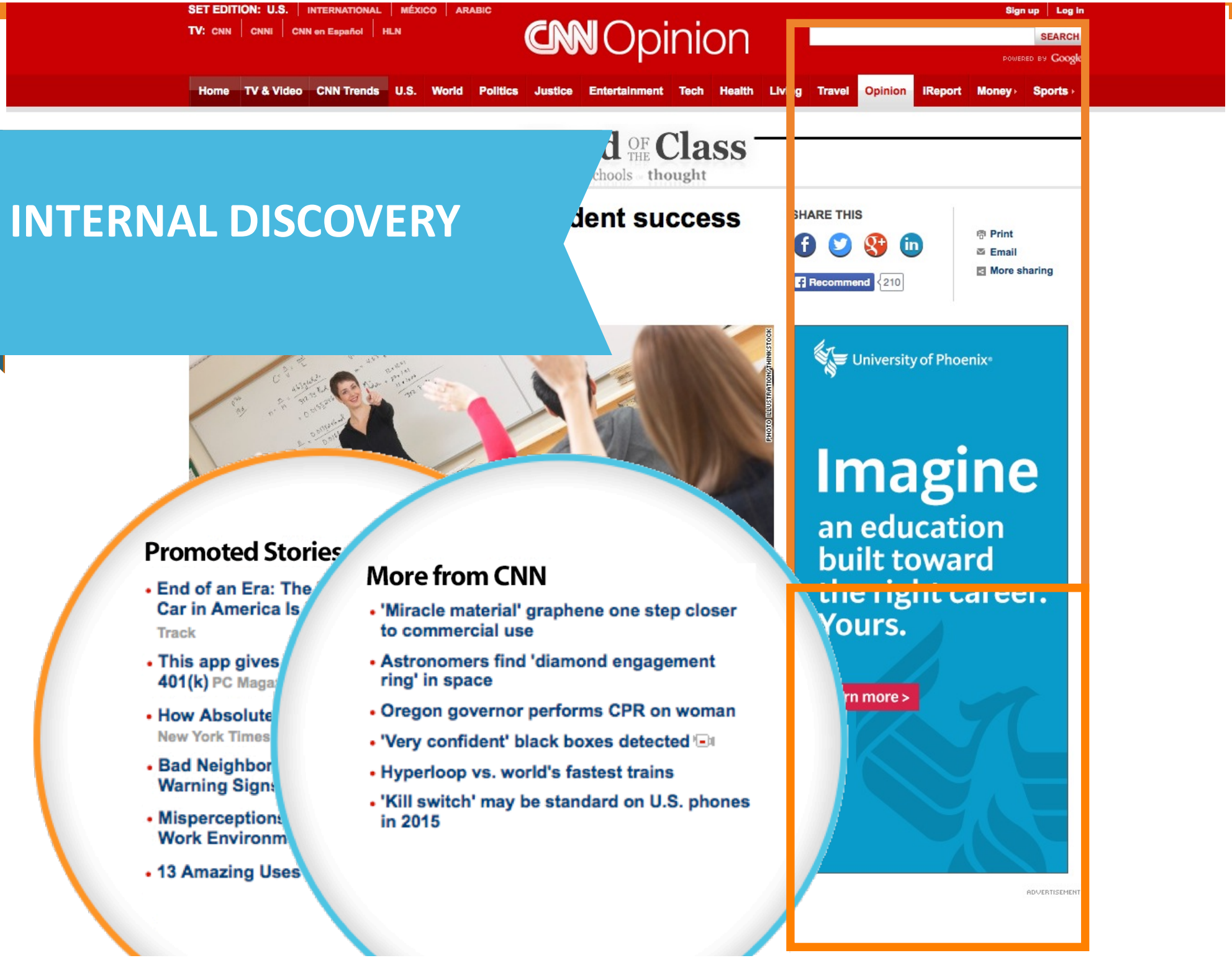
Dynamic Online Profiling of Users' Content Consumption Preferences

Rani Nelken

Director of Research, Outbrain

What Is

INTERNAL DISCOVERY



Distribution Partners

560 MILLION

MONTHLY UNIQUE VISITORS GLOBALLY

Over

25

BILLION

PAGE VIEWS

PER MONTH

Over

200 BILLION

RECOMMENDATIONS

SERVED PER MONTH

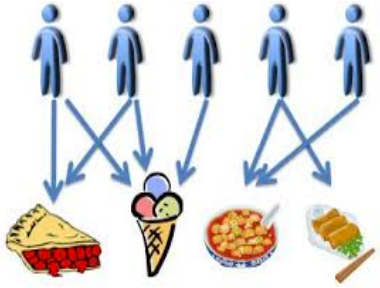

The Lighthouse
Help people discover content they can trust to be interesting, relevant, and timely for them

Personalization

We use 2 approaches:

Content Based

Collaborative Filtering



This talk

User profiles lifecycle

- Extract predictive features from documents
- Construct user profiles incrementally
- Match content with user profile

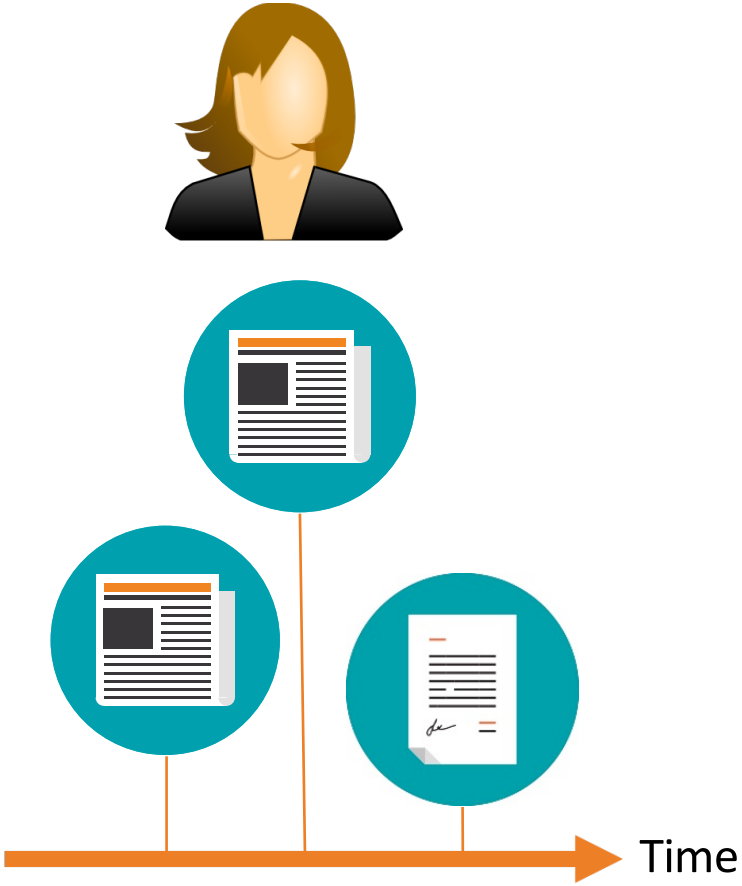
Constructing the profile

Interesting algorithmic questions due to:

- Content understanding challenges
- Online updates
- Dynamic interests
- Scale

outbrain

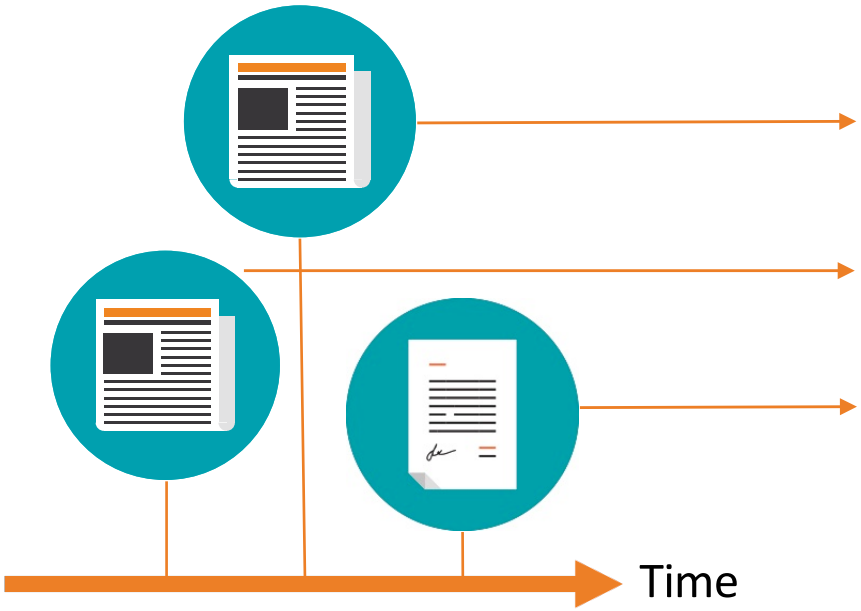
User interacts with documents



Interaction types:

Page Views, Clicks, Shares, Likes, Video play/pause, …

Extract rich features from documents



- Sites
- NLP features: categories, topics, entities
- Cartesian products of features

NLP: What is a document **about**?

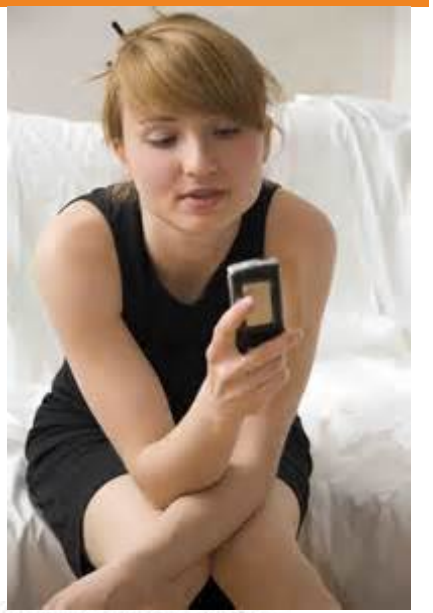
- Classic "Bag of words" approach is too high-dimensional

- We want to encode document meaning in a more compact space

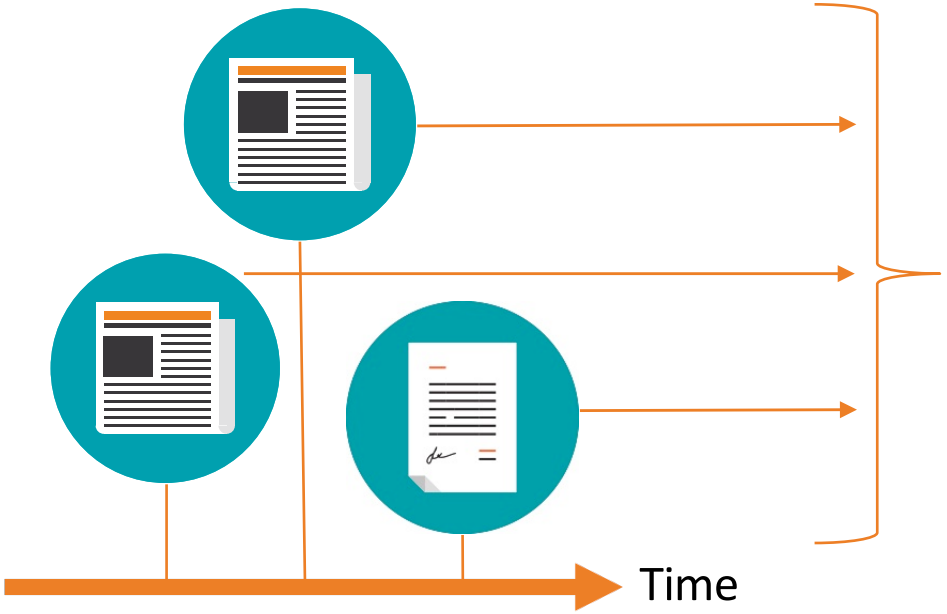
Reduced dimensionality

- Categories
  - *Supervised* classification relative to a fixed set of categories
- Topics
  - *Unsupervised* probabilistic model
  - topic = distribution over words
  - doc = distribution over topics
- Most *relevant* Named entities (person, organization, location)

Illustration: topics



Aggregate features into a profile



User profile

Learning to Aggregate: requirements

- Incremental
- Dynamic
- Low space-complexity

Learning to Aggregate: solutions

- Streaming algorithms:

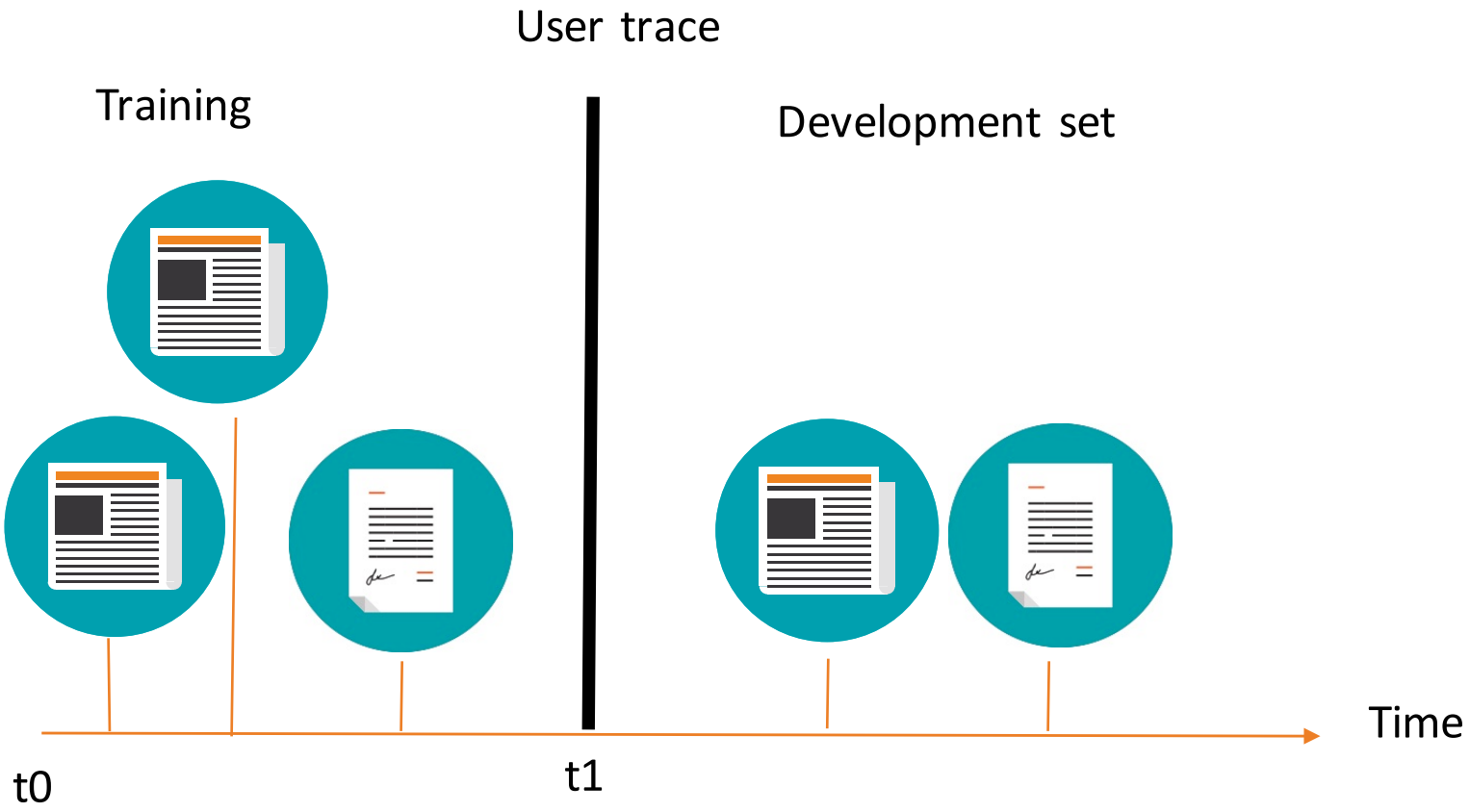
"Lossy" counting over streams: count items but *selectively* forget/decay

- Sketching algorithms:

Compact hashing schemes with partial overlap and redundancy

outbrain

Offline Evaluation



Incrementally build profile

outbrain

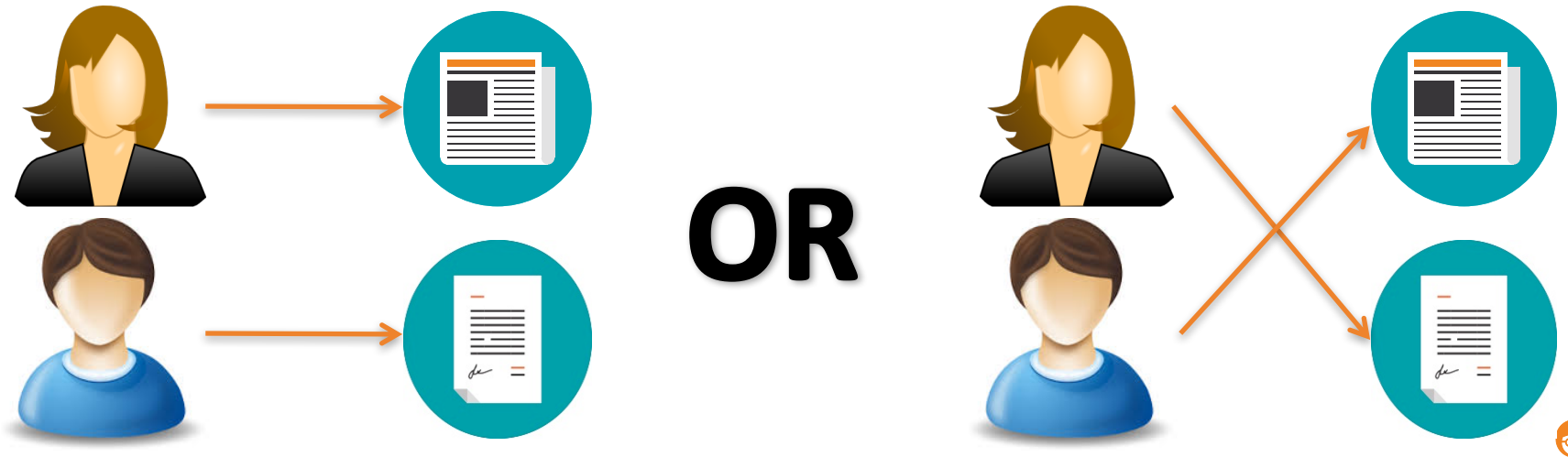
Offline evaluation: difficulties

- Positive interactions are observed but negative examples are not
- Heavy bias due to publishers' editorial staff

outbrain

Offline prediction: Solution

- Pick a pair of users and a pair of docs from their development set
- Does the model prefer actual pairing or counterfactual one?



Conclusion

- Content based profiling is extremely powerful
- Interesting algorithmic challenges in building dynamic online profiles
- Interesting software architecture challenges for a scalable implementation

outbrain

Thank you

rani at outbrain dot com,
@RaniNelken

We're hiring
www.outbrain.com/about/careers