Social Media-Driven News Personalization

O'Banion, Birnbaum, Hammond
Knight News Innovation Lab
Northwestern University

Goal

- News and media content recommendation based on an analysis of users' social media activity.

- Contributions:
  1. User-modeling technique based on analysis of Twitter content.
  2. News story recommendation system providing stories from The Huffington Post.
  3. Unique evaluation technique using existing Twitter data as a basis for comparison.

Twitter as an Information Source

- Popularity
  - 100+ million active users, 200+ million tweets per day [1]
- Unique practices promote **frequent and concise** posts
- Continuous and real-time information stream

[1] http://blog.twitter.com/2011/09/one-hundred-million-voices.html

Approach

- Identify named entities in text.

(Indiana Pacers, Organization, Sports)

"Indiana Pacers guard Paul George throws down a 360-degree, through-the-legs dunk"

(Paul George, Person, Sports)

Approach

- Build vertically aligned lexicons.
  - E.g., cooking terms (ingredients, equipment, actions)

"Recipe: Vegan White Cheese Party Dip bit.ly/U93r3s"

"Futures dip on manufacturing data: U.S. stock market futures are falling after the Commerce Department reported … bit.ly/OrorOY"

- Problems:
  - Ambiguity
  - Not always "newsworthy"

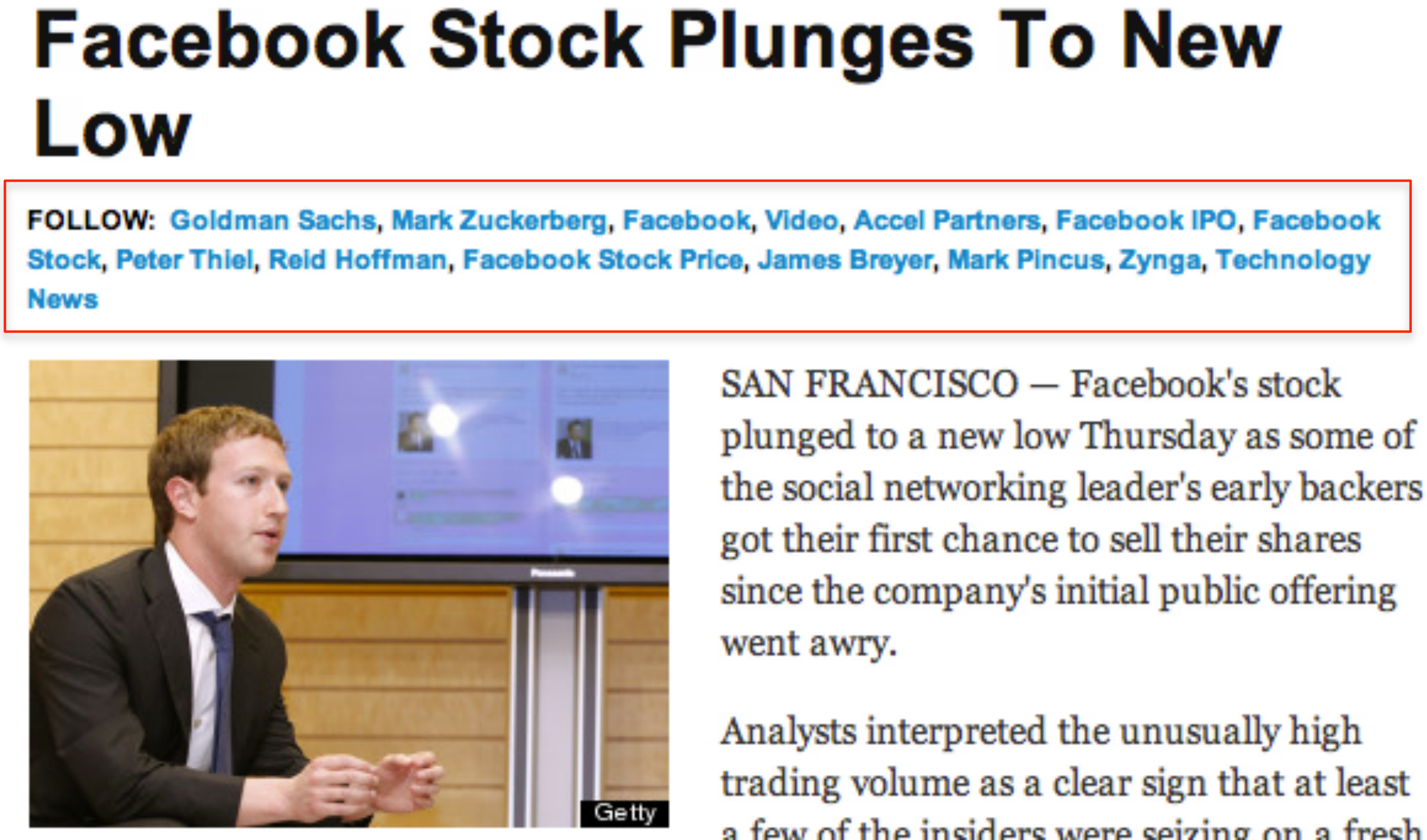
Approach

- Identify **categories** and **tags** in tweets.
  - Derived from the The Huffington Post’s taxonomy.

Categories	**Tags**
Arts	*People, Organizations,*
Books	*Topics, etc.*
Business	
College	Chicago Bulls
Education	Mitt Romney
Entertainment	Entrepreneurship
Food	Music
Green	Beer
…	…

Dataset

- The Huffington Post news story corpus.



Facebook Stock Plunges To New Low

FOLLOW: Goldman Sachs, Mark Zuckerberg, Facebook, Video, Accel Partners, Facebook IPO, Facebook Stock, Peter Thiel, Reid Hoffman, Facebook Stock Price, James Breyer, Mark Pincus, Zynga, Technology News

SAN FRANCISCO — Facebook's stock plunged to a new low Thursday as some of the social networking leader's early backers got their first chance to sell their shares since the company's initial public offering went awry.

Analysts interpreted the unusually high trading volume as a clear sign that at least a few of the insiders were seizing on a fresh

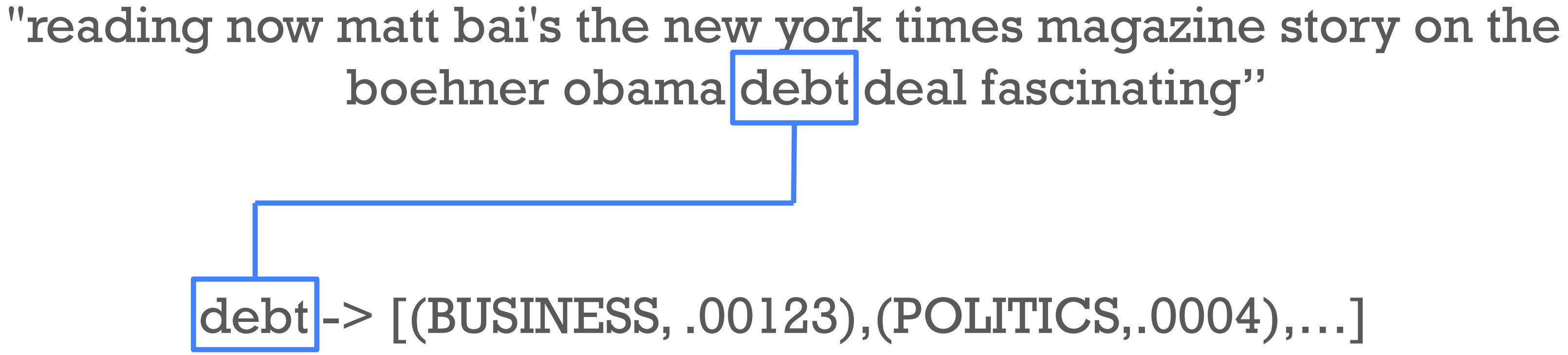
Classifiers

- Centroid-based classification algorithm[1].
- A category/tag is represented as a term vector with tfidf weights.
- Inverted index maps distinct terms to weighted categories for quick lookup.

[1] E.-H. S. Han and G. Karypis. Centroid-based document classification: Analysis and experimental results.

Classifiers

"Reading now: Matt Bai's @NYTimes magazine story on the Boehner/ Obama debt deal. Fascinating."



"reading now matt bai's the new york times magazine story on the boehner obama debt deal fascinating”

debt -> [(BUSINESS, .00123),(POLITICS,.0004),…]

Classifiers

"Reading now: Matt Bai's @NYTimes magazine story on the Boehner/ Obama debt deal. Fascinating."

Category: **Business**

Tags: **John Boehner, Debt Ceiling, Barack Obama, Debt Limit**

User Profiling

A profile for user *u* consists of categories *c* and tags *t*:

$$P(u) = \{(c, f(c)) \mid c \in C\}, \{(t, f(t), maxdate(t)) \mid t \in T\}$$

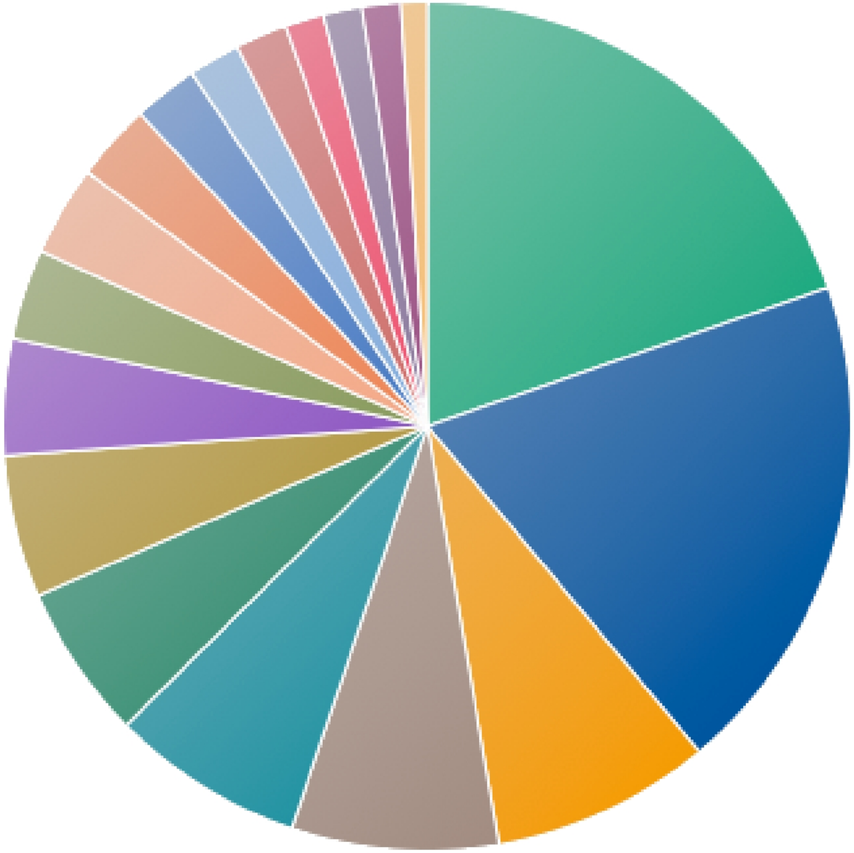
Each tag *t* is assigned a score:

$$TS(u,t) = f(t) * .9^{d}$$

Each category *c* is assigned a score:

$$CS(u,c) = \sum_{t \in TC} TS(u,t)$$

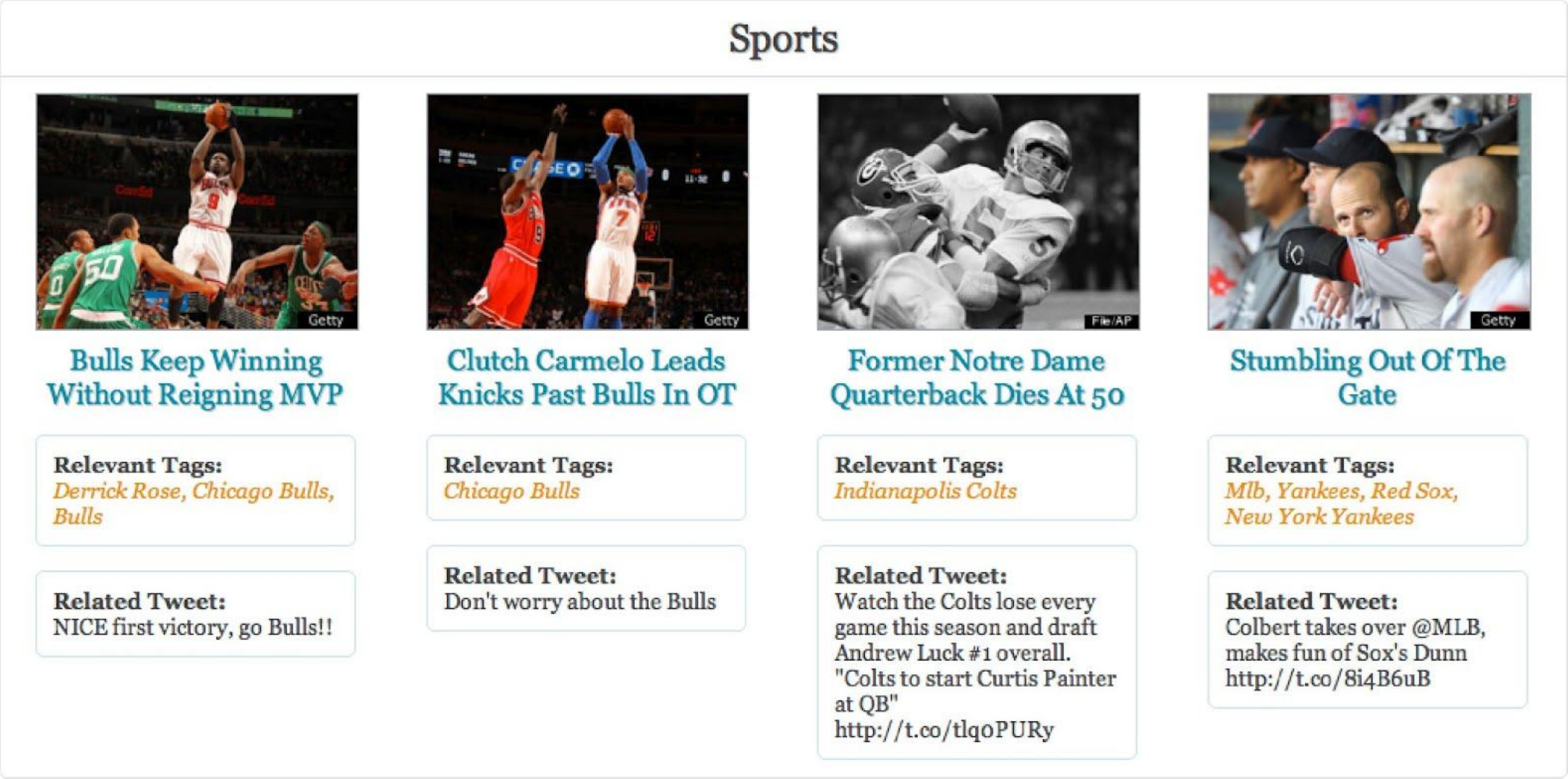
User Profiling



Food	40	World	7
Sports	39	Politics	6
Entertainment	17	Books	5
College	16	Arts	4
Technology	15	Travel	4
Education	12	Style	3
Healthy living	11	Business	3
Media	9	Religion	3
Green	7	Impact	2

Presentation

A story s with tags $t < T$ is assigned a score for user u: $\sum_{t \in T} TS(u,t)$



Sports

Bulls Keep Winning Without Reigning MVP

Relevant Tags:
Derrick Rose, Chicago Bulls, Bulls

Related Tweet:
NICE first victory, go Bulls!!

Clutch Carmelo Leads Knicks Past Bulls In OT

Relevant Tags:
Chicago Bulls

Related Tweet:
Don't worry about the Bulls

Former Notre Dame Quarterback Dies At 50

Relevant Tags:
Indianapolis Colts

Related Tweet:
Watch the Colts lose every game this season and draft Andrew Luck #1 overall. "Colts to start Curtis Painter at QB" http://t.co/tlq0PURy

Stumbling Out Of The Gate

Relevant Tags:
Mlb, Yankees, Red Sox, New York Yankees

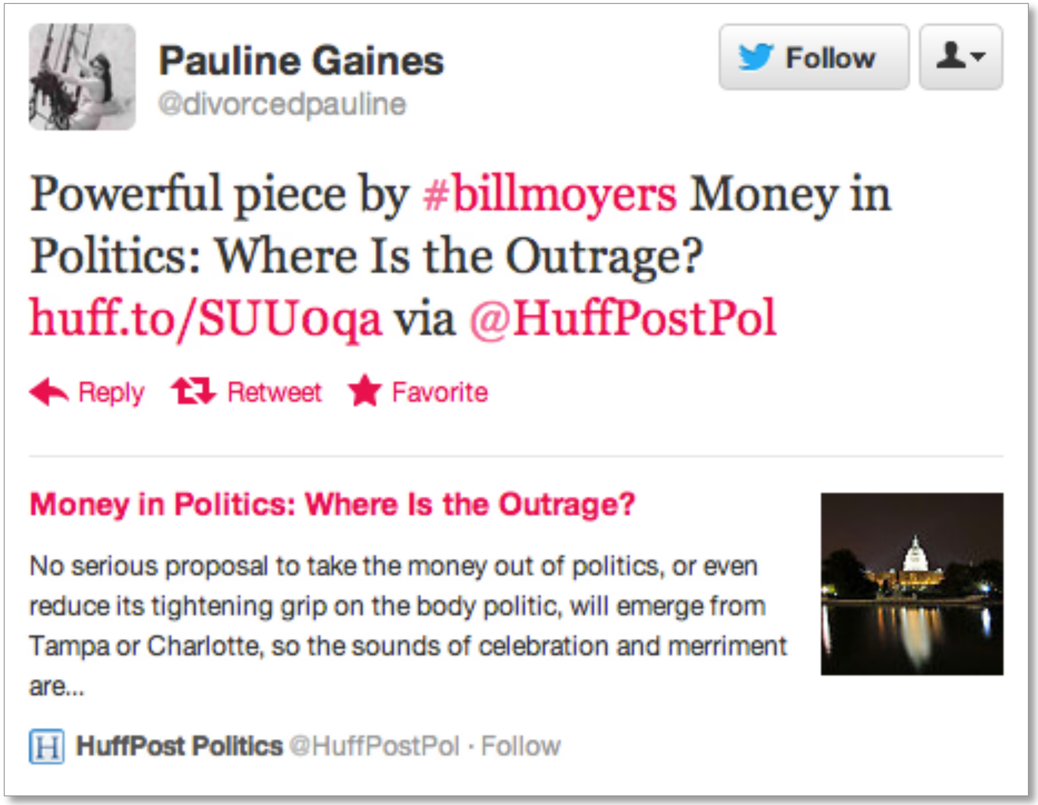
Related Tweet:
Colbert takes over @MLB, makes fun of Sox's Dunn http://t.co/8i4B6uB

Evaluation

- Gathering data on user preferences is difficult.
- Field Study
  - Requires a deployed system with significant user-base.
- Controlled Experiment
  - Potentially time consuming and costly.

Evaluation

- Our solution: Mine existing data for explicit indications of interest.
- For example:



Pauline Gaines
@divorcedpauline

Powerful piece by #billmoyers Money in Politics: Where Is the Outrage? huff.to/SUUoqa via @HuffPostPol

Money in Politics: Where Is the Outrage?

No serious proposal to take the money out of politics, or even reduce its tightening grip on the body politic, will emerge from Tampa or Charlotte, so the sounds of celebration and merriment are...

HuffPost Politics @HuffPostPol · Follow

Twitter Dataset

Characteristic	**Value**
Users	1,000
Total user tweets	1,082,441
Avg. tweets per user	1,082
Avg. valid tweets per user	774
Distinct stories	729
Story tweet date range	1/12/12 - 2/11/12
Avg. tags per story	11.69
Categories represented	18

Evaluation

1. Identify explicit indications of interest.

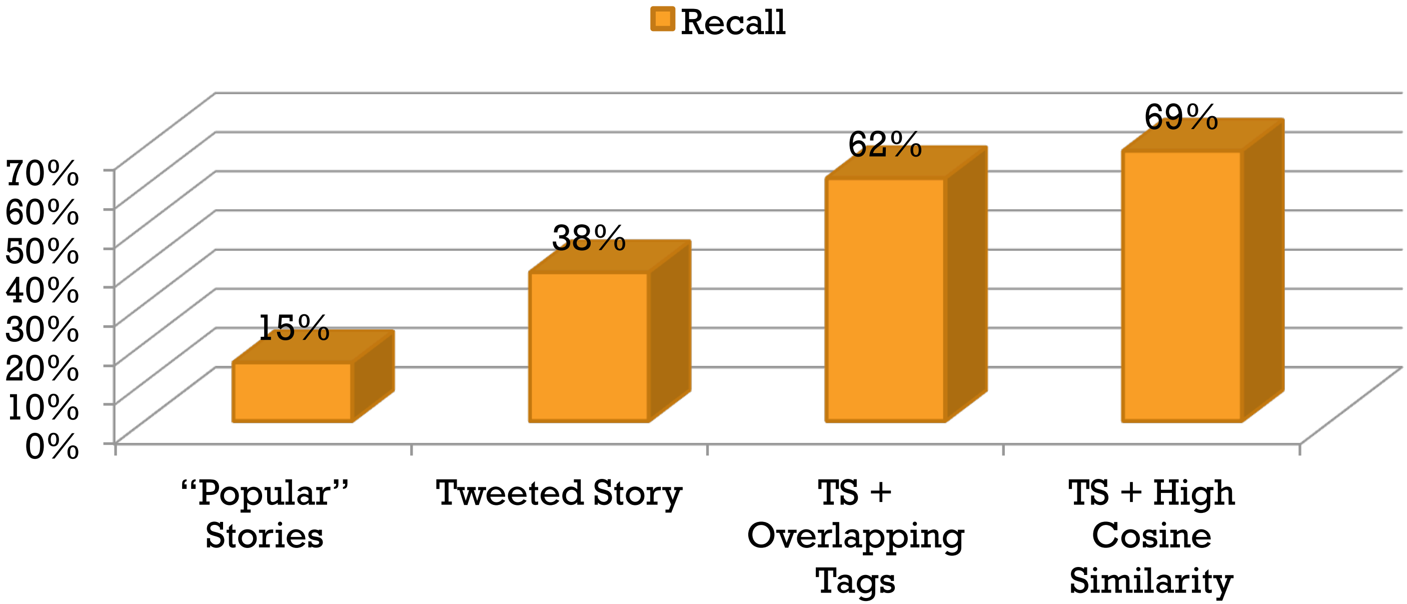
2. Perform retroactive analysis.

3. Answer the question:
    - Would our system have recommended this story or a *similar* one to this user?

- Similar stories:
    - Overlapping Tags
    - High Cosine Similarity (represented as term vectors with tfidf weighting)

Results

- Recall: Fraction of instances where a correct story was recommended
- Baseline: Stories featured on the "most popular" section of the Huffington Post.



Acknowledgements

- I’m grateful for the support of:
  - National Science Foundation
  - John S. and James L. Knight Foundation
  - Cognitive Science Fellowship from Northwestern University

- Website:
  - http://twxray.knightlabprojects.com/

Related Work

- User Modeling based on Social Media
  - F. Abel, Q. Gao, G.-J. Houben, and K. Tao. Analyzing user modeling on twitter for personalized news recommendation.
  - O. Phelan, K. McCarthy, and B. Smyth. Using twitter to recommend real-time topical news.
  - J. Chen, R. Nairn, L. Nelson, M. Bernstein, and E. Chi. Short and tweet: experiments on recommending content from information streams.
- News Media Content Recommender Systems
  - J. Liu, P. Dolan, and E. R. Pedersen. Personalized news recommendation based on click behavior.
  - A. S. Das, M. Datar, A. Garg, and S. Rajaram. Google news personalization: scalable online collaborative filtering.