Human Centered Data Science

DATA 512 — Jonathan T. Morgan

Human centered algorithm design | Week 8 | November 15, 2018

Overview of the day

- Final project plan refresher
- Guest Lecture: Kelly Franznick, BlinkUX
- Reading reflections
- Design process and organizational structures for HCAD
- Evaluating algorithms for bias
- Evaluating algorithms for user experience
- In-class activity: Ethical OS Risk Assessment

Homework due next week (NO CLASS SESSION)

Reading reflection

- Hill, B. M., Dailey, D., Guy, R. T., Lewis, B., Matsuzaki, M., & Morgan, J. T. (2017). Democratizing Data Science: The Community Data Science Workshops and Classes

Final Project plan

- Due Week 9 (November 22)
- 10 points
- Min. 1000 words
- Jupyter Notebook or .md file on GitHub, link submitted to Canvas

Final project plan

- **Why are you planning to do this analysis?** Provide background information about the topic, research questions/hypotheses, (imagined) business goals, and anything else that will be required to properly contextualize your study.
- **What is your plan?** Describe *and link to* the data sources will you collect, how data will be collected and processed, the analysis you intend to perform, and the outcomes and deliverables you anticipate.
- Are there any unknowns or dependencies that might affect your ability to complete this project?
- How do human-centered design considerations inform...
    a. your decision to pursue this project
    b. your approach to performing the work?

Final project plan

What type of data and analysis?

- Must use publicly-available and appropriately licensed dataset(s)
- Can be a ‘classic’ statistical analysis, or the design and/or evaluation of a machine learning model
- Use your own definition of ‘big data’
- Choose datasets and analyses that are likely to support reproducibility
- Choose datasets and methods that let you answer questions that you find interesting and important
- Visualizations aren’t necessary, but encouraged as an effective way of communicating your findings

Final project plan - open data

You can only use a dataset for your project if the license or terms of use allow you to collect the data, analyze it, and re-publish it publicly.

- Some licenses and terms of use specifically prohibit that.
- Some TOU say it's okay for non-commercial purposes (like academic research).
- Some data sources don't specify a license *or* terms of use for their data (hint: avoid these).

Possible exceptions (explicit permission needed)

  - if your project is an audit-style analysis of an existing algorithm
  - If you can publish a persistent identifier for each datapoint (e.g. Tweet ID)

How to document your data

When your dataset has an explicit license

1. State the license of your data (e.g. “CC-By-SA 4.0”) in your report.
2. When possible, link to the license deed, e.g. https://creativecommons.org/licenses/by-sa/4.0/

When data re-use is covered under the provider’s Terms of Use

1. Quote the relevant section of the terms of use in your report
2. Link to the terms of use page

If possible, link to the *original source* of the data, which may be different from where you found it.

- E.g. MovieLens data on the GroupLens website vs. MovieLens data on Kaggle

Be careful with Kaggle data

Many of those datasets are not explicitly licensed. If you cannot find appropriate license information for the data, you cannot use it. You'll fail the assignment, even if the rest of your work is really good. :/

Many of those datasets have already been analyzed by other Kagglers. Many of those analyses are public on Kaggle.com.

- Your analysis should not simply duplicate analysis that Kagglers have already done on this dataset (e.g. "do more data scientists use Python or R?").
- It's perfectly fine to build off of the analysis that others have done, just make sure you cite the original analysis.
- Tip: Avoid even *looking like you might be* plagiarizing someone else's analysis.

Talking about gender

Many projects are centred on gender as a variable. Since the assignment requires you to think about ethics, consider what ‘gender’ is:

- Not the same as sex (‘man’, ‘woman’ not ‘male’, ‘female’)
- Not a binary (people have identities other than ‘man’ or ‘woman’)
- Not impermeable (people transition between genders)
- Something that creates myriad different experiences of life that are likely to be reflected in your data.

Talking about gender

Questions to ask:

- Does your data only contain binary options? If so, note that as a limitation - and note why it is a limitation (it excludes people outside that binary).
- Does your data include trans people? If so, incorporate that into the gender research. If not, highlight *that* as a limitation.

Final project plan

- Due Week 9 (November 22)
- 10 points
- Min. 1000 words
- Jupyter Notebook or .md file on GitHub, link submitted to Canvas

Examples

- Erin Orbits
  - Plan
  - Presentation
- Rex Thompson
  - Plan
  - Presentation

Guest lecture

Kelly Franznick

CIO @ BlinkUX

Reading reflections

Reading reflections

If we were to start from a pre-trained learner and further adapt it to a new environment, how should we handle pre-existing biases the learner adopted while running in the previous environment? Does the development of interfaces that allow users to query the learner with specific examples to probe for known biases resolve this?

-Edmund

Reading reflections

I'm concerned that the use of interactive ML systems may lead to greater bias in models, induced by end users either consciously on unconsciously and/or to a better understanding of how to seem transparent but really hide a model's biases and limitations. How would we go about making sure that research in interactive ML systems is extended to cover such "attacks"? Or the larger question that is: how would a white-hat interactive ML hacker approach the evaluation of these systems to make them secure and fair? - *Javier*

This is certainly a step in the right direction, but how do you weed out the bad actors, in this case, people who have malicious intents and who might feed the model wrong inputs? Should there be a screening process instituted to prevent this from happening? - *Tejas H.*

Reading reflections

With growing concerns over privacy and people in favor of concealing their identity, Iterative Machine learning approach leads to building very customized models much closer to its end user, how is this approach takes care of privacy concerns? Is it not potentially more intruding and thus can be exploited more conveniently than a generalized Applied machine learning model?

-Purshottam

Reading reflections

One thing I would be concerned about as a data scientist after reading this paper is that in some cases the authors are suggesting that users have control over intricate parts of models such as classifier decision boundaries, etc. As the authors state, these things are complex even for experts, and it seems allowing laypersons who don't have a firm background in machine learning to have control over these aspects could lead to degradation of the model or even inappropriate applications…

The authors state that interfaces for interacting with machine learning should be evaluated properly to ensure this doesn't happen, which would lead me to ask, what constitutes appropriate evaluation for a novel interface before deployment, and who should be making these evaluations?

- Kenton

Reading reflections

We know that algorithms can not be designed to be perfect after all they are designed by humans. We also know that in most cases the problems can only be discovered only after real deployment. Considering this, don’t you think we should allow room for mistakes by algorithms if they are unexpected and unintentional?

- Tejas J.

Human centered algorithm design

Human centered systems

- Based on an analysis of human tasks
- Built to take account of human skills
- Designed to address human needs
- Monitored in terms of human benefits

Source: Rob Kling & Leigh Star, 1997 “Human Centered Systems in the Perspective of Organizational and Social Informatics (in week 1 ‘Resources’ on wiki)

Human Centered Design

- The design is based upon an explicit understanding of users, tasks and environments.
- Users are involved throughout design and development.
- The design is driven and refined by user-centred evaluation.
- The process is iterative.
- The design addresses the whole user experience.
- The design team includes multidisciplinary skills and perspectives.

Source: International Organization for Standardization:
https://www.iso.org/standard/52075.html

Today’s focus

- Human centered approaches to avoiding/identifying/addressing *harmful bias*
- Human centered approaches to understanding/improving *user experience*

Assessing and avoiding bias: methods and trade-offs

Three entry points for bias

Source: https://piret.gitlab.io/fatrec2018/program/fatrec2018-garciagathright-slide.pdf

Different types of outcome biases

metrics + outcomes

"[W]e use the term bias to refer to computer systems that systematically and unfairly discriminate against certain individuals or groups of individuals in favor of others."

Friedman & Nissenbaum (1996)

Friedman, B. and Nissenbaum, H., 1996. Bias in computer systems. *ACM Transactions on Information Systems (TOIS)*, *14*(3), pp.330-347.

Harms of allocation withhold opportunity or resources from certain groups

Harms of representation reinforce subordination along the lines of identity / stereotypes

Kate Crawford, "The Trouble With Bias" keynote at NIPS 2017

https://www.youtube.com/watch?v=fMym_BKWQzk

Source: https://piret.gitlab.io/fatrec2018/program/fatrec2018-garciagathright-slide.pdf

Allocation bias examples (speech recognition)

English dialects

Play you da baddest

Source: https://piret.gitlab.io/fatrec2018/program/fatrec2018-garciagathright-slide.pdf

Non-english & code switching

Play Dile Que Tu Me Quieres

Source: https://piret.gitlab.io/fatrec2018/program/fatrec2018-garciagathright-slide.pdf

Understanding fairness

Challenges in assessing ‘fairness’.

Some content gaps & biases are **intentional:**

- New music playlists: recency bias

Some content gaps & biases can be argued to be **unfair:**

- Under-index of certain genres over others

Source: https://piret.gitlab.io/fatrec2018/program/fatrec2018-garciagathright-slide.pdf

Different stakeholders can have different perspectives on ‘fairness’

Decision-maker: of those I’ve labeled high-risk, how many will recidivate?

Predictive value

Defendant: what’s the probability I’ll be incorrectly classified high-risk?

False positive rate

Society [think hiring rather than criminal justice]: is the selected set demographically balanced?

Demography

	Labeled low-risk	Labeled high-risk
Did not recidivate	TN	FP
Recidivated	FN	TP

Arvind Narayanan: 21 fairness definitions and their politics
https://www.youtube.com/channel/UCO19zyFNtkbcTQwERVVZB0Q

Source: https://piret.gitlab.io/fatrec2018/program/fatrec2018-garciagathright-slide.pdf

Understanding fairness

- ‘Fair’ is not an inherent quality of a system or an algorithm.
- Fairness is normative: based on values and social expectations
- Fairness is situated: based on audience, purpose, and context

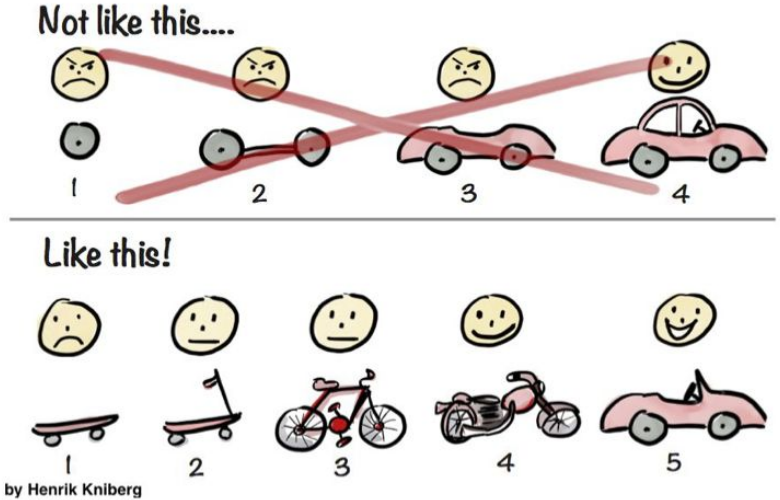
Designing for fairness

- Building algorithms, and algorithmically-powered systems, is a *design* activity.
- Design involves making choices
- Making choices involves value judgements: option A is better than option B *because [value judgement]*

Qualitative and process-oriented approaches for anticipating & avoiding bias

Minimum viable products and continuous improvement

- What would an MVP for algorithmic bias assessment and mitigation look like?



Source: https://piret.gitlab.io/fatrec2018/program/fatrec2018-garciagathright-slide.pdf

Questions and checklists

1. Will this algorithm influence—or serve as the basis of—decisions with the potential to negatively impact people's lives?
2. Can the available data actually lead to a good outcome?
3. Is the algorithm fair?
4. How will the results (really) be used by humans?
5. Will people affected by these decisions have any influence over the system?

Williams and Gunn, 2018. *Math Can't Solve Everything: Questions We Need To Be Asking Before Deciding an Algorithm is the Answer. Electronic Frontier Foundation.*

Better documentation

Anatomy of a data statement

1. *Curation rationale*
2. *Language variety*
3. *Speaker demographics*
4. *Annotator demographics*
5. *Speech situation*
6. *Text characteristics*
7. *Recording quality*

"Drawing on value sensitive design, this paper contributes one new professional practice— called data statements—which we argue will bring about improvements in engineering and scientific outcomes while also enabling more ethically responsive NLP technology.

A data statement is a characterization of a dataset which provides context to allow developers and users to better understand how experimental results might generalize, how software might be appropriately deployed, and what biases might be reflected in systems built on the software."

Bender and Friedman, 2018. *Data Statements for NLP: Toward Mitigating System Bias and Enabling Better Science.*

Stakeholder interviews

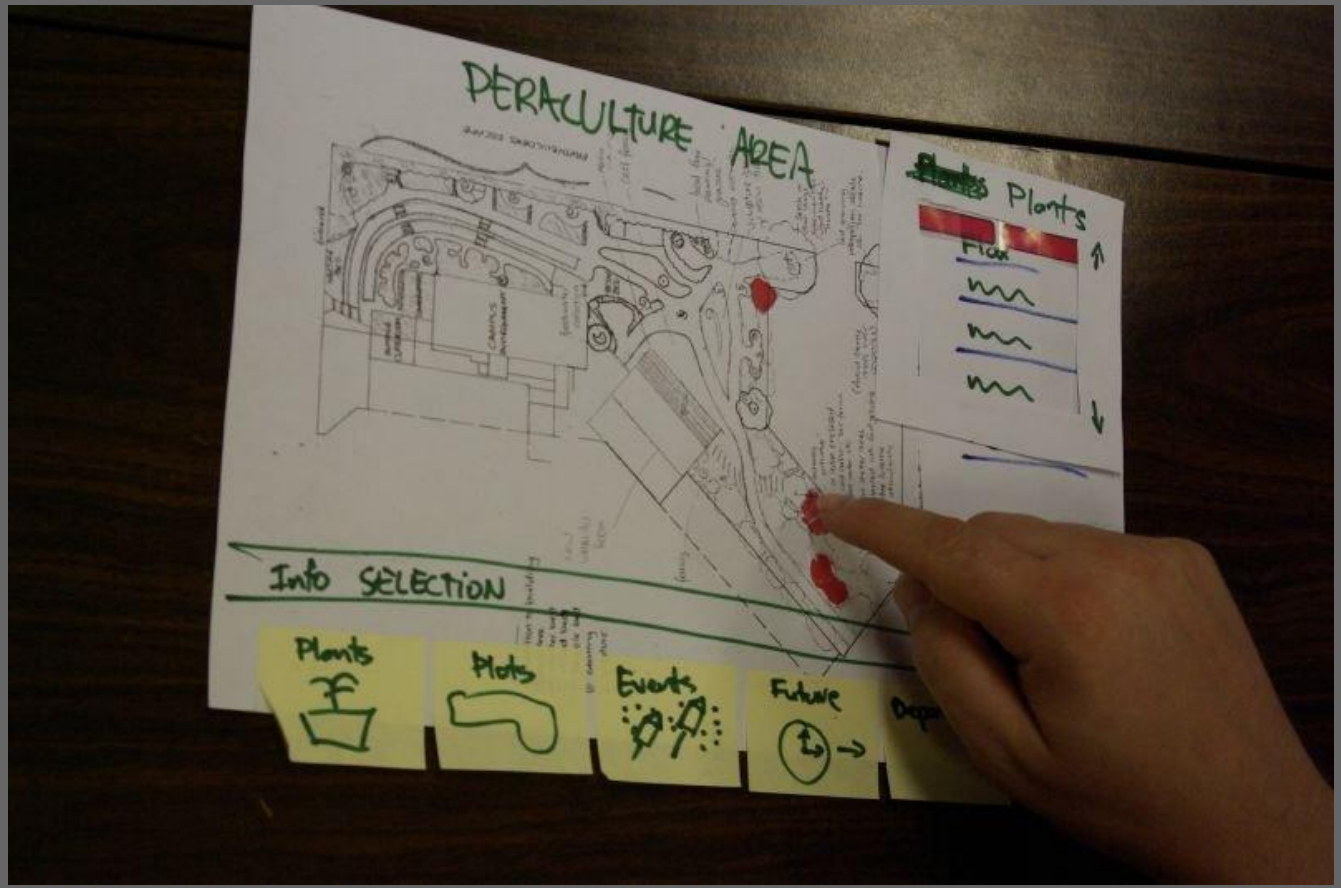
Scenarios and personas

Scenario: a short, detailed story that describes how a particular kind of user might interact with a system. Sort of like an Agile ‘user story’, but with significantly more context. Includes relevant information about the user themself, their motivations and goals, and a step-by-step description of their hypothetical usage of the system.

Persona: a detailed profile of a (fictitious) system user, including extensive information about their background and motivations, as well as relevant demographic info and their relationship (if any) to the system.

Ved & Morgan, 2018. GLAM Personas for Structured Data on Commons
Morgan, 2012. Wikipedia Teahouse - new user scenarios

Low fidelity prototypes



User studies

Organizational and policy considerations

Organization, policy, and processes

What can orgs and teams do to avoid unintended consequences?

- Develop and use checklists
- Create a data ‘red team’
- Implement a dissent channel
- Implement post mortems and share lessons learned
- Create case studies
- In interviews, ask an ethics question

Organization, policy, and processes

What can individual contributors do to encourage organizations to be good citizens?

- Demand diversity
- Being interviewed? Ask how the company handles ethical issues and do they have a dissent channel

Organization, policy, and processes

“Unanticipated or undetected biases should be further reduced by including members of diverse social groups in both the planning and evaluation of AI systems and integrating community outreach into the evaluation process. Behavioral scientists and members of the target populations will be particularly valuable when devising criterion tasks for system evaluation. Such tasks would assess, for example, whether the [system] applies norms in discriminatory ways to different races, ethnicities, genders, ages, body shapes, or to people who use wheelchairs or prosthetics, and so on.”

IEEE *Ethically-Aligned Design*. 2018.

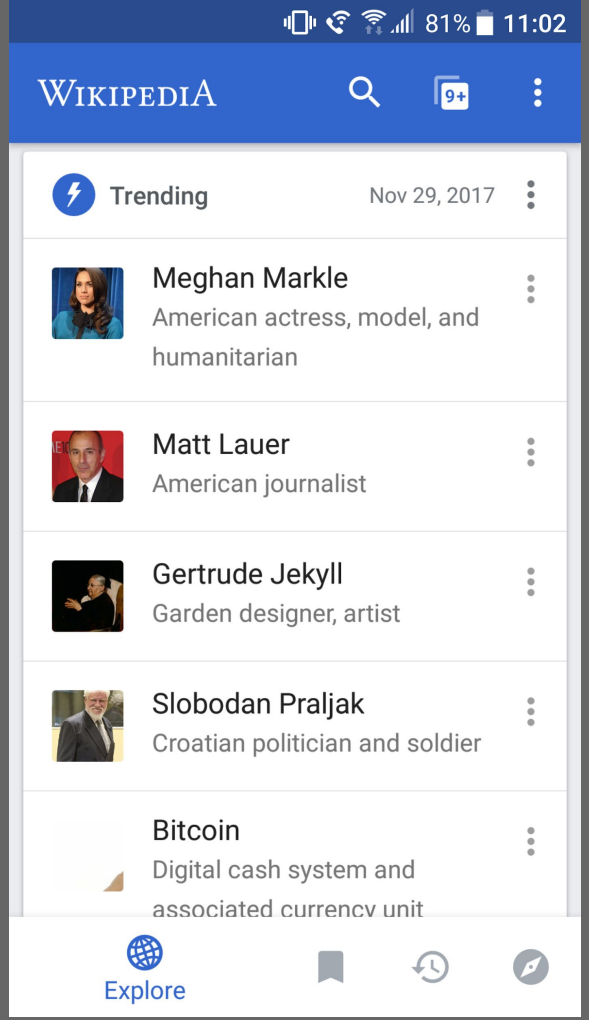
Avoiding bad outcomes - societal measures

“Every student working with data needs to be trained in ethics & security as part of the core curriculum”

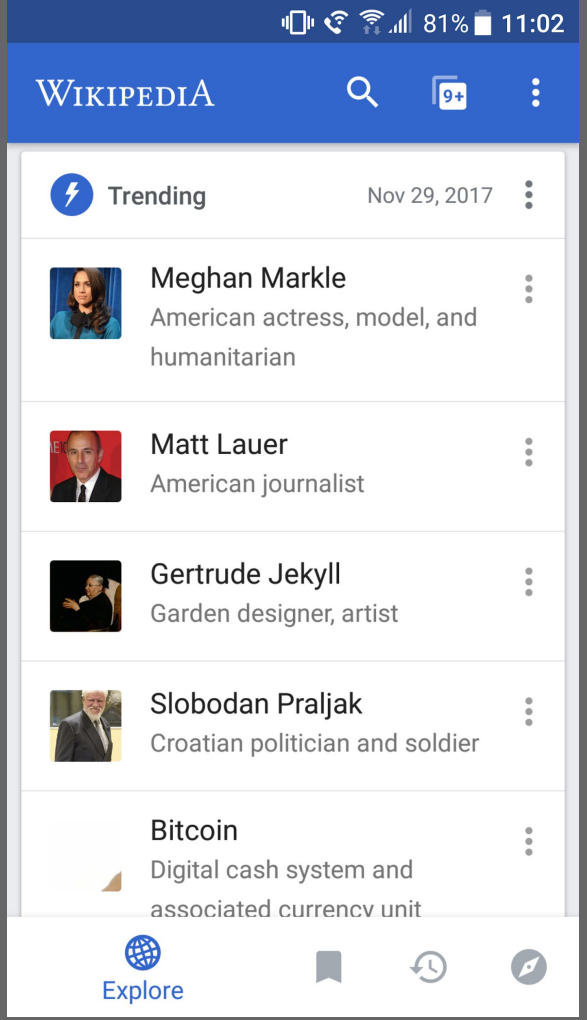
Partnership on AI Annual meeting - November 14, 2018

Evaluating for bias

Case study: Evaluating ‘top articles’



https://meta.wikimedia.org/wiki/Research:Comparing_most_read_and_trending_edits_for_Top_Articles_feature

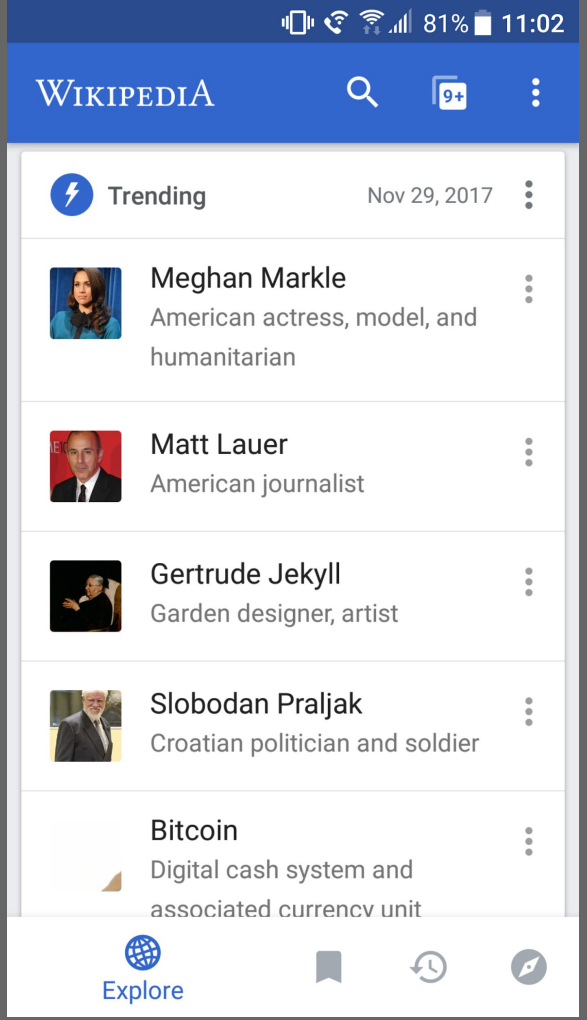


One feed, two algorithms

1. **Top read:** articles with most page views as of ~24 hours ago
2. **Trending edits:** articles that have experienced an unusual ‘bump’ in edits/editors (near real-time)

Which algorithm works best? What do we mean by best? Who does it work best *for*?

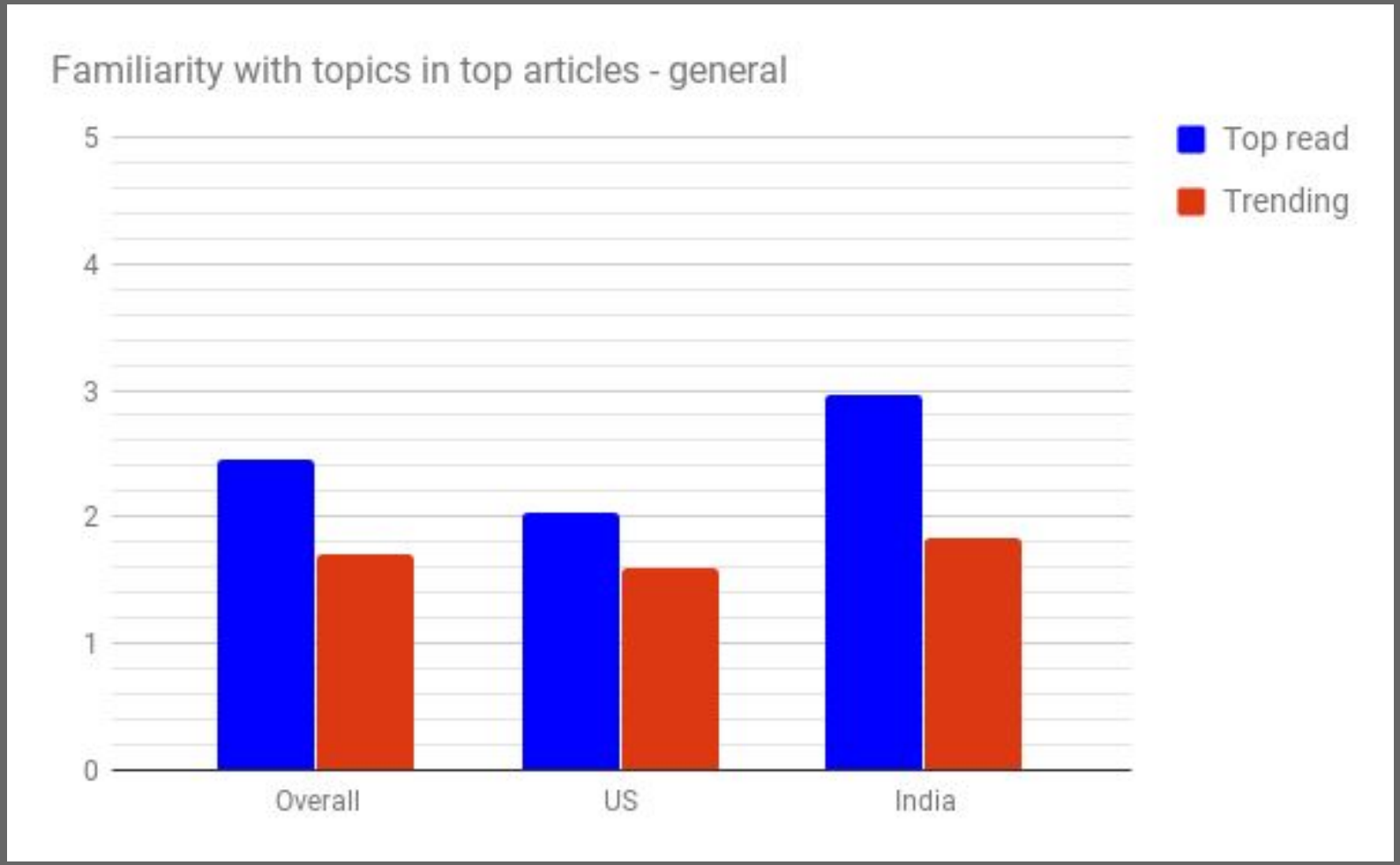
https://meta.wikimedia.org/wiki/Research:Comparing_most_read_and_trending_edits_for_Top_Articles_feature

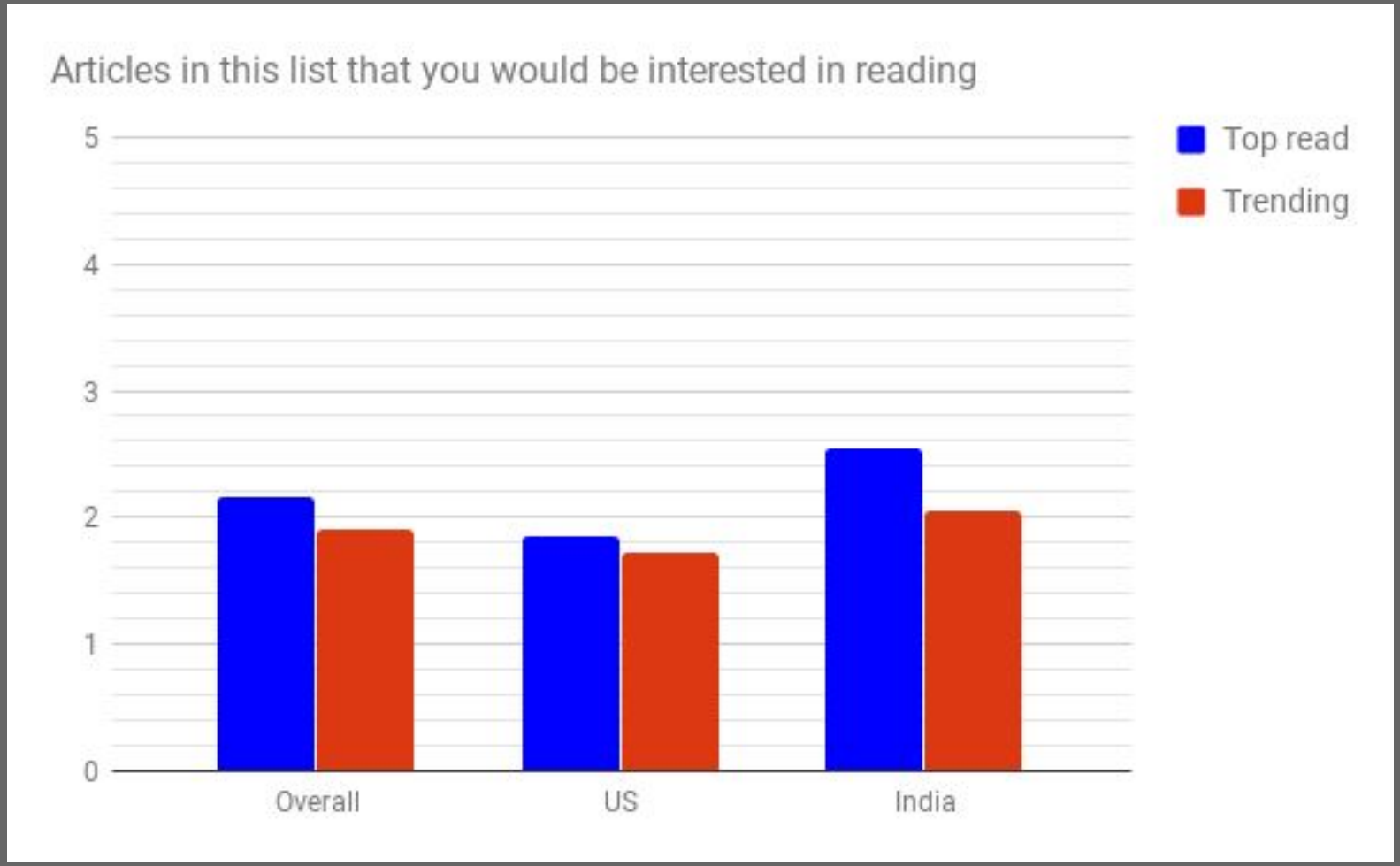


Rating questions for Turkers

1. How many articles in this list are *clearly related* to topics that you are familiar with?
2. How many articles in this list are *clearly related* to topics that you have seen or read about on other websites (not Wikipedia) within the past 24 hours?
3. How many articles in this list would you be interested in reading right now?
4. If there was a list of trending articles *like the ones on this list* on the home screen of a Wikipedia app for mobile devices, *how often* would you use it to look for new articles to read?
5. Why would you (choice from question #4) use a list that contained articles like these to find new articles to read?

https://meta.wikimedia.org/wiki/Research:Comparing_most_read_and_trending_edits_for_Top_Articles_feature


Familiarity with topics in top articles - general
5
4
3
2
1
0
Overall
US
India
Top read
Trending


Articles in this list that you would be interested in reading
5
4
3
2
1
0
Overall
US
India
Top read
Trending

Evaluating for user experience

Human centered algorithm design case study

Recommender systems

Recommender systems

- One of the most familiar types of AI (thanks, Amazon and Facebook)
- Many different sources of signal to build from
  - E.g. user behavior, user demographics, item characteristics
- Many established AI-based approaches
  - E.g. Collaborative filtering, content filtering, naive Bayes, LSI
- ‘dumb’ approaches are also ubiquitous and can be effective
  - E.g. raw item popularity, recency
- Which approach to choose?

Approaches to rec sys design

Engineering-centric

- use whatever sources of data are easiest to access and process
- build whatever models provide the best performance and offline accuracy

Business-centric

- use whatever sources of data we have
- build whatever model provides the best ROI, per pre-established metrics

Human-centric

- use sources of data that are ethically appropriate
- Allow users to understand, and potentially control, model behavior
- build models that take into account user characteristics, needs/goals, and context of use
- Evaluate in terms of metrics that are meaningful to end users

Human-recommender interaction model

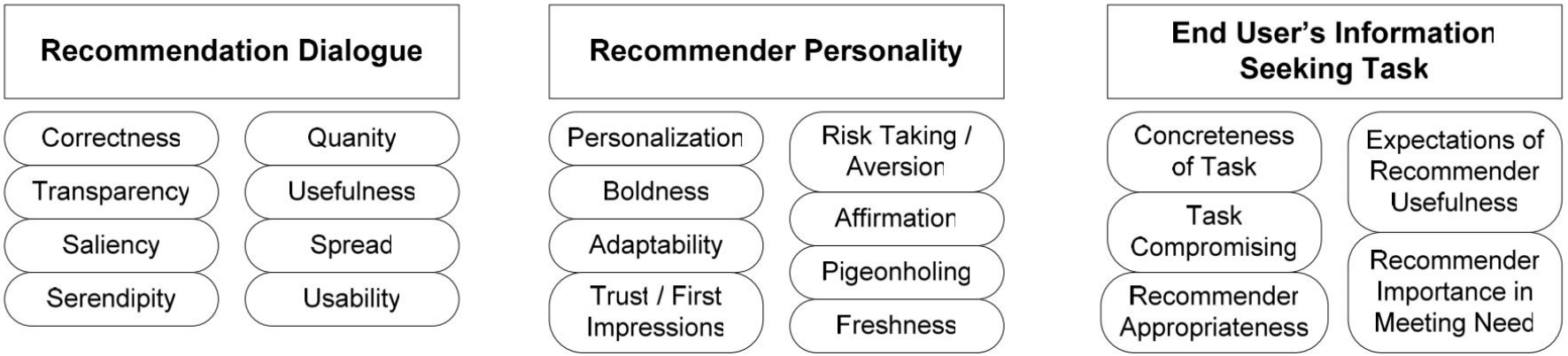


Figure 1-1: Aspects of Human-Recommender Interaction. The Aspects are divided into three 'Pillars'

Sean M. McNee, John Riedl, and Joseph A. Konstan. *Making recommendations better: an analytic model for human-recommender interaction. CHI EA 2006.*

How people evaluate recommendations

Aspects of the recommender dialogue

Correctness: user judges the rec to be high quality

Usefulness: the rec helps the user with their task

Transparency: user understands why they received this rec

Salience: the rec stands out, generates an emotional response (pos or neg!)

Serendipity: the rec is unexpected, in a good way

Spread: the rec list represents the domain well (completeness/recall)

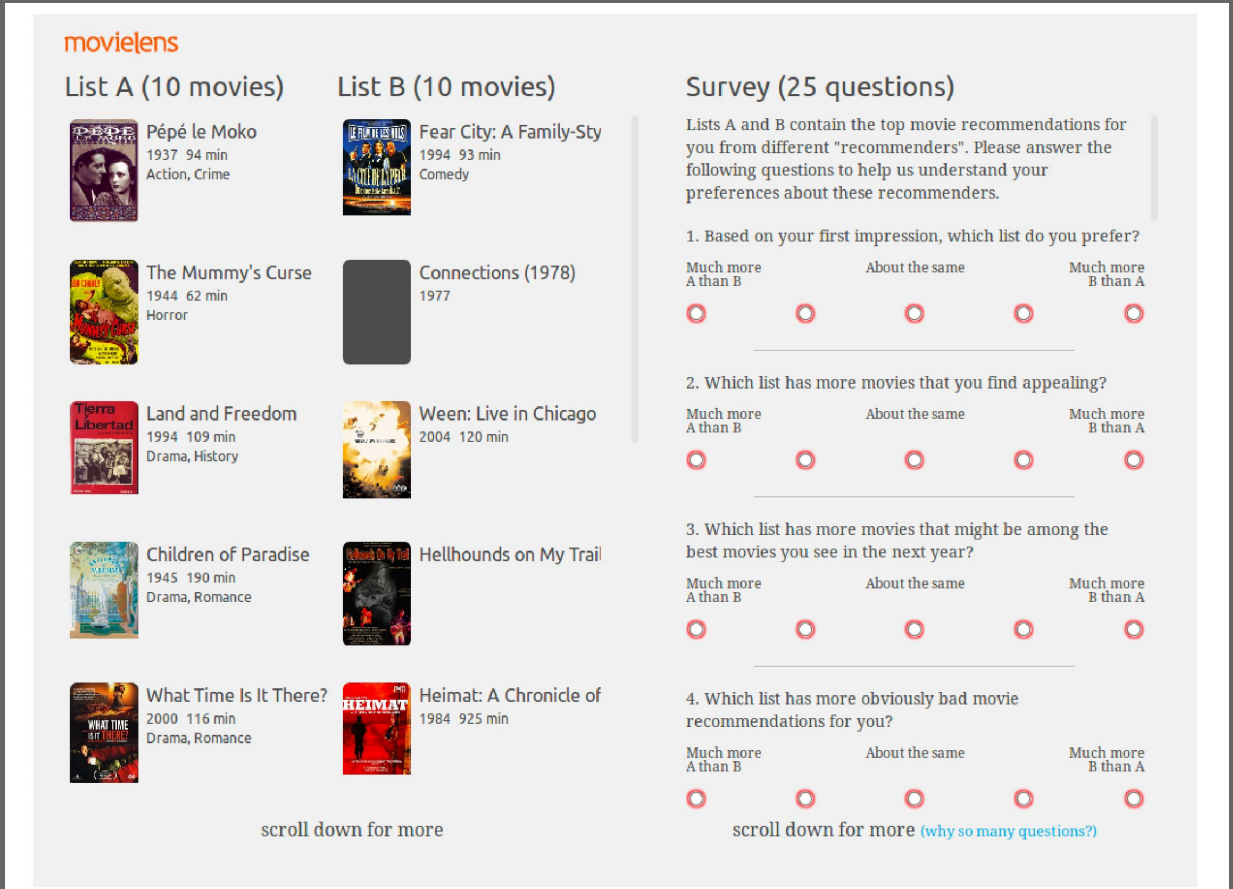
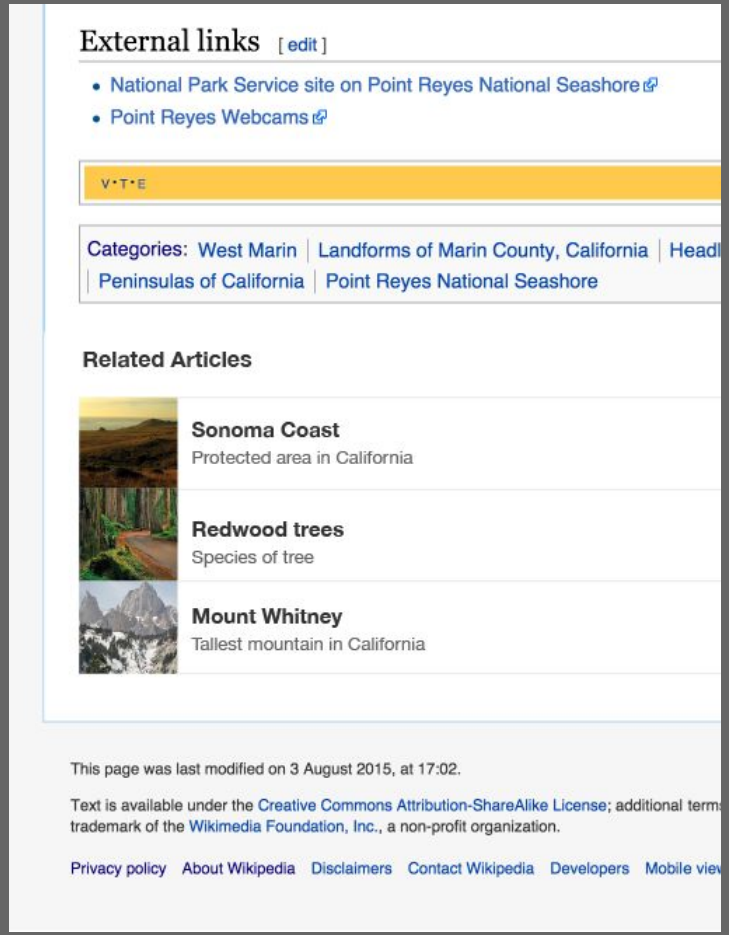


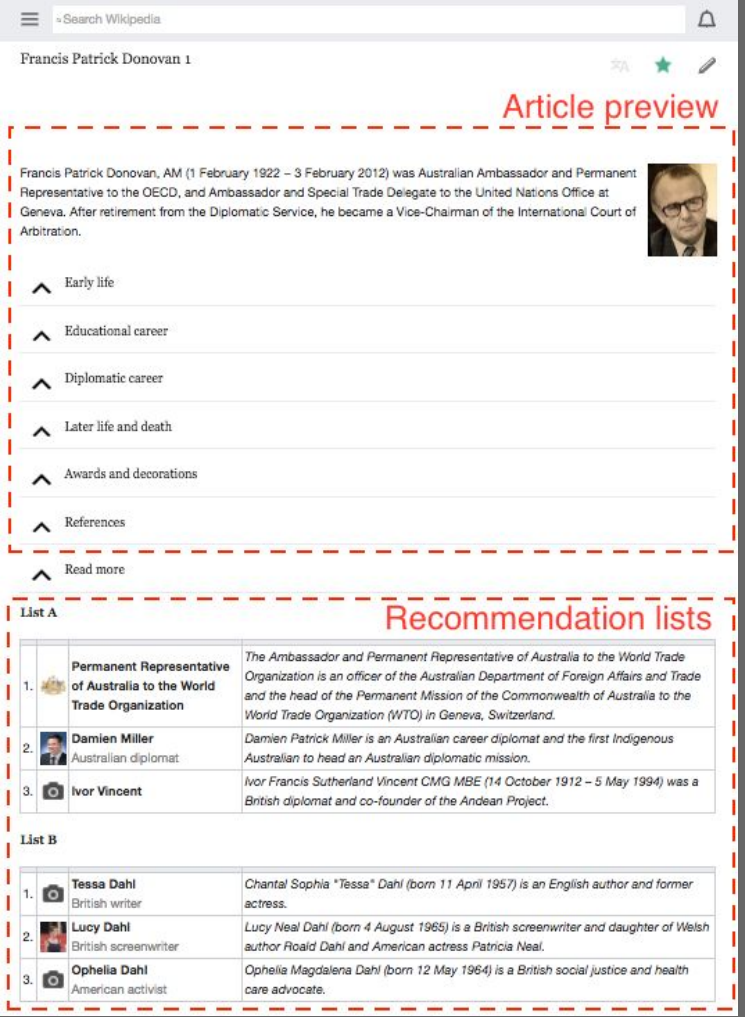
Figure 1: Screen shot of the experiment interface. Clicking on a movie in the list opens a pop-over with additional movie details.

Michael D. Ekstrand, F. Maxwell Harper, Martijn C. Willemsen, Joseph A. Konstan. *User perception of differences in recommender algorithms.* RecSys 2014.



Case study: Evaluating *related articles*

https://meta.wikimedia.org/wiki/Research:Evaluating_RelatedArticles_recommendations



1. **MoreLike:** articles with similar words (TF/IDF) in the lead section
2. **WikiVectors:** articles that tend to be read in the same browsing session ('link embeddings' a la Word2Vec)

Which algorithm works best? What do we mean by best? Who does it work best *for*?

https://meta.wikimedia.org/wiki/Research:Evaluating_RelatedArticles_recommendations

Search Wikipedia

Francis Patrick Donovan 1

Article preview

Francis Patrick Donovan, AM (1 February 1922 – 3 February 2012) was Australian Ambassador and Permanent Representative to the OECD, and Ambassador and Special Trade Delegate to the United Nations Office at Geneva. After retirement from the Diplomatic Service, he became a Vice-Chairman of the International Court of Arbitration.

- Early life
- Educational career
- Diplomatic career
- Later life and death
- Awards and decorations
- References
- Read more

Recommendation lists

List A

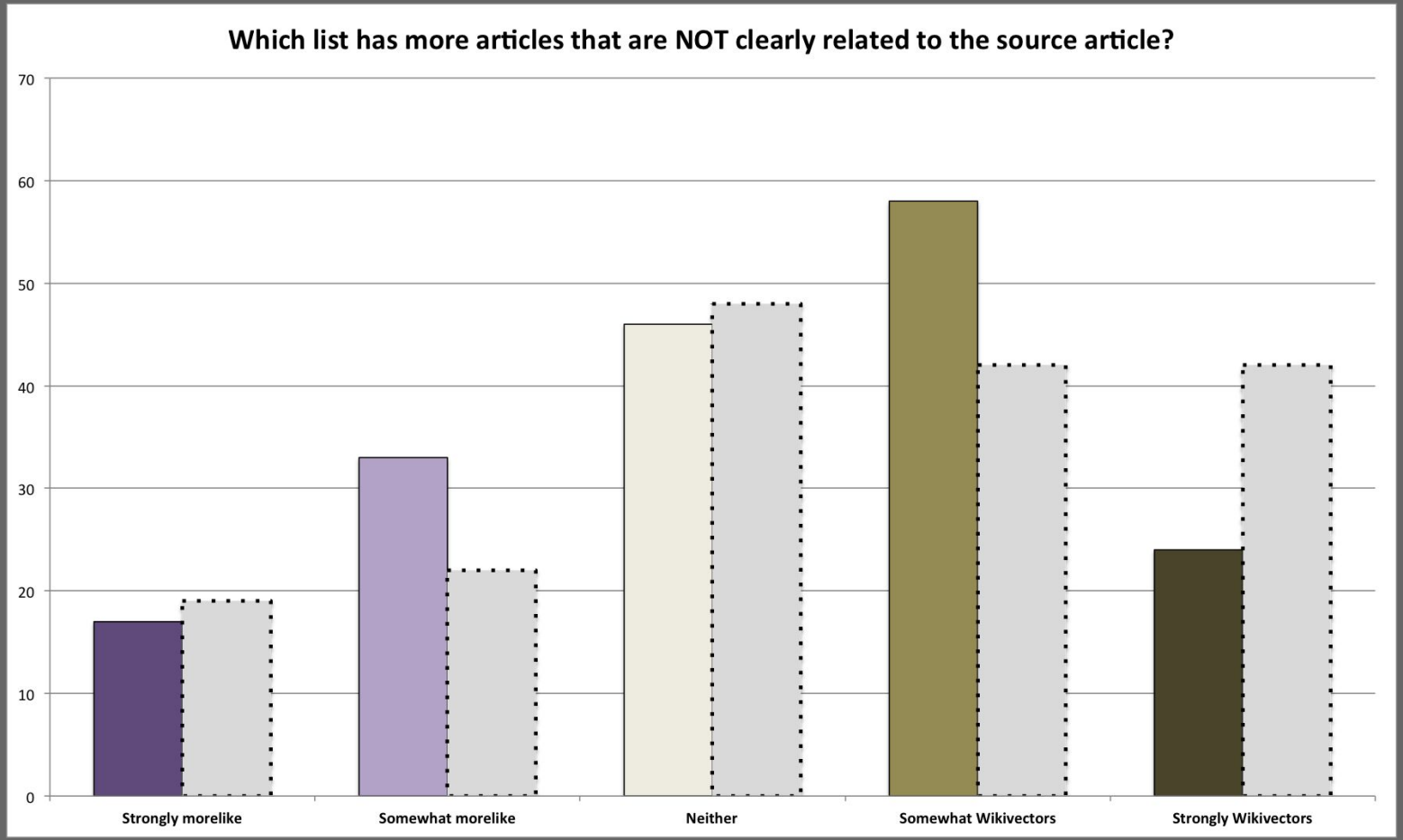
1.	**Permanent Representative of Australia to the World Trade Organization**	*The Ambassador and Permanent Representative of Australia to the World Trade Organization is an officer of the Australian Department of Foreign Affairs and Trade and the head of the Permanent Mission of the Commonwealth of Australia to the World Trade Organization (WTO) in Geneva, Switzerland.*
2.	**Damien Miller** Australian diplomat	*Damien Patrick Miller is an Australian career diplomat and the first Indigenous Australian to head an Australian diplomatic mission.*
3.	**Ivor Vincent**	*Ivor Francis Sutherland Vincent CMG MBE (14 October 1912 – 5 May 1994) was a British diplomat and co-founder of the Andean Project.*

List B

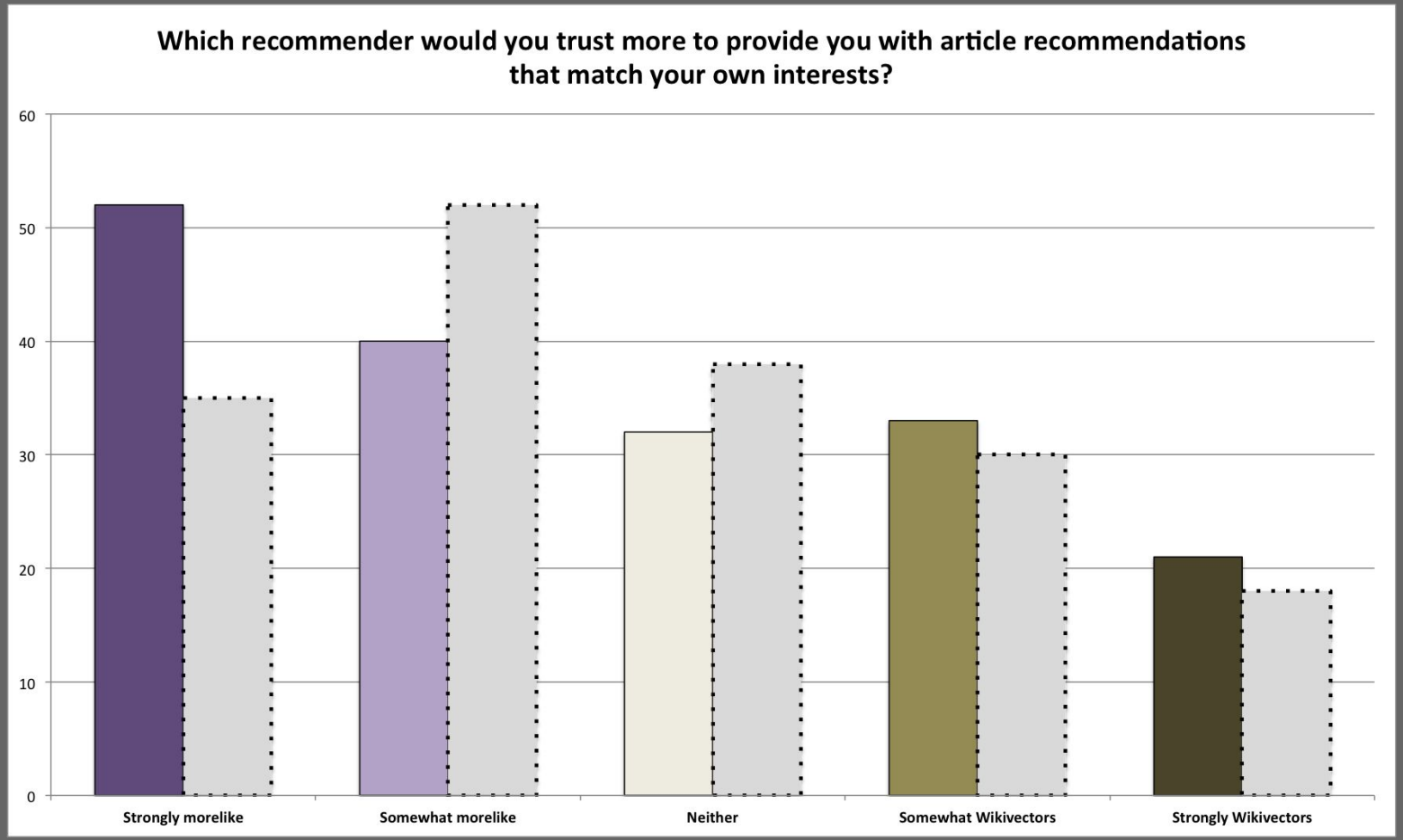
1.	**Tessa Dahl** British writer	*Chantal Sophia "Tessa" Dahl (born 11 April 1957) is an English author and former actress.*
2.	**Lucy Dahl** British screenwriter	*Lucy Neal Dahl (born 4 August 1965) is a British screenwriter and daughter of Welsh author Roald Dahl and American actress Patricia Neal.*
3.	**Ophelia Dahl** American activist	*Ophelia Magdalena Dahl (born 12 May 1964) is a British social justice and health care advocate.*

1. Which list has more articles that you would be interested in reading?
2. Which list has more articles that are similar to each other?
3. Which list has more articles that are NOT clearly related to the source article?
4. Which list contains the article that you would be most likely to read next?
5. Based on these two lists, which recommender would you trust more to provide you with article recommendations that match your own interests?
6. In a sentence or two, please describe why you do (or do not) prefer one list over the other?

https://meta.wikimedia.org/wiki/Research:Evaluating_RelatedArticles_recommendations



https://meta.wikimedia.org/wiki/Research:Evaluating_RelatedArticles_recommendations



https://meta.wikimedia.org/wiki/Research:Evaluating_RelatedArticles_recommendations

Neil Gorsuch 2

Neil McGill Gorsuch (/ˈgɔːrsətʃ/; born August 29, 1967) is an American federal appellate judge on the United States Court of Appeals for the Tenth Circuit. On February 1, 2017, President Donald Trump nominated Gorsuch to be an Associate Justice of the U.S. Supreme Court, to fill the seat left vacant after the death of Justice Antonin Scalia eleven months earlier, after announcing the nomination the previous day. Gorsuch is a proponent of textualism in statutory interpretation and originalism in interpreting the U.S. Constitution. Gorsuch clerked for Judge David B. Sentelle on the U.S. Court of Appeals for the D.C. Circuit from 1991 to 1992, and then for U.S. Supreme Court Justices Byron White and Anthony Kennedy, from 1993 to 1994.

- Early life and education
- Career
- Personal life
- Awards and honors
- Bibliography
- See also

https://meta.wikimedia.org/wiki/Research:Evaluating_RelatedArticles_recommendations

References

External links

Read more

List A

1.	**Harris Hartz** American judge	*Harris L Hartz (born January 20, 1947 in Baltimore, Maryland) is a federal judge on the United States Court of Appeals for the Tenth Circuit.*
2.	**David M. Ebel** United States federal judge	*David Milton Ebel (born June 3, 1940) is a federal appellate judge who has served on the United States Court of Appeals for the Tenth Circuit since 1988.*
3.	**Jerome Holmes** U.S. federal judge	*Jerome A. Holmes (born November 18, 1961 in Washington, DC) is a federal judge on the United States Court of Appeals for the Tenth Circuit.*

List B

1.	**Donald Trump Supreme Court candidates**	*With the advice and consent of the United States Senate, the President of the United States appoints the members of the Supreme Court of the United States, which is the highest court of the federal judiciary of the United States.*
2.	**Timothy Tymkovich** American judge	*Timothy Michael Tymkovich (born November 2, 1956) is the Chief Judge of the United States Court of Appeals for the Tenth Circuit.*
3.	**Neil Gorsuch Supreme Court nomination**	*In February 2016, Associate Justice Antonin Scalia of the Supreme Court of the United States died, leaving a vacancy on the highest federal court in the United States.*

https://meta.wikimedia.org/wiki/Research:Evaluating_RelatedArticles_recommendations

Inspiring trust

or WTF?!? Don’t look stupid

Inspiring trust in recommendations

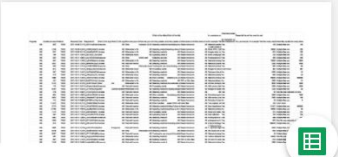
Fundamental questions:

- *Why should I, as a user,* trust that this algorithm understands who I am, what I like, and what I'm doing?
- *How can I, as a system designer,* communicate to the user that they *should* trust this algorithm to make decisions about what information they want (or need) to see?

Quick Access

Notes for themes
You open around this time

p4 interview notes
You edited this week



Responses as of 10-17-2017
You edited this week

Interview Analysis
You opened this week

Name ↓	Last modified	File size
survey	Oct 12, 2017 me	–
Personas / Scenarios	Jul 19, 2017 Niharika Ved	–
interview notes	Aug 31, 2017 me	–
Consent forms	Jul 18, 2017 me	–
Notes for themes	Oct 17, 2017 me	–

Re: Reminder and more: Q3 Goals are due soon

Inbox x

Deborah Tankersley
to tech-dept-info, me

from:	**Deborah Tankersley** <dtankersley@wikimedia.org>
to:	tech-dept-info@wikimedia.org
cc:	Jonathan Morgan <jmorgan@wikimedia.org>
date:	Fri, Dec 15, 2017 at 6:31 AM
subject:	Re: Reminder and more: Q3 Goals are due soon
mailed-by:	wikimedia.org
signed-by:	wikimedia.org
security:	Standard encryption (TLS) Learn more
:	Important according to our magic sauce.

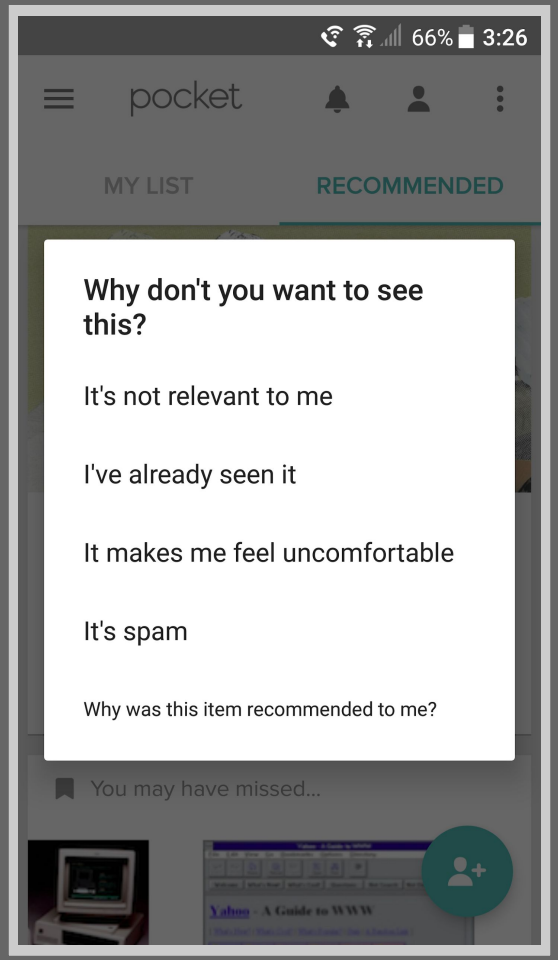
Yay, it's finally Frid o on vacation until January 2018!

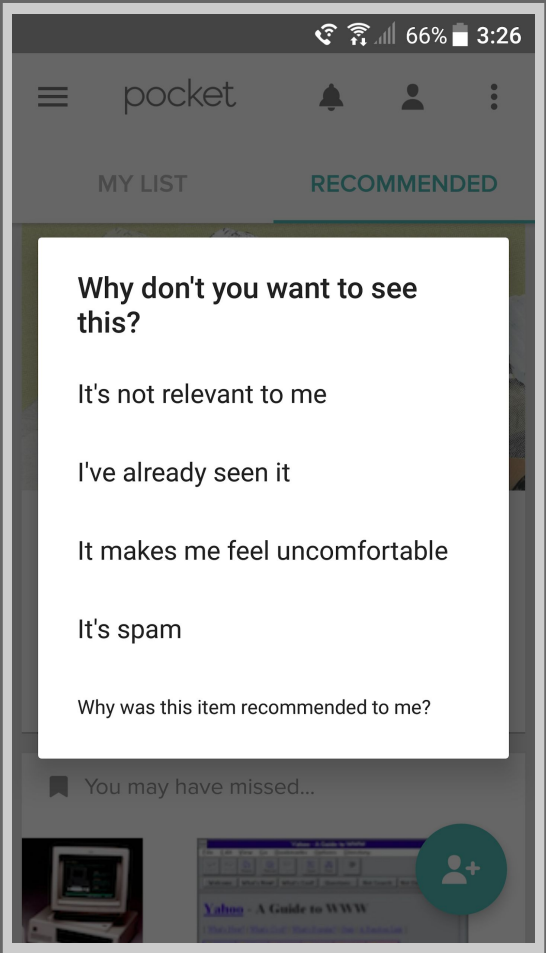
Got your Q3 goals written (and posted) yet? :)

Cheers,

Deb

--
deb tankersley
Program Manager, Engineering
Wikimedia Foundation


66% 3:26
pocket
MY LIST
RECOMMENDED
Why don't you want to see this?
It's not relevant to me
I've already seen it
It makes me feel uncomfortable
It's spam
Why was this item recommended to me?
You may have missed...
Yahoo - A Guide to WWW



“Recommendations are articles and videos that we think you’ll be interested in, sourced from the millions of items that are being saved to Pocket every day. The more you save and interact with Pocket, the more personalized your Recommendations will be.

Recommendations also come from the people you follow on Pocket. When someone you follow recommends something, it’ll appear in your feed alongside your personalized recommendations from Pocket.”

Case study: Citation recommendations

Recommending additional articles to cite in a research paper, based on the articles that are already cited.

McNee, S. M., Kapoor, N., & Konstan, J. A. (2006). *Don’t look stupid: avoiding pitfalls when recommending research papers.*

Case study: Citation recommendations

Recommending additional articles to cite in a research paper, based on the articles that are already cited.

“Bayes and PLSI perform well as recommenders in offline simulation experiments... Users, however, were not satisfied with these recommendation lists.

These results suggest that the research community’s dependence on offline experiments have created a disconnect between algorithms that score well on accuracy metrics and algorithms that users will find useful.”

McNee, S. M., Kapoor, N., & Konstan, J. A. (2006). *Don’t look stupid: avoiding pitfalls when recommending research papers.*

Case study: Citation recommendations

“In previous work, we argued that showing one good recommendation in a list of five was enough to satisfy users.

It is not that simple: showing one horrible recommendation in five is enough for users to lose confidence in the recommender.

We call this the Don’t Look Stupid principle: only show recommendation lists to users when you have some confidence in their usefulness.”

McNee, S. M., Kapoor, N., & Konstan, J. A. (2006). *Don’t look stupid: avoiding pitfalls when recommending research papers.*

how long do mls soccer games last

All | News | Shopping | Videos | Maps | More | Settings | Tools

About 207,000,000 results (0.47 seconds)

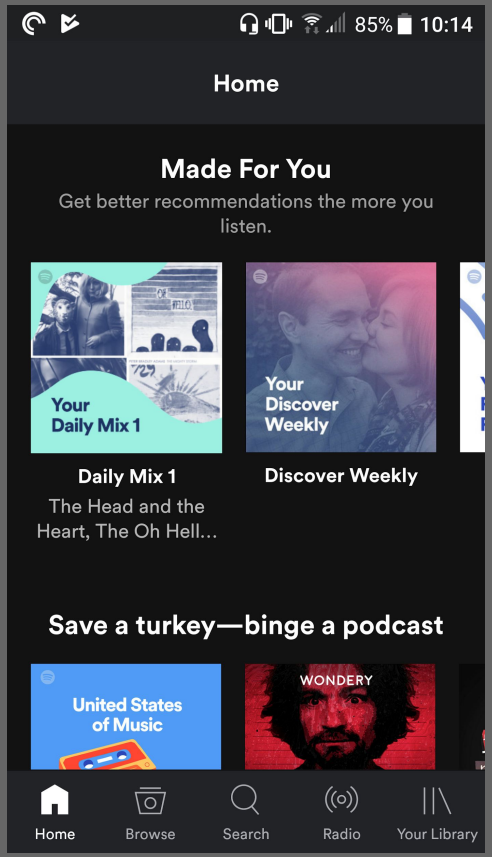
about 15 minutes

A: Major League Soccer matches consist of two 45-minute halves, with a halftime break of **about 15 minutes**. Stoppage time may be added to the end of each half at the referee's discretion to compensate for delays during the game. In general, most games last **about two hours**.

Frequently Asked Questions - Gameday | LA Galaxy
https://www.lagalaxy.com/es/club/faq/gameday

About this result Feedback

Case study: Music recommendations



Case study: Music recommendations

Comparing music recommendations between Echo Nest (Spotify), Google Instant Mix, and iTunes Genius by “WTF score”

https://musicmachinery.com/2011/05/14/how-good-is-googles-instant-mix/

Case study: Music recommendations

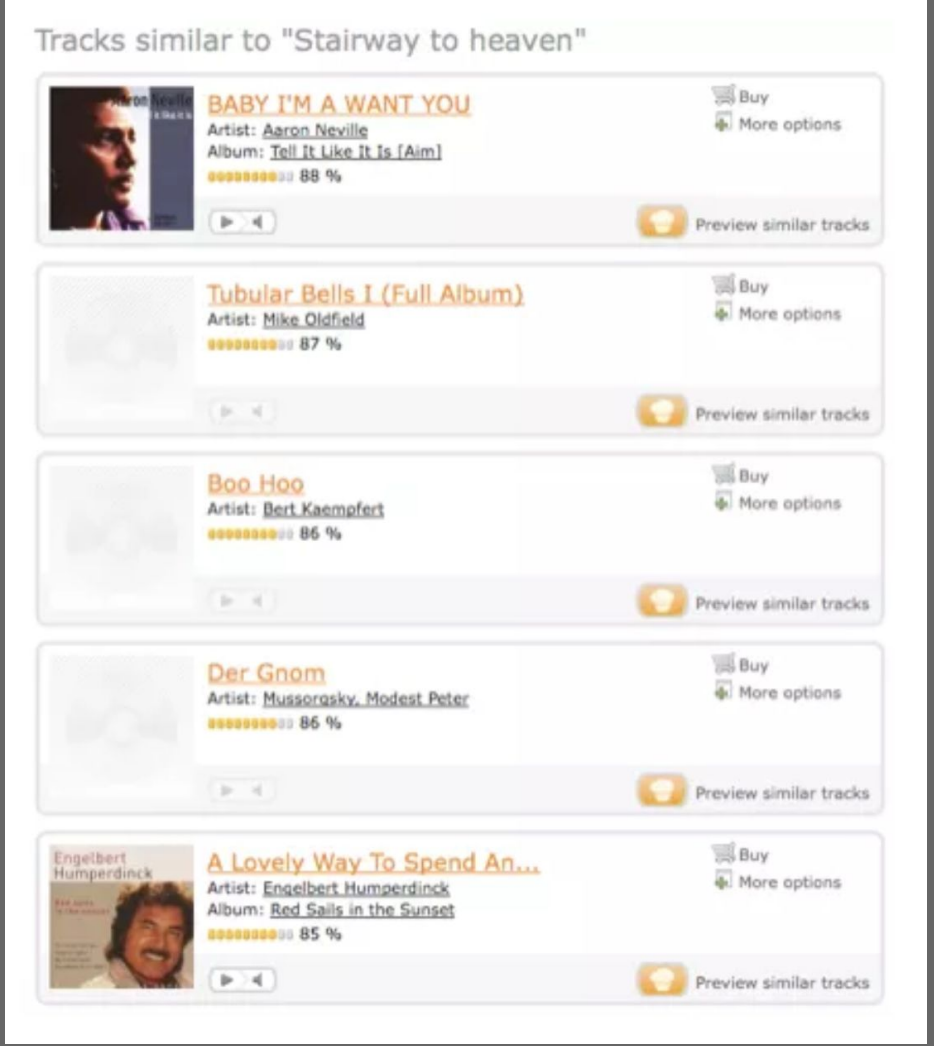
Comparing music recommendations between Echo Nest (Spotify), Google Instant Mix, and iTunes Genius by "WTF score"

"Evaluating playlists is hard. However, there is something that we can do that is fairly easy to give us an idea of how well a playlisting engine works compared to others.

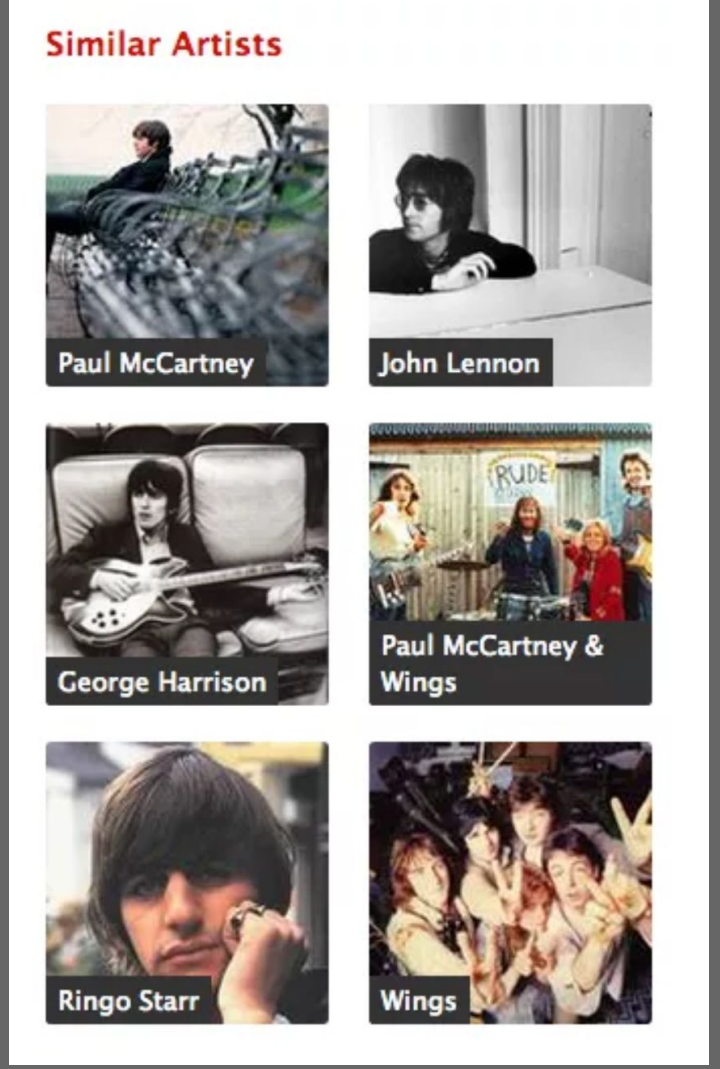
I call it the WTF test. It is really quite simple. You generate a playlist, and just count the number of head-scratchers in the list. If you look at a song in a playlist and say to yourself 'How the heck did this song get in this playlist' you bump the counter for the playlist. The higher the WTF count the worse the playlist."

https://musicmachinery.com/2011/05/14/how-good-is-googles-instant-mix/

“Show me tracks similar to *Stairway to Heaven*”



“Show me artists similar to *The Beatles*”



Human-recommender interaction model

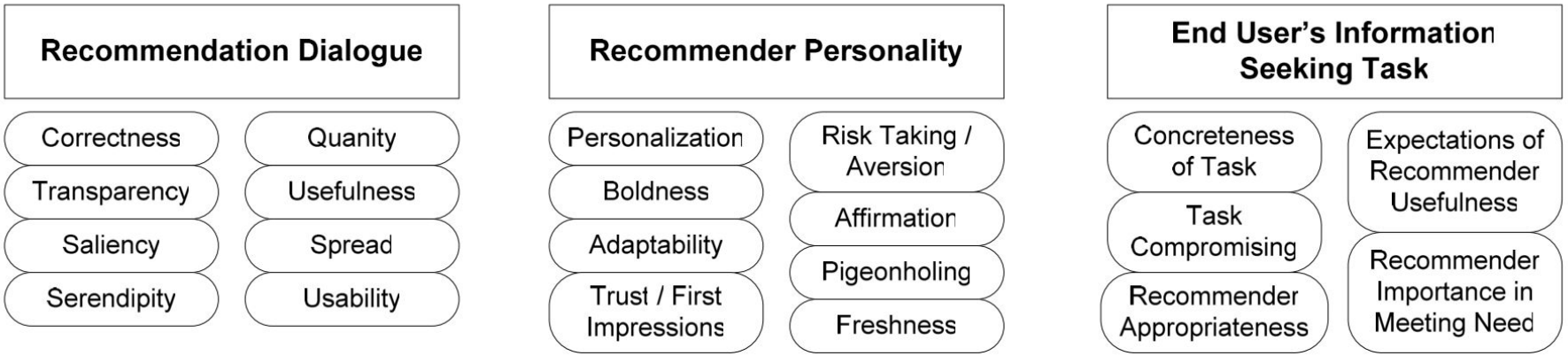


Figure 1-1: Aspects of Human-Recommender Interaction. The Aspects are divided into three 'Pillars'

Sean M. McNee, John Riedl, and Joseph A. Konstan. *Making recommendations better: an analytic model for human-recommender interaction. CHI EA 2006.*

Wrap up: Recommendations

In order to make recommendations that people will actually want to use, **inspire trust**.

Your users should...

- feel the recommendations meet their current needs (audience, purpose, context)
- feel like they understand how the recommendation was made (interpretability)
- not have a ‘WTF’ moment (don’t look stupid)
- not feel like the recommendation is invasive or embarrassing (don’t be creepy)
- feel like they have control over their experience (ask, don’t tell)

Questions to ask yourself

User experience questions

1. How do you know that you're making human-centered recommendations?
2. How does the presentation of your recommendations affect user trust?

Large scale risk/benefit questions

1. How do you know you're building human-centered technologies?
2. How might the way your technologies are developed, deployed, or used lead to harmful bias or other unintended consequences, for the end-user, indirect stakeholders, or society as a whole?

In-Class Activity: Ethical OS risk assessment

Groups of 3-4

In-class activity: EthicalOS Scenarios

1. Download the Ethical OS *Toolkit* & *Risk Mitigation Checklist* from Canvas
2. Read through the scenario(s) for the *Risk Zone* your group has been assigned
    a. Pages 16-29 of the *Toolkit*
3. Imagine that your team is in charge of developing and deploying some tech similar to what's described in the scenario you chose (but not explicitly for evil)
4. As a group, read through the Risk Mitigation checklist questions that correspond to your Risk Zone (e.g. "Truth, Disinformation, and Propaganda")
5. Answer these questions as a group, and record your answers.
    a. You will need to invent *additional details* about the tech, the design process, and/or the application in order to answer these questions.

Homework due next week (NO CLASS SESSION)

Reading reflection

- Hill, B. M., Dailey, D., Guy, R. T., Lewis, B., Matsuzaki, M., & Morgan, J. T. (2017). Democratizing Data Science: The Community Data Science Workshops and Classes

Final Project plan

- Due Week 9 (November 22)
- 10 points
- Min. 1000 words
- Jupyter Notebook or .md file on GitHub, link submitted to Canvas

Questions?