Issues with anonymization:

studying LibraryThing users to improve book search engines

Marijn Koolen Data Privacy and Ethics workshop 13 October 2015, Amsterdam, NL

Use Case

- Background: Social Book Search
- Anonymizing user data for research
- Questions on privacy and ethics

Search on LT Forums



Add books

Talk

Groups

Local

More

Zeitgeist

MarinusFDT | Sign out | Help



Search site

Politics of Multiculturalism Recommendations?

Political Philosophy

11 messages | * Star this topic | * Ignore topic | * Jump to bottom (0 unread)

1 steve.clason

Sep 26, 2010, 11:32pm 7

I'm new, and would appreciate any recommended reading on the politics of multiculturalism. Parekh's Rethinking Multiculturalism: Cultural Diversity and Political Theory (which I just finished) in the end left me unconvinced, though I did find much of value I thought he depended way too much on being able to talk out the details later. It may be that I found his writing style really irritating so adopted a defiant skepticism, but still...

Anyway, I've read Sen, Rawls, Habermas, and Nussbaum, still don't feel like I've wrapped my little brain around the issue very well and would appreciate any suggestions for further anyone might offer.

Reply More

Group: Political Philosophy



212 members

87 messages

You are not a member of this group.

About

This topic is not marked as primarily about any work, author or other topic.

Add...

Touchstones

Works

Rethinking Multiculturalism: Cultural Diversity and

BHIKHU PAREKH RETHINKING MULTICULTURALISM Cultural Oversity and Political Theory SECOND EDITION MEMORIPA BARBARITION

Main page
Work details

Reviews (1)

Recommendations

Members

Descriptions (2)

Conversations (1)

Common Knowledge

Popularity

Cover images

Editions

Rethinking Multiculturalism: Cultural Diversity and Political Theory

by Bhikhu Parekh

Your books

| Members R | eviews Popularit | y Average rating | Mentions |
|-----------|------------------|------------------|----------|
| 57 | 1 207,436 | ★★★ (2.75) | 1 |

Add to Your books Add to wishlist

Members

Recently added by saulibrary, jelledebont, imran_taib, LSEHumanRights, mi_hua, openomniproductions, bioethicscentre

▼ Tags

numbers show all

all members

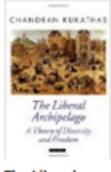
(B) anthropology British literature contemporary theory cultural studies diversity E E3 educational politics identity immigration liberalism Mread

multiculturalism multiculturalism fall 2005 Multiculturalism Studies non-fiction philosophy pluralism political philosophy political theory politics race read in 2012 religion & secularism SOC: Politics/Policy SS theory theory|philosophy TS

LibraryThing Recommendations



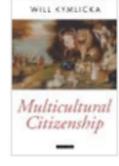
Culture and equality :... Brian Barry



The Liberal Archipelag... Chandran Kukathas



Multicultural Jurisdic... Ayelet Shachar



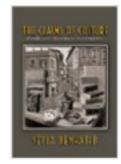
Multicultural Citizensh... Will Kymlikca



Multiculturalism (Theme... Tariq Modood



Multiculturalism: Exami... Charles Taylor



The Claims of Culture:... Seyla Benhabib



Vernac... Will Kymlikca



√

Strange Multiplicity:... James Tully

Show more ▼ | Vote | See complete recommendations

Member recommendations

refresh



Will you like it? Conversations (About links) No current Talk conversations about this book. » See also 1 mention ▼ Member reviews Although the book has a lot to offer an educated reader, the book is by an academic, written for academics, is written in an academic language with academic rhetorical flourishes and what seems to be an academic's meanness of spirit. Parekh certainly contributes to the discussion of how liberal democracies might evolve to become multicultural rather than assimilationist, most of us general readers will close this book long before we assimilate much. I read the last half very, very fast. Look elsewhere, in short. (**) steve.clason | Sep 26, 2010 | ®® 🖒 | flag ▼ Published reviews no reviews | add a review Other authors » Add/edit other authors (1 possible) ▼ Work-to-work relationships » Add a work-to-work relationship view history Common Knowledge Click the pencil to edit. Use the plus-sign to add multiple entries. Do not put quotes around entries. For more help see the Common Knowledge help page. Series (with order)

Canonical title

Original title

Alternative titles

People/Characters

Original publication date

Personal Profile

LibraryThing

Home

Groups

Zeitgeist

Profile Recommendations

Reviews

Stats/Memes

Clouds Authors

Member: steve.clason

Collections Your library (445), Currently reading (9), Lost, Misplaced, or Left Behind (31), eBook (85), Boulder Public Library (151), Norlin

Library (6), Prospector Libraries (23), Looks interesting (192), Read but unowned (106), Wishlist (13), Didn't Finish (1),

Programming etc. (41), As If (7), All collections (883)

Talk

Reviews 169 reviews

Tags fiction (132), philosophy (90), history (79), political theory (50), science fiction (39), thriller (35), boats (33), science (28),

democracy (19), politics (17) — see all tags

Media Not set (1), Book (878), Paper Book (810), Audiobook (3), Ebook (35), Unknown (4)

Clouds tag cloud, author cloud, tag mirror

Recommendations 10 recommendations

Groups American History, American Revolution & Founding Fathers History, Armchair Travellers, Books in 2025: The Future of the Book World, Colorado Bibliophiles,

Crime, Thriller & Mystery, History: On learning from and writing history, Philosophy and Theory, Political Philosophy, Progressive & Liberal! —show all groups

Favorite authors Joseph Conrad, Jürgen Habermas, Charles S. Peirce (Shared favorites)

Homepage http://www.frogsforsnakes.com

Real name Steve Clason

Account type public, lifetime

URLs /profile/steve.clason (profile)

/catalog/steve.clason (library)

Member since Aug 29, 2007

Currently reading The Political Culture of Leadership in the United Arab Emirates by Andrea B. Rugh

The Way of Water and Sprouts of Virtue (S U N Y Series in Chinese Philosophy and Culture) by Sarah Allan

Metaphors We Live By by George Lakoff

Shipwreck with spectator: paradigm of a metaphor for existence by Hans Blumenberg

Executive Air by Robert Roughton



Search s

Q

Search

Random

Misquoting

Legacy by

The pragr

Angle of F

Rights of I

Classics) |

Oceanogr

Religion a

Member

weighted

....giice

Hempel (5 ThufirHaw BoEberle

(53/2553)

adelapaz (62/1724)

tropology

(39/1070) markell (9

(100/4370

karen-s (4

(54/1621)

(show mo

Member

Personal Catalogue



User Data

- This is derived from publicly available data.
- Why useful for search engine research?
 - Forum requests help understand search needs and behaviour
 - User tags & reviews help index books
 - User profiles help personalise search results

NETFLIX

Netflix Prize



Home

Rules

Leaderboard

Update

The Netflix Prize Rules

For a printable copy of these rules, go here.

Overview:

We're quite curious, really. To the tune of one million dollars.

Netflix is all about connecting people to the movies they love. To help customers find those movies, we've developed our worldclass movie recommendation system: CinematchSM. Its job is to predict whether someone will enjoy a movie based on how much they liked or disliked other movies. We use those predictions to make personal movie recommendations based on each customer's unique tastes. And while Cinematch is doing pretty well, it can always be made better.

Now there are a lot of interesting alternative approaches to how Cinematch works that we haven't tried. Some are described in the literature, some aren't. We're curious whether any of these can beat Cinematch by making better predictions. Because, frankly, if there is a much better approach it could make a big difference to our customers and our business.

So, we thought we'd make a contest out of finding the answer. It's "easy" really. We provide you with a lot of anonymous rating data, and a prediction accuracy bar that is 10% better than what Cinematch can do on the same training data set. (Accuracy is a measurement of how closely predicted ratings of movies match subsequent actual ratings.) If you develop a system that we judge most beats that bar on the qualifying test set we provide, you get serious money and the bragging rights. But (and you knew there would be a catch, right?) only if you share your method with us and describe to the world how you did it and why it works.

Serious money demands a serious bar. We suspect the 10% improvement is pretty tough, but we also think there is a good chance it can be achieved. It may take months; it might take years. So to keep things interesting, in addition to the Grand Prize, we're also offering a \$50,000 Progress Prize each year the contest runs. It goes to the team whose system we judge shows the most improvement over the previous year's best accuracy bar on the same qualifying test set. No improvement, no prize. And like the Grand Prize, to win you'll need to share your method with us and describe it for the world.

Netflix Data Issues

- De-anonymization of Netflix users (Narayanan & Shmatikov, 2008)
- 'The issue is not "Does the average Netflix subscriber care about the privacy of his movie viewing history?,"
 - but "Are there any Netflix subscribers whose privacy can be compromised by analyzing the Netflix Prize dataset?"

Anonymizing LT Profiles

- Anonymize username, remove member info
- Remove day from cataloguing date (year-month)
 - alternative: remove dates, retain order
- Swap 1-5 books in catalogue with random other user
 - adds noise but item popularity remains stable

Example Profiles

| book-id | entry-date | rating | tags |
|---------|---|---|---|
| 952822 | 2012-02 | 0.0 | |
| 3344349 | 2012-02 | 0.0 | |
| 2317257 | 2012-02 | 10.0 | |
| 6415999 | 2012-02 | 8.0 | |
| 5525956 | 2012-09 | 0.0 | |
| 842432 | 2012-02 | 8.0 | |
| 3171103 | 2012-02 | 4.0 | |
| 542201 | 2009-05 | 10.0 | tarot |
| 5403381 | 2010-11 | 10.0 | potter, philosopher, Harry, stone |
| | 952822 3344349 2317257 6415999 5525956 842432 3171103 542201 | 952822 2012-02 3344349 2012-02 2317257 2012-02 6415999 2012-02 5525956 2012-09 842432 2012-02 3171103 2012-02 542201 2009-05 | 952822 2012-02 0.0 3344349 2012-02 0.0 2317257 2012-02 10.0 6415999 2012-02 8.0 5525956 2012-09 0.0 842432 2012-02 8.0 3171103 2012-02 4.0 542201 2009-05 10.0 |

. . .

Gathered Data

- 130,000 forum threads
 - posts, dates, book/author mentions
- 94,000 user profiles
 - books, cataloguing months, ratings, tags
- Use in research to improve algorithms
 - scientific practice urges sharing data

Remaining Issues

- profiles are highly unique
 - sparse data make identification easy
- Idiosyncratic tags: personal, unusual
 - perfect for identifying specific users
- What if users change/remove their profiles?

Questions

- Can this be used for research? Under what conditions?
- Does anonymisation/distortion change privacy status?
- Under what conditions can this be shared with peers?