

Algorithms as Data: A Case Study for Turning Algorithmic User Experiences into Research Data Objects

Jason A. Clark @jaclark

Montana State University
csv,conf,v6

What's Ahead

- Research Update
- Beyond Definition and Into Practice
- Algorithmic UX
- Algorithm audit techniques
- Research Implications

The Research

Institute of Museum and Library Services
Laura Bush 21st Century Librarian Program (LB21)



“RE:Search” - Unpacking the Algorithms That Shape Our UX

<https://github.com/jasonclark/algorithmic-awareness>

Practices of Algorithms

Documenting First Principles of Algorithms

Exercises in Writing Algorithms in Pseudocode

Pedagogy for the Algorithm (META)

1. Frame with a material consequence.
2. Introduce “First Principle”, not math formula.
3. Demonstrate the algorithm in action.
4. Leave space for group to discuss.

Algorithm as Symptom

A ghost, something unseen

What if?

Algorithms weren't the Ghost in the Machine.

Algorithms were understood **as** part of the **open data** continuum.

The UX of Algorithms

masking

filtering

searching

ranking

natural language processing

parsing

counting

sorting

vectorizing

How do you quantify Algorithmic UX?

Algorithms were under fire at a Senate hearing on social media

"Algorithms have great potential for good. They can also be misused."

TIM DE CHANT - 4/28/2021, 2:20 PM



Al Drago-Pool/Getty Images

Enlarge / Chairman Senator Chris Coons (D-Del.) (right) speaks with Sen. Ben Sasse (R-Neb.) during a Senate Judiciary Subcommittee on Privacy, Technology, and the Law hearing April 27, 2021 on Capitol Hill in Washington, DC. The committee is hearing testimony on the effect social media companies' algorithms and design choices have on users and discourse.

[Algorithms & Amplification - Subcommittee on Privacy, Technology, and the Law](#)

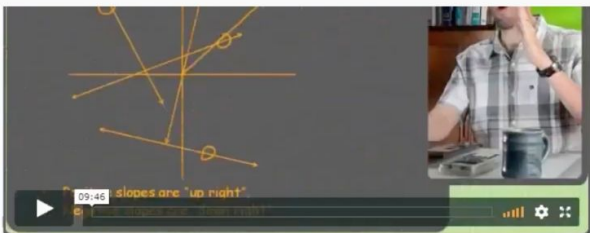
At the 2018 CES conference, YouTube's chief product officer said the site's recommendation engine is responsible for more than 70% of users' time spent watching videos on the platform.

Source: [Pew Research Analysis of YouTube Recommendation Algorithm](#)

How to Find the Slope of a Line

Calculating slope, graphing slope. This video also gets into the slope special cases a bit, such as the slopes of vertical, horizontal, and perpendicular lines.

Add to playlist

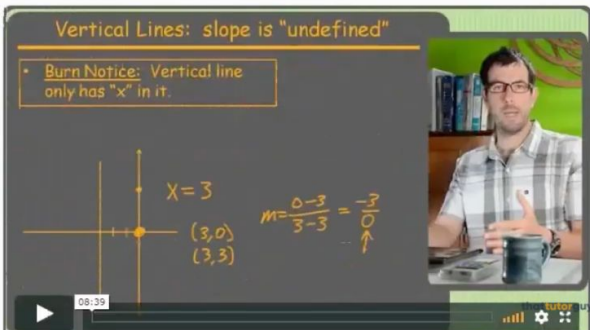


Horizontal & Vertical Lines

This video gets into the nitty gritty of horizontal and vertical lines. Why is there only one letter in them? How do you find the equation of a vertical line through a point? Why is the slope of a horizontal line zero, while vertical is undefined?

Add to playlist

Resume video | Start from the beginning



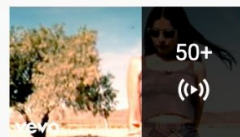
Check out the new ThatTutorGuy.com -- blue backgrounds, trial mention and a slide-up CTA!

Unlisted

403,745 views • Jun 29, 2019

695 likes 188 comments SHARE SAVE ...

All Recently uploaded



My Mix YouTube



TOP 20 ACOUSTIC GUITAR INTROS OF ALL TIME Rick Beato 12M views • 1 year ago



Celebrity Jeopardy! Kathie Lee, Tom Hanks, Sean Connery, Bur... Saturday Night Live 31M views • 7 years ago



Who is Getting Rich Off of High Lumber Prices?!? DIY with Dave 1M views • 3 weeks ago



Guardians of The Galaxy Awesome Mix - Vol. 1 & Vol. 2... Music For Memory 8.5M views • 2 years ago

Query for "algebra" on YouTube

Auditing Algorithms (Case Study)

Setting up software and scripting to understand and classify YouTube search recommendations.

Research Partners

- Marielle Korringa
- Julian Kaptanian
- Tyler Bass

Step One:

Simulating a User

Creating a user agent

Web scraping and harvesting

`webscraper.py`

Step Two:

Analyzing results

Evaluating the recommendation.

What are the tells in a recommendation?

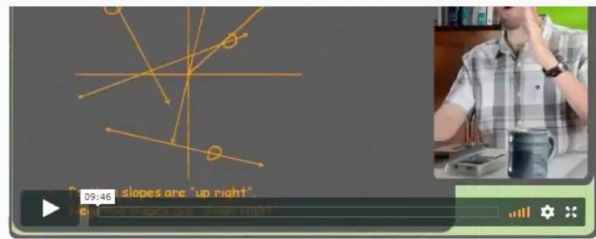
- ALL CAPS
- Excessive links in description
- Shift in content subject types

`analysis.py`

How to Find the Slope of a Line

Calculating slope, graphing slope. This video also gets into the slope special cases a bit, such as the slopes of vertical, horizontal, and perpendicular lines.

Add to playlist

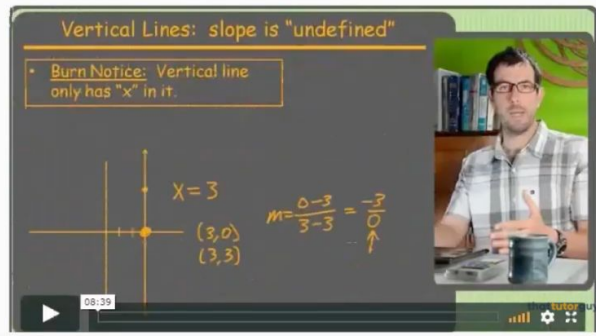


Horizontal & Vertical Lines

This video gets into the nitty gritty of horizontal and vertical lines. Why is there only one letter in them? How do you find the equation of a vertical line through a point? Why is the slope of a horizontal line zero, while vertical is undefined?

Add to playlist

Resume video | Start from the beginning



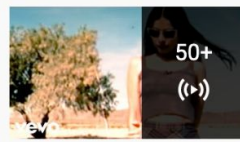
Check out the new ThatTutorGuy.com – blue backgrounds, trial mention and a slide-up CTA!

Unlisted

403,745 views • Jun 29, 2019

695 likes 188 comments SHARE SAVE ...

All Recently uploaded



My Mix YouTube



TOP 20 ACOUSTIC GUITAR INTROS OF ALL TIME Rick Beato 12M views • 1 year ago



Celebrity Jeopardy! Kathie Lee, Tom Hanks, Sean Connery, Bur... Saturday Night Live 31M views • 7 years ago



Who is Getting Rich Off of High Lumber Prices?!? DIY with Dave 1M views • 3 weeks ago



Guardians of The Galaxy Awesome Mix - Vol. 1 & Vol. 2... Music For Memory 8.5M views • 2 years ago

Query for "algebra" on YouTube

Step Three:

Packaging for Reuse

Creating structured data.

Encoding the analysis.

Setting up further analysis.

schema.org

Home **Schemas** Documentation

DataFeed

[Thing](#) > [CreativeWork](#) > [Dataset](#) > [DataFeed](#)

A single feed providing structured information about one or more entities or topics.

[more...]

Property	Expected Type	Description
Properties from DataFeed		
dataFeedElement	DataFeedItem or Text or Thing	An item within in a data feed. Data feeds may have many elements.
Properties from Dataset		
distribution	DataDownload	A downloadable form of this dataset, at a specific location, in a specific format.
includedInDataCatalog	DataCatalog	A data catalog which contains this dataset. Supersedes catalog , includedDataCatalog . Inverse property: dataset .
issn	Text	The International Standard Serial Number (ISSN) that identifies this serial publication. You can repeat this property to identify different formats of, or the linking ISSN (ISSN-L) for, this serial

<https://schema.org/Dataset>

<https://schema.org/DataFeed>

ro-crate



Research Object Crate

[View the Project on GitHub](#)
ResearchObject/ro-crate

This project is maintained by
[ResearchObject](#)

Hosted on GitHub Pages — Theme by orderedlist

RO-Crate Metadata Specification 1.0

- Permalink: <https://w3id.org/ro/crate/1.0>
- Status: Recommendation
- JSON-LD context: <https://w3id.org/ro/crate/1.0/context>
- This version: <https://w3id.org/ro/crate/1.0>
- Previous version: <https://w3id.org/ro/crate/0.2>
- Published: 2019-11-15
- Publisher: researchobject.org community
- Cite as: <https://doi.org/10.5281/zenodo.3541888> (this version)
<https://doi.org/10.5281/zenodo.3406497> (any version)
- Editors: Peter Sefton, Eoghan Ó Carragáin, Stian Soiland-Reyes
- Authors: Peter Sefton, Eoghan Ó Carragáin, Stian Soiland-Reyes, Oscar Corcho, Daniel Garijo, Raul Palma, Frederik Coppens, Carole Goble, José María Fernández, Kyle Chard, Jose Manuel Gomez-Perez, Michael R Crusoe, Ignacio Eguinoa, Nick Juty, Kristi Holmes, Jason A. Clark, Salvador Capella-Gutierrez, Alasdair J. G. Gray, Stuart Owen, Alan R Williams, Giacomo Tartari, Finn Bacall, Thomas Thelen

1. Introduction & definition of an RO-Crate

1. Terminology
2. Linked Data conventions

2. RO-Crate Structure

1. RO-Crate Metadata File ([ro-crate-metadata.jsonld](#))
2. RO-Crate Website ([ro-crate-preview.html](#) and [ro-crate-preview_files/](#))

a lightweight approach to
packaging research data with
their metadata

<https://w3id.org/ro/crate/1.0>

[#digital-library-and-repository
-content](#)

Auditing an Algorithm (META)

1. Create a user agent and profile.
2. Map out your harvesting data points.
3. Collect and analyze the data.
4. Encode results in structured data for reuse.

Implications for Research Data Practice

Implications

Identifying algorithms as datasets.

Encoding algorithms as datasets.

Prioritizing hidden experiences.

Moving beyond naming and defining.

Auditing additional Algorithmic UX.
(Amazon, Instagram, TikTok, etc.)

Code Repository

[Auditing tool](#)

[Algorithms in Search X-Ray app](#)

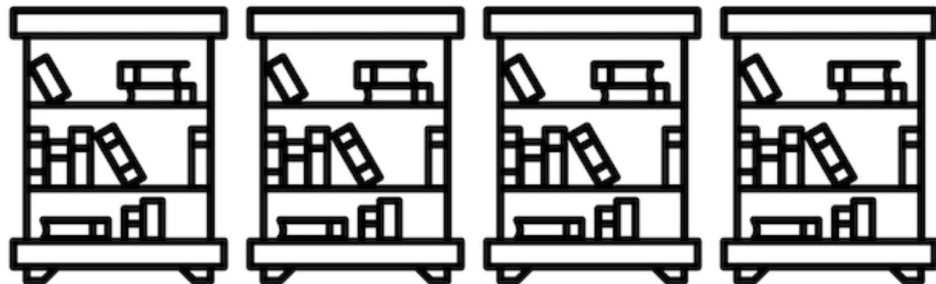
[Curriculum, Syllabus, Teaching Resources](#)

Research Inspiration

ALWAYS ALREADY COMPUTATIONAL - COLLECTIONS AS DATA

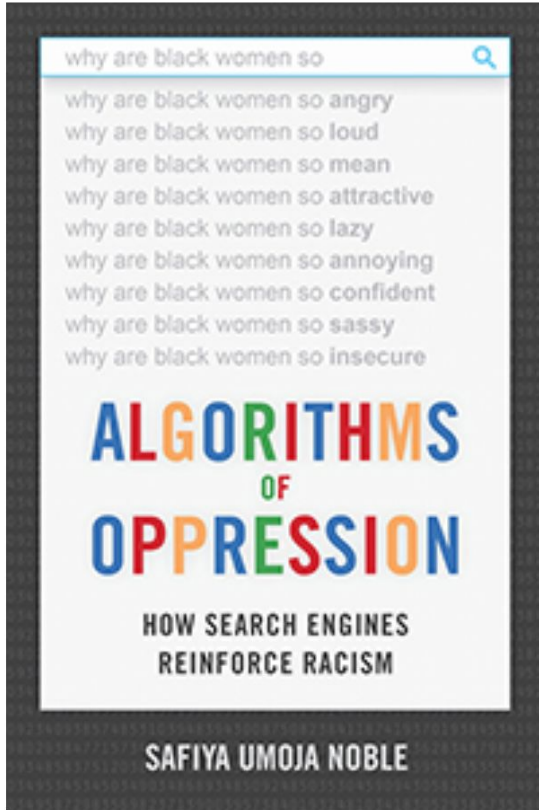


Always Already Computational



Thomas Padilla
Laurie Allen

[CollectionsAsData](#)



Algorithms of Oppression

How Search Engines Reinforce Racism

Safiya Umoja Noble

256 pages

57 b/w photos

February, 2018

ISBN: 9781479837243



Table of Contents



Introduction

<https://safiyaunoble.com/>

The Markup

Big Tech Is Watching You. We're Watching Big Tech.

About Us

The Markup is a nonprofit newsroom that investigates how powerful institutions are using technology to change our society. We are a new kind of media organization, staffed with an unparalleled roster of quantitative journalists who pursue meaningful, data-driven investigations.

Our approach is scientific: We build datasets from scratch, bulletproof our reporting, and show our work. We call this [The Markup Method](#).

themarkup.org/about

****Thank you****



Jason A. Clark
@jaclark

References and Follow-up Resources

Critical Algorithm Studies - Social Media Collective

<https://socialmediacollective.org/reading-lists/critical-algorithm-studies/>

AlgoTransparency Manifesto

<https://www.algotransparency.org/our-manifesto.html>

The Bigot in the Machine - Barbara Fister

<https://barbarafister.net/political/the-bigot-in-the-machine/>

Information Literacy in the Age of Algorithms - Project Information Literacy

https://projectinfolit.org/pubs/algorithm-study/pil_algorithm-study_2020-01-15.pdf

Can Auditing Eliminate Bias from Algorithms? - Ask The Markup

<https://themarkup.org/ask-the-markup/2021/02/23/can-auditing-eliminate-bias-from-algorithms>

List of audit algorithms for getting insights from black-box algorithms - GitHub

<https://github.com/erwanlemerrer/awesome-audit-algorithms>