Fashion, Art and Big Data: Mining the Vogue Archive

Lindsay King
Robots Reading *Vogue*: a collection of data mining experiments on the ProQuest *Vogue* Archive

http://dh.library.yale.edu/projects/vogue
The Vogue Archive as big data

If you read one page a minute for 8 hours a day, 40 hours a week, it would take you over three years to read every issue of Vogue.
The ProQuest Vogue Archive

- 125 years of American Vogue: 1892 - present
- 2,800+ issues
- 450,000+ pages
- Articles and images marked up with names of photographers, editors, designers, authors, advertisers, etc.
Standard ProQuest searching
Standard ProQuest searching
Olivia Vinall saw her first play at London’s National Theatre when she was fifteen. David Leveaux’s high-flying revival of "Jumpers," Tom Stoppard’s meta-physical love story, starring Simon Russell Beale. I remember being completely baffled, she says. But at the same time, it kind of blew my brain.

Now the 26-year-old is returning to the National and Stoppard; but this time on the other side of the footlights as she gets set to star in "The Hard Problem," the author’s first new play since 2006’s "Rock & Roll." Outgoing director Nicholas Hytner describes it as a 90 minutes straight through demanding, diamond hard, fierce. As it turns out, Vinall, named after the "Twelfth Night" beauty, is no stranger to the national stage. As Desdemona in Hytner’s "Othello" and Cordelia opposite Russell Beale in Sam Mendes’s "King Lear," she captivated audiences and critics with performances that were remarkably self-assured yet completely fresh, intricately detailed while still holding something back.
Beyond search, what can we do with full text and robust metadata?

- Sorting and filtering on metadata fields
- N-gram search (mapping frequency of words over time)
- Topic modeling (letting the text organize itself)
Sorting and filtering on metadata
N-gram search with Bookworm

- Date-stamped text: month, day, year
- Long timespan: 125 years
- Open-source code: Bookworm
- Close vs. distant reading: what's behind that peak?
“lovely” / “pretty” / “beautiful” / “sexy”
“career” / “education” / “family” / “business” / “college”
“career” / “family” / “college,” articles vs. ads
What’s behind the “college” peak in 1944…
Topic modeling

Topic Modeling Vogue

Computer-generated topics use statistical methods to suggest themes that emerge from term cooccurrence — how often certain words appear close to one another.

The topics on this page were generated by a computer program which automatically “read” all of the articles ever published in Vogue, and grouped together those words which were clustered more frequently than a normal distribution would predict.

Click on a the timeline displayed in each topic to see the highest-saturated articles for that year. Permalinks to articles will only resolve if your institution subscribes to the Vogue Archive.
Women’s health topic

“Women's Health”

Women's Health Words

cancer dr body medical
people make
cancer health women exercise
weight percent
time don’t good
don’t m.d.
medical

Women's Health Phrases

years ago health fitness
breast cancer men women
united states weight loss
lose weight plastic surgeon

Women's Health over Time

% > 20% saturation

1900 1920 1940 1960 1980 2000

1980 Articles

Janice Kaplan, “Beauty and Health: The Dirt" Diblma: Export Views 1 May, 1900: 519
"Beauty & Health: Fitness Now" 1 Sep, 1986: 401
"Fitness Q & A: Aerobics Vs. Anaerobic/Thired Muscles..." 1 Nov, 1993: 173
"Fitness Q & A: Motivation/Weight Training" 1 Mar, 1989: 936

Yale
Beyond text, what can we do with images and associated metadata?

- “Big picture” viewing
- Colormetric analysis
- Face detection
- Image similarity
Vogue covers, 1910s
Vogue covers, 1960s
Vogue covers, 1970s-80s
Vogue covers, 2000s
Patterns emerge

“The Americana Issue”
Patterns emerge

“Smart Clothes for College Girls”
Saturation by year
Saturation and hue by month
Color sorter: Dark Slate Blue #483d8b over time

Colors ordered by frequency:
74 covers with darkslateblue / #483D8B / rgb(72, 61, 139) / hsv(248, 56%, 55%) / hsl(248, 39%, 39%)
Highest percentage of Slate Grey #708090

Colors ordered by frequency:
114 covers with slategrey / #708090 / rgb(112, 128, 144) / hsv(210, 22%, 56%) / hsl(210, 13%, 50%)

Distribution of this color over time from 1911-2013:
Face Detection + Time

Christiana Wong, *Analysis of Vogue Fashion Photography’s Implications about the Female Face*
High Face Percentage + High Saturation =
Visual Similarity: Face Cream
Visual Similarity: Puffed Sleeves
Visual Similarity: Cars
What are we doing with this project?

- Demonstrating what digital humanities approaches can do
- Engaging different audiences from librarian/technologist/subject perspectives
- Experimenting with digital methods and tools for research and teaching
How can students work with this data?

- Using existing tools (n-gram search, colormetrics)
- Getting inspired to think about design history and change over time as a springboard for their work
- Creating their own computer science projects
Student projects using Vogue data

Robots Reading Vogue
Data Mining Is in Fashion

Student Work

What can you do with over 2,700 covers, 400,000 pages, 6 TB of data? Yale students are exploring the digital archive of Vogue to explore questions in fields from gender studies to computer science. We highlight some of these student projects below:

- **Pattern Detection through Vogue Cover Analysis over Periods of Social Change**
  - Clyde Edwards (Computer Science)
  - Read Paper [pdf]

- **Female image in Vogue magazine: A pictorial analysis of facial and body language**
  - Christiana Wong (Applied Math and Statistics)
  - See Presentation [pdf]

- **Tones in Vogue: An Analysis of Skin Color in Vogue Fashion Photography**
  - Saan Gabriel Petegorsky (Computer Science)
  - See Presentation
How can these ideas translate to other collections?

• Consider intrinsic metadata of available collections: e-journals, digitized special collections

• Increasing numbers of text and image collections available: museums, government collections, social media
Thank you!

Questions?