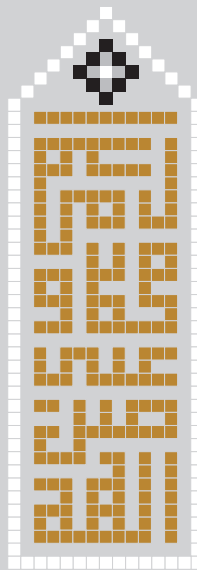


اسماء الوحمير المحبب

سرشناسه: مانوویچ، لی یف Manovich, Lev
عنوان و نام پدیدآور: تحلیل کلان‌داده‌های فرهنگی/ لو مانوویچ؛
مترجمان حسین حسینی، زهرا اسدی؛ ناظر علمی مجید سلیمانی ساسانی؛
ویراستار رضا صادقیان.
مشخصات نشر: تهران: سازمان تبلیغات اسلامی، حوزه هنری،
پژوهشکده فرهنگ و هنر اسلامی، ۱۴۰۱.
مشخصات ظاهری: ۴۹۴ ص.؛ ۵/۱۴×۲۱/۲۱ س.م.
شابک: ۹۷۸-۶۲۲-۹۴۷۱۶-۱-۶
یادداشت: عنوان اصلی: [2020]. Cultural analytics.
یادداشت: کتاب حاضر با حمایت مرکز ملی فضای مجازی، پژوهشگاه
فضای مجازی منتشر شده است.
موضوع: فرهنگ -- تحقیق -- روش‌های آماری
Culture -- Research -- Statistical methods
فرهنگ -- تحقیق -- داده‌پردازی
Culture -- Research -- Data processing
رسانه‌های گروهی -- تحقیق -- روش‌های آماری
Mass media -- Research -- Statistical methods
رسانه‌های گروهی -- تحقیق -- داده‌پردازی
Mass media -- Research -- Data processing
مصورسازی اطلاعات Information visualization
شناسه افزوده: حسینی، حسین، ۱۳۶۰-، مترجم
شناسه افزوده: زهرا، اسدی، ۱۳۴۷-، مترجم
شناسه افزوده: سلیمانی ساسانی، مجید، ۱۳۶۲-
شناسه افزوده: سازمان تبلیغات اسلامی. حوزه هنری. پژوهشکده فرهنگ و هنر اسلامی
شناسه افزوده: پژوهشگاه فضای مجازی
رده بندی کنگره: HM۶۲۲
رده بندی دیویی: ۲۸۵/۳۰۶
شماره کتابشناسی ملی: ۹۰۵۹۹۱۶



تحلیل کلان داده‌های فرهنگی

لو مانوویچ

مترجمان: حسین حسینی / زهرا اسدی





مرکز بین‌رشته‌ای مطالعات
پژوهشگاه فضای مجازی



پژوهشکده فرهنگ و هنر اسلامی
و پژوهشگاه فضای مجازی
تحلیل کلان داده‌های فرهنگی
لو مانوویچ

مترجمان: حسین حسینی
عضو هیئت علمی پژوهشگاه فرهنگ، هنر و ارتباطات
و زهرا اسدی

ناظر علمی: مجید سلیمانی ساسانی
عضو هیئت علمی دانشگاه تهران

ویراستار: رضا صادقیان
طراح جلد و گرافیک: صادق جمالی
صفحه‌آرا: احمد الهی

نشانی:

تهران، خیابان سمیه،
حوزه هنری، پژوهشکده
فرهنگ و هنر اسلامی
کدپستی: ۱۵۹۹۷۱۹۵۱۳
شماره تماس: ۸۸۸۹۶۳۳



با اسکن
رمزینسه از
دیگر آثار
پژوهشکده
آگاه
شوید

نقل و چاپ نوشته‌ها منوط به اجازه رسمی از ناشر است.

شماره تماس نشر پژوهشکده فرهنگ و هنر اسلامی: ۰۹۱۰۸۱۰۲۴۲۸

چاپ اول: ۱۴۰۱
شمارگان: ۵۰۰ نسخه
شابک: ۹۷۸-۶۲۲-۹۴۷۱۶-۱-۶

فهرست مطالب



- ۸ پیشگفتار
- ۱۰ مقدمه
- ۱۰ چگونه می‌توان یک میلیارد تصویر را دید؟
- ۱۳ نگرستن به فرهنگ، به وسیله رایانه‌ها
- ۱۹ تحلیل کلان داده‌های فرهنگ؛ پنج ایده
- ۳۳ تحلیل کلان داده‌های فرهنگ؛ دوازده مسئله پژوهشی
- ۳۷ تحلیل کلان داده‌های فرهنگ چه چیزی نیست؟
- ۳۹ تحلیل کلان داده‌های فرهنگ؛ نظریه رسانه‌ها و مطالعات نرم‌افزاری
- ۴۵ استفاده از این کتاب در کلاس‌ها



۴۸ بخش نخست؛ مطالعه فرهنگ در مقیاس بزرگ



- ۴۹ فصل نخست؛ از رسانه‌های جدید به رسانه‌های بیشتر
- ۵۷ از رسانه‌های جدید به رسانه‌های بیشتر (۲۰۰۸)
- ۶۳ مشاهده فرهنگ جهانی در زمان واقعی
- ۶۹ تحلیل کلان داده‌های فرهنگ در زمینه تاریخی
- ۷۳ فصل دوم؛ علم فرهنگ؟
- ۷۴ نمونه‌هایی از تحلیل و مصورسازی داده‌های فرهنگ و تعامل با آن‌ها
- ۸۵ گذشته در برابر زمان حال؛ حرفه‌ای‌ها در برابر تازه‌کارها
- ۹۰ عادی در برابر خاص
- ۹۳ علم فرهنگ؟! قوانین قطعی، مدل‌های آماری، شبیه‌سازی
- ۱۰۱ فصل سوم؛ صنعت فرهنگ و تحلیل کلان داده‌های رسانه‌ای
- ۱۰۷ مرحله جدیدی در تاریخ فناوری رسانه
- ۱۰۸ نمونه‌هایی از تحلیل کلان داده‌های رسانه‌ای
- ۱۱۳ دو بخش تحلیل کلان داده‌های رسانه‌ای
- ۱۱۷ خودکاربودن؛ تحلیل رسانه‌ای
- ۱۲۴ خودکارکردن؛ کنش‌های رسانه‌ای
- ۱۳۱ تحلیل کلان داده‌های رسانه‌ای و تحلیل کلان داده‌های فرهنگی



بخش دوم، بازنمایی فرهنگ به عنوان داده ۱۳۹

- ۱۳۹ فصل چهارم؛ انواع داده‌های فرهنگی
۱۴۴ رسانه؛ شبکه‌های اجتماعی و شبکه‌های حرفه‌ای
۱۵۳ رفتار؛ ردپاهای دیجیتال و فیزیکی
۱۶۴ بازنمایی تعامل
۱۶۸ رویدادها، مکان‌ها، سازمان‌ها
- ۱۷۳ فصل پنجم؛ نمونه‌گیری فرهنگی
۱۷۷ جزایر و اقیانوس
۱۸۳ موزه‌ها در برابر کتابخانه‌ها
۱۸۷ ایجاد نمونه‌های معرف
۱۹۲ چگونه می‌توان چیزهای نامرئی را دید؟
۲۰۳ محدودیت‌های نمونه‌های تصادفی
۲۰۷ آمار به عنوان تقلیل
۲۱۳ چرا برای مطالعه فرهنگ‌ها به کلان داده‌ها نیاز داریم؟
۲۱۶ آیا نمونه‌گیری ضروری است؟
- ۲۲۳ فصل ششم؛ فراداده و ویژگی‌های آن
۲۲۵ از یک جهان به یک مجموعه داده
۲۳۰ فراداده و ویژگی‌های آن
۲۳۶ داده‌ها = اشیا + ویژگی‌ها
۲۴۶ آمار در قرن‌های نوزدهم و بیستم؛ از یک متغیر تا چندین متغیر
۲۵۲ تفسیر، تبیین، خودکاری
۲۵۷ شکاف معنایی
- ۲۶۵ فصل هفتم؛ زبان، مقوله‌ها، حس‌ها
۲۶۶ انواع داده‌ها
۲۷۲ مقیاس‌های اندازه‌گیری
۲۸۰ زبان و حواس
۲۹۰ حواس و اعداد
۲۹۵ اندازه‌گیری دریافت‌ها
۳۰۰ تحلیل بالا به پایین و پایین به بالا
۳۰۶ زیبایی‌شناسی تجویزی و مدرنیسم
۳۱۳ نمونه‌های تحلیل: ۷۷۶ نقاشی ون‌گوگ و یکمیلیون صفحه مانگا
۳۲۵ نمونه‌های بیشتر: یکمیلیون اثر هنری و ۴۲،۵۷۱ فیلم
۳۳۱ جامعه مقوله‌ها

۳۳۹ بخش سوم؛ مطالعه کلان داده‌های فرهنگی



۳۳۹ فصل هشتم؛ مصورسازی اطلاعات

۳۴۳ مصورسازی اطلاعات

۳۴۳ مصورسازی چیست؟

۳۴۸ تقلیل و فضا

۳۵۸ مصورسازی بدون تقلیل

۳۶۱ مصورسازی رسانه‌های هنری

۳۶۷ سری‌های زمانی فرهنگی

۳۷۰ فراتر از مصورسازی اطلاعات

۳۷۵ فصل نهم؛ تحلیل اکتشافی رسانه‌ها

۳۷۸ در برابر جست‌وجو

۳۸۷ رابط

۳۹۲ پردازش تصویر و بینایی رایانه

۳۹۷ استفاده از ویژگی‌های تصویر برای تحلیل اکتشافی رسانه

۴۰۲ دیدن در مقابل تحلیل کردن

۴۰۵ فصل دهم؛ روش‌های مصورسازی رسانه

۴۰۹ مونتاژ تصویر

۴۲۲ نمونه‌گیری در مقابل خلاصه‌سازی

۴۲۸ نمونه‌گیری زمانی

۴۳۱ نمونه‌گیری فضایی

۴۳۵ تصویرگری مجدد

۴۴۳ نتیجه‌گیری: آیا می‌توانیم بدون مقوله‌ها بیندیشیم؟



۴۴۸ آیا ما می‌خواهیم فرهنگ را «تبیین» کنیم؟

۴۵۲ آیا هدف تحلیل کلان داده‌های فرهنگی مطالعه الگوهاست؟ (بله و خیر)

۴۵۶ چگونه بدون مقوله‌ها بیندیشیم؟

۴۶۰ یادگیری دیدن در مقیاسی جدید



۴۶۲ یادداشت‌ها



پیش‌گفتار



رسانه‌های اجتماعی جهانی جدید را خلق و نظام ارتباطات اجتماعی را متحول کرده‌اند. آنچنان‌که استفاده از رسانه‌های اجتماعی به‌عنوان امری روزمره تلقی شده و بسیاری از کنش‌های سیاسی، اجتماعی، فرهنگی و اقتصادی و خدمات از طریق این رسانه‌ها رخ می‌دهد. به نظر می‌رسد که عطف به همین مسئله، رسانه‌های اجتماعی به‌واسطهٔ انقلاب دیجیتال، نه‌تنها شئون زندگی فردی، بلکه ساختارهای سیاسی و اقتصادی را نیز دگرگون کرده است. از این رو، لازم است با خوانش انتقادی به سراغ سازوکارهای جدید تولید سرمایه‌داری در شکل‌گیری و ادامهٔ حیات پلتفرم‌های رسانه‌های اجتماعی رفت.

در همین زمینه، گروه مطالعات رسانهٔ پژوهشکدهٔ فرهنگ و هنر اسلامی و گروه مطالعات فرهنگی و اجتماعی پژوهشگاه فضای مجازی نیز فعالیت گسترده را در حوزهٔ «مطالعات رسانه‌های اجتماعی» آغاز کرده‌اند تا اندکی از فقر و فقدان ادبیات نظری در این حوزه کاسته شود. بخشی از این فعالیت پژوهشی متوجه نقش رسانه‌های اجتماعی در ظهور مرجعیت‌های جدید و حوزه‌های اثرگذاری

سلبریتی‌ها و بخش دیگر نیز بر مطالعات نظری، تجربی و انتقادی رسانه‌های اجتماعی تمرکز دارد. کتاب حاضر نیز یکی از آثار این مجموعه است که به مسئله کلان‌داده‌های رسانه‌های اجتماعی پرداخته است. با این حال، حوزه مطالعات رسانه‌های اجتماعی به این اثر محدود نخواهد شد.

از تلاش‌های آقای دکتر حسین حسنی و خانم دکتر زهرا اسدی مترجمان این اثر و همچنین آقای دکتر مجید سلیمانی ساسانی، ناظر علمی کتاب حاضر کمال قدردانی را داریم. امید است که فعالیت‌های پژوهشی در حوزه «مطالعات انتقادی رسانه‌های اجتماعی» با جدیت ادامه یابد تا از این رهگذر امکان‌های جدیدی برای فهم مسائل جدید ارتباطات اجتماعی فراهم شود.

در پایان، از حمایت‌های بی‌دریغ «پژوهشگاه فضای مجازی» که در تولید و چاپ این اثر، مشارکت مؤثر داشته‌اند، کمال تشکر را داریم.

پژوهشکده فرهنگ و هنر اسلامی

پژوهشگاه فضای مجازی



مقدمه



اگرچه در بلندمدت، تأثیر رایانه بر علوم انسانی [نسبت به علوم فیزیکی و زیستی] تحول آفرین تر به نظر می‌رسد، یکی از دلایل آن به‌طور ساده ناشی از تمایل به استفاده از این فرصت است یا دلیل آن وجود پیش‌زمینه‌ای است که در اثر استفاده گسترده از پردازش و به‌ویژه پردازش آماری ایجاد شده است و رایانه‌ها در تسهیل آن نقش داشته‌اند. علاوه بر این، شاید به این دلیل باشد که رایانه در یک معنا همانند یک تلسکوپ یا میکروسکوپ به ما امکان می‌دهد تا مشاهده کنیم. به عبارت دیگر، اگر پردازش رایانه‌ای به‌درستی انجام شده باشد، ما را قادر می‌سازد که در انبوه داده‌ها به بررسی رابطه‌ها و الگوهایی بپردازیم که پیش‌تر فراتر از درکمان بود یا اینکه پیامدهای حقیقی ایده‌ای را که بر داده‌ها اعمال می‌شود، مشاهده کنیم؛ پیامدهایی که گرچه منحصر به فرد نیستند، یقیناً اجتناب‌ناپذیر و سریع‌اند. دل‌اچ. هایمز، «مقدمه»،

در کتاب استفاده از رایانه‌ها در انسان‌شناسی، ۱۹۶۵: ۱.

چگونه می‌توان یک میلیارد تصویر را دید؟

این کتاب، فصل مشترک علوم داده‌ها و مطالعات رسانه‌ها و مطالعات فرهنگ دیجیتال است که مفاهیم و روش‌های برگزیده را برای تحلیل محاسباتی داده‌های

فرهنگی، ارائه می‌کند. از این روش‌ها می‌توانیم برای مطالعه مصنوعات تاریخی دیجیتال شده^۱ و رسانه‌های دیجیتال معاصر بهره ببریم و نیز می‌توانیم برای مطالعه یک مصنوع واحد یا چند نمونه از آن استفاده کنیم؛ اما اگر این روش‌ها را به‌طور خاص برای مطالعه میلیون‌ها مصنوع به کار ببریم، از اهمیت بیشتری برخوردار می‌شوند.

در واقع، مقیاس حیرت‌انگیز فرهنگ دیجیتال مراوداشت تا در سال ۲۰۰۵ مطالعه این روش‌ها را آغاز کنم؛ مطالعه‌ای که در نهایت به نوشتن این کتاب منجر شد. چگونه می‌توانیم عکاسی مردم‌پسند معاصر^۲ را که هر روز با تولید میلیاردها تصویر رشد می‌کند، درک کنیم یا موسیقی معاصر را که با هم‌رسانی صدها میلیون ترانه در ساوندکلاد^۳ به وسیله بیست میلیون سراینده بازنمایی می‌شود یا محتوای چهارمیلیون بُرد^۴ در بین‌ترست (۲) یا الگوهای علایق فکری مردم که ماهانه در ۱۹۰ کشور، در ۳۳۰ هزار گروه و ۸۴ هزار رویداد در میتاپ^۵ (حدود سال ۲۰۱۹) بازتاب داده می‌شود؟ (۳) این‌ها هم «فرهنگ دیجیتال» هستند؛ زیرا این رویدادهای فیزیکی به وسیله پلتفرم وبی میتاپ میسر می‌شود. به نظر من، در چنین مقیاسی، تنها راه برای مطالعه الگوها و روندها و پویایی‌های فرهنگ معاصر، استفاده از روش‌های «علم داده» است.

برای استفاده از این کتاب، لازم نیست پیش‌زمینه‌ای در علوم داده یا برنامه‌نویسی یا آمار یا ریاضی داشته باشید. مخاطبان موردنظر من، پژوهشگران دانشگاهی و دانشجویان رشته‌های هنر و طراحی و علوم انسانی و علوم اجتماعی و مطالعات رسانه و علوم داده و علوم رایانه هستند؛ همچنین متخصصانی که در زمینه طراحی، عکاسی، فیلم‌سازی، طراحی شهری، معماری، روزنامه‌نگاری،

1. Digitized Historical Artifacts

2. Contemporary Popular Photography

3. Sound Cloud

4. Board

5. Meetup Groups

حوزه‌های موزه و کتابخانه، مجموعه‌داری و مدیریت فرهنگی کار می‌کنند و نیز همه کسانی که در رسانه‌های اجتماعی و وب در هر نقشی فعالیت دارند؛ مانند تولیدکننده، وبلاگ‌نویس، استراتژیست، مدیر، توسعه‌دهنده، بازاریاب.

حتی اگر به تحلیل مجموعه داده‌های فرهنگی علاقه‌ای نداشته باشید، باز هم روزانه با چنین تحلیل‌هایی مواجه می‌شوید. شاید در پی تحلیل کلان داده‌ای فیس‌بوک، اینستاگرام، ویدیو یا باشید یا بخواهید از ابزار گوگل آنالیتیکس^۱ برای تحلیل وبلاگ یا وبسایت خود استفاده کنید یا در محل کارتان از پیشخوان نظارت^۲ بر رسانه اجتماعی استفاده می‌کنید.

اگر هم به چنین داده‌هایی توجه نکنید، هنگام انجام هر کار دیجیتالی، مدام با نتایج تحلیل محاسباتی سروکار دارید؛ مثلاً هر بار که عکسی می‌گیرید، الگوریتم‌های دوربین گوشی همراهمان، به‌طور خودکار، نوردهی را انتخاب و کنتراست عکس را تنظیم می‌کند و نوع صحنه و اشیای موجود در عکس را تشخیص می‌دهد.^(۴)

آنچه جست‌وجوی وب و پیشنهادهای فیلترکردن و سفارشی‌سازی^۳ و تعامل با دستگاه‌های دیجیتال و تبلیغات رفتاری^۵ و دیگر عملیاتی که «واژگان» فرهنگ دیجیتال را شکل می‌دهد امکان‌پذیر می‌سازد، تحلیل محاسباتی مصنوعات رسانه‌ای و تعاملات کاربر است؛ برای مثال، موتورهای جست‌وجوی وب، مانند بایدو^۶ یا بینگ^۷ یا یاندکس^۸ یا گوگل، دائماً به تحلیل محاسباتی محتوای میلیاردها صفحه وب و تصویرهای برخط و دیگر محتواهای وب متکی‌اند تا نتایج مرتبط را برای شما نمایش دهند.

1. Weibo

2. Google Analytics

3. monitoring dashboard

4. Customization

5. Behavioral Advertising

6. Baidu

7. Bing

8. Yandex

فکر می‌کنم برای اینکه در جامعه کنونی باسواد شناخته شوید، باید آن دسته از مفاهیم مرکزی و اصولی‌ای را بشناسید که چنین عملیاتی را امکان‌پذیر می‌کند. این کتاب، مقدمه‌ای ساده و غیرفنی است برای دانستن بعضی از این مفاهیم. به‌طور خلاصه، این کتاب به شما می‌آموزد که چگونه مجموعه داده‌های فرهنگی را مطالعه کنید و نیز توضیح می‌دهد که جامعه ما چگونه با استفاده از داده‌ها و الگوریتم‌ها می‌اندیشد.

نگریستن به فرهنگ، به وسیله رایانه‌ها

در این کتاب، نمونه‌های زیادی از تحلیل رایانشی کلان‌داده‌های فرهنگی ارائه می‌کنم که توسط بسیاری از پژوهشگران انجام شده یا در آزمایشگاه خودم به دست آمده است. اکنون می‌خواهم با شرح دو مثال، امکانات و دشواری‌های این تحلیل را نشان بدهم.

نخستین مثال، طرحی با نام «جای دیگر»^۱ است که من و همکارانم از سال ۲۰۱۸ در حال تحقیق درباره آن هستیم. در این طرح، رشد و انتشار فرهنگ معاصر را در تعداد زیادی از شهرهای کوچک، در مقایسه با چند نمونه از پایتخت‌های جهانی، مطالعه می‌کنیم.

امروزه شماری از این پایتخت‌ها در توجه به رسانه‌ها و مطالعات تحقیقاتی و رتبه‌بندی‌های مختلف، همچنان ضعیف عمل می‌کنند؛ بنابراین، به سادگی می‌توانیم فرض کنیم که تعداد اندکی از شهرهای «برتر»^۲ همچنان «مرکز» به شمار می‌آیند و بقیه شهرهای جهان هنوز در حاشیه قرار دارند و ایده‌های جدید را با تأخیر دریافت می‌کنند. اما چگونه می‌توانیم به تصویر واقعی دست پیدا کنیم؟ چگونه جهانی شدن و ظهور فناوری‌های نوین ارتباطی، جغرافیای فرهنگ^۱ را تغییر داد؟ آیا امروزه

1. Elsewhere

2. top

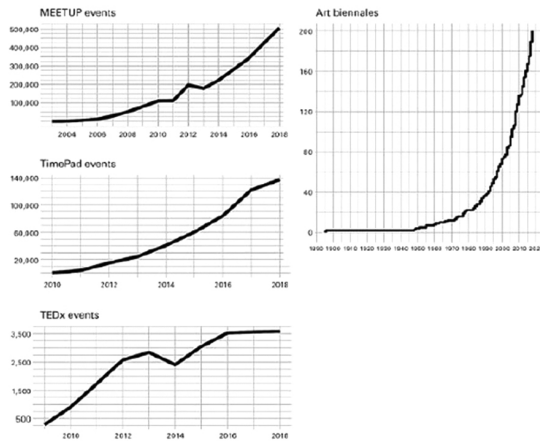
می‌توانیم روند فرهنگی معاصر را در هزاران شهر کوچک‌تر نیز بیابیم؟ آیا ممکن است شماری از این شهرها صرفاً به سبب فاصله داشتن با پایتخت و کوچک‌تر بودن، از نظر فرهنگی نوآورتر باشند؟ آیا ممکن است هنوز بخش‌های زیادی از جهان از این روندها آگاه نباشند و نوآوری نداشته باشند؟ چگونه از آغاز جهانی شدن، فرهنگ معاصر در سراسر جهان توسعه و گسترش یافته است؟ آیا این رشد در برخی از دوره‌ها یکنواخت یا ناهموار بوده است یا تند یا کند؟ آیا این الگوهای رشد^۲ در عرصه‌های فرهنگی^۳ مختلف، یکسان‌اند یا هریک به شیوه خاص خود توسعه می‌یابند؟

مطمئناً هیچ طرحی نمی‌تواند به همه این پرسش‌ها پاسخ دهد. هدف طرح «جای دیگر»، توسعه و آزمودن روش‌شناسی جدیدی است. این روش‌شناسی از داده‌های عمومی^۴ در زمینه رویدادهای فرهنگی^۵ و مکان‌هایی بهره می‌گیرد که این رویدادها را سازمان‌دهی می‌کنند. تعداد این مکان‌ها و رویدادها در جهان امروز به حدی زیاد شده است که اکنون می‌توانیم با آن‌ها مثل «کلان‌داده‌ها»^۶ رفتار کنیم و روش‌های علم داده^۷ را برای تحلیلشان به کار بگیریم. این چشم‌انداز به ما این امکان را می‌دهد که نقشه‌ها و جدول‌های زمانی بسیار مفصل‌تری از فرهنگ معاصر (نسبت به آنچه در مطالعات موجود درباره صنایع فرهنگ یا فهرست‌های مؤسسات فرهنگی ارائه می‌شود) تهیه کنیم. افزون بر این، می‌توانیم تحلیل کلان‌داده‌ای متن را برای همه متونی که برگزارکنندگان درباره رویدادهای خود منتشر می‌کنند، به کار ببریم؛ مانند توصیف میلیون‌ها نمایشگاه و سخنرانی و کارگاه و جشنواره و نشست‌های گروه‌های علاقه‌مند. این کار به ما این امکان

- | | |
|-----------------------------|--------------------------------|
| 1. The geography of Culture | 2. Contemporary Cultural Trend |
| 3. Growth Patterns | 4. Cultural Fields |
| 5. Public Data | 6. Cultural Events |
| 7. Big Data | 8. Data Science |

را می‌دهد که الگوهایی را در مضامین^۱ و علایق و «کلیدواژه‌ها» در عرض‌های جغرافیایی و بازه‌های زمانی طولانی جست‌وجو کنیم.

شکل ۱.۱ رشد تعداد رویدادهای فرهنگی را طی زمان نشان می‌دهد. این شکل، با استفاده از پلتفرم‌ها و شبکه‌های منتخب ترسیم شده است که از طریق آن‌ها چنین رویدادهایی ترویج یا سازمان‌دهی می‌شود. این مجموعه داده‌ها که از طریق اعلام رویدادها در شش پلتفرم و شبکه گردآوری شده است، ۴,۳۸۰,۹۴۶ رویداد را در ۲۱,۰۷۲ شهر و دو بیست کشور و شش قاره شامل می‌شود. این پلتفرم‌ها و شبکه‌ها عبارت‌اند از: بی‌هنس؛^۲ ای-فلاکس؛^۳ شبکه هنر و آموزش؛^۴ میتاپ؛^۵ رویدادهای محلی تِد؛^۵ تایم پد.^۶ (در بی‌هنس، ثبت هر حساب کاربری جدید یک رویداد به شمار می‌رود.)

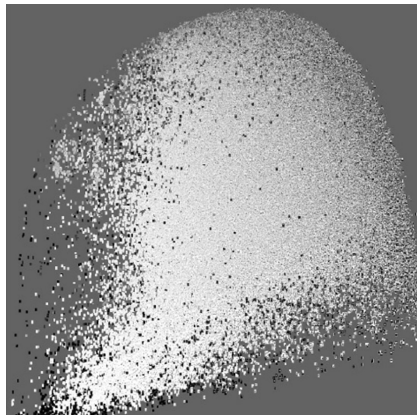


شکل ۱.۱. رشد تعداد رویدادهای فرهنگی طی زمان که به وسیله منابع داده‌ای خاص طرح «جای دیگر» به نمایش درآمده است.

- | | |
|---------------------|-------------------------------|
| 1. Themes | 2. Behance |
| 3. E-Flux | 4. Arts and Education Network |
| 5. TED Local Events | 6. TimePad |

همان طور که در نمودارها مشاهده می‌کنیم، تعداد رویدادهای فرهنگی در هر پلتفرم و شبکه با گذشت زمان در حال رشد بوده است. در صورتی که همه منابع را با هم جمع کنیم، مجموعه داده ما برای سال ۲۰۰۶، ۱۱،۶۴۲ رویداد، برای سال ۲۰۰۹، ۱۰۲،۲۱۱ رویداد و برای سال ۲۰۱۸، ۷۸۱،۶۹۷ رویداد دارد. البته در پلتفرم جهانی، واحدی وجود دارد که تمام رویدادهای فرهنگی را فهرست می‌کند و ما باید در تعمیم نتایج به دست آمده از منابع خاص خود، احتیاط کنیم. این طرح، دشواری‌های استفاده از پدیده‌های دیجیتالی را برای یادگیری دیگر پدیده‌ها با ذکر مثال نشان می‌دهد. آیا رشدی که با گذشت زمان در تمام شش منبع داده‌ای مشاهده می‌کنیم، رشد واقعی رویدادهای فرهنگی را در بسیاری از کشورها بازنمایی می‌کند؟ یا فقط نشانه‌ای از انتشار خودبه‌خود فرهنگ دیجیتال است که نشان می‌دهد سازمان‌های بیشتری در کشورهای متعدد، به تدریج، از پلتفرم‌ها و شبکه‌هایی استفاده می‌کنند که ما برای ترویج رویدادهایشان انتخاب کرده‌ایم؟ آیا ما اثر «برنده همه را برمی‌دارد»^۱ را که طی آن پلتفرم‌های اطلاعاتی خاصی غالب می‌شوند و همه شروع به استفاده از آن‌ها می‌کنند، به آسانی می‌بینیم؟ احتمال دارد تمام این تأثیرها در داده‌های ما بازنمایی شود. بخشی از رشدهایی که ما طی زمان مشاهده می‌کنیم، ناشی از محبوبیت فزاینده خود پلتفرم‌هاست. رشد دیگر نیز افزایش واقعی تعداد اماکن فرهنگی^۲ و کنشگران و رویدادها را منعکس می‌کند. «جای دیگر» از داده‌ها و موقعیت‌ها و مقوله‌بندی‌ها و محتوای متن اعلان رویدادها استفاده می‌کند. در مثال دوم که می‌خواهم بیان کنم، مصنوعات واقعی رسانه^۳ را با استفاده از علم داده و مصورسازی بررسی می‌کنیم. هدف این است که با نگرستن به آثار بسیار بیشتری (در مقایسه با آثاری که مطالعات رسانه‌ای به طور معمول می‌آزمایند) عرصه‌های فرهنگی متنوع معاصر را مطالعه کنیم. طرح یادشده

شامل یک میلیون صفحه مانگا است که جرمی داگلاس و ویلیام هوبر و من در سال ۲۰۰۹ کار روی آن را شروع کردیم. شکل ۲.I یکی از مصورسازی‌هایی را نشان می‌دهد که ما هنگام مطالعه مجموعه داده‌ای از ۱,۰۷۴,۷۹۰ صفحه منحصربه‌فرد از ۸۸۳ مجموعه مانگا ساختیم. این صفحات از محبوب‌ترین وبگاه طرف‌داران مانگا در آن زمان، یعنی وان مانگا^۲، به دست می‌آید. این وبگاه شامل بیشترین صفحات این مجموعه‌ها بود که هواداران، آن‌ها را اسکن و به زبان‌های مختلف ترجمه کردند. طولانی‌ترین مجموعه در حال پخش مانگا که در وبگاه وان مانگا در دسترس است، از سال ۱۹۷۶ به‌طور پیوسته انتشار یافته است. محبوب‌ترین مجموعه این وبگاه، ناروتو^۳ (دارای ۸,۸۳۵ صفحه برای دوره ۱۹۹۹-۲۰۰۹) و وان پیس^۴ (دارای ۱۰,۵۶۲ صفحه برای دوره ۱۹۹۷-۲۰۰۹) بود. در کنار این مجموعه‌های طولانی مانگا، مجموعه داده‌های ما شامل مجموعه‌های کوتاه‌تری نیز هستند که در دهه ۲۰۰۰ شروع و فقط به مدت یک تا سه سال پخش شدند.



شکل ۲.I. مصورسازی ۱,۰۷۴,۷۹۰ صفحه مانگا

که به وسیله دو مشخصه بصری به صورت الگوریتمی در هر صفحه محاسبه شدند: انحراف استاندارد مقدار سطح خاکستری (محور X) و آنتروپی مقدار سطح خاکستری (محور Y).

برای سنجش هر صفحه، از نرم‌افزار تحلیل تصویر خودمان بهره گرفتیم و مشخصات بصری^۱ منتخب آن را به ویژگی‌های عددی^۲ تبدیل کردیم. مصورسازی، صفحات را مطابق دو ویژگی روی محورهای x و y ترسیم می‌کند. محور x ، انحراف استاندارد^۳ «مقادیر سطح خاکستری پیکسل‌ها»^۴ را بازنمایی می‌کند که در هر صفحه اندازه‌گیری می‌شود. محور y ، آنتروپی^۵ همهٔ مقادیر سطح خاکستری^۶ پیکسل‌ها^۷ را بازنمایی می‌کند که آن نیز در هر صفحه اندازه‌گیری می‌شود.

این اندازه‌گیری‌ها در عمل به چه معناست؟ این صفحات بیشترین گرافیک را در بخش پایینی مصورسازی دارند و دارای کمترین میزان جزئیات‌اند. صفحاتی که در بخش بالایی سمت راست قرار دارند، دارای جزئیات و بافت زیادی هستند. صفحاتی هم که بیشترین کنتراست^۸ را دارند در سمت راست قرار گرفته‌اند و همچنین صفحات دارای کمترین کنتراست در سمت چپ قرار دارند.

بین این چهار حالت بی‌نهایت، هرگونه تغییر سبکی^۹ ممکن را می‌یابیم. توجه به این موضوع، این نکته را به ذهن متبادر کرد که وقتی مجموعهٔ کلان داده‌های فرهنگی مطرح است، شاید مفهوم بنیادینی که از «سبک» در ذهن داریم، مناسب نباشد. طبق فرض این مفهوم، ما می‌توانیم مجموعه‌ای از مصنوعات فرهنگی را به چند مقولهٔ مجزا تقسیم کنیم. دربارهٔ مجموعه داده‌های یک میلیون صفحهٔ مانگا، عملاً تغییرات گرافیکی نامحدودی پیدا می‌کنیم که اگر سعی کنیم این فضا را به

1. Visual Characteristics 2. Numerical Features 3. Standard Deviation

4. Pixels' Grayscale Values 5. Entropy

۶. سطح خاکستری (Grayscale) به تصویری گفته می‌شود که در آن تنها از طیف رنگی خاکستری استفاده شده است و شدت نور هر پیکسل آن می‌تواند از مشکی تا سفید تغییر کند. در تصاویر خاکستری، برای نمایش عددی، شدت نور هر پیکسل می‌تواند با مقداری بین یک کران پایین و بالا مشخص شود. م.

۷. منظور از پیکسل که از «Picture Element» گرفته شده است، نقطه‌های بسیار ریز مربع‌شکلی است که از تجمع آن‌ها، تصویر روی صفحه نمایش یا کاغذ توسط چاپگر شکل می‌گیرد. م.

8. Contrast

9. Stylistic Variation

مقوله‌های سبکی مجزا تقسیم کنیم، هرگونه تلاشی در این زمینه اختیاری خواهد بود. همچنین مصورسازی نشان می‌دهد که هنرمندان مانگا بیشتر از کدام انتخاب‌های گرافیکی استفاده می‌کنند (بخش مرکزی «ابر» صفحات) و کدام یک به ندرت ظاهر می‌شوند (بخش‌های پایین و چپ). می‌توانیم پیرسیم چرا مانگا از نظر بصری به روش‌های خاصی تطور یافته است؛ به طوری که بعضی از انتخاب‌ها به تواتر استفاده شده، برخی دیگر کمتر تکرار شده و بعضی تقریباً هرگز به کار نرفته‌اند.

اگر می‌خواهیم اصالت بصری^۲ یک مجموعه مانگای جدید را درک کنیم (همان گونه که با ویژگی‌های بصری منتخب و قابل اندازه‌گیری بازنمایی می‌شود)، می‌توانیم صفحات آن را به چنین مصورسازی‌ای بیفزاییم و اگر دوست داشته باشیم، حتی می‌توانیم این اصالت را کمی سازی^۳ کنیم.

بعدتر در این کتاب، تحلیل دیگری از این مجموعه داده ارائه می‌کنم؛ به این صورت که حین تحلیل، به ارتباط بین سبک‌های بصری^۴ و جنسیت مخاطبان مانگا و ژانرهای مانگا توجه می‌کنم. نک: شکل‌های ۱-۷ و ۲-۷ و صفحه ۹.

تحلیل کلان داده‌های فرهنگی؛ پنج ایده

در پاییز ۲۰۰۵، ابتدا شروع کردم به تفکر درباره الگوهای تحلیل و مصورسازی فرهنگ دیجیتال معاصر در مقیاسی وسیع. در آن زمان، نمونه‌هایی از تحلیل مجموعه وبگاه‌ها و وبلاگ‌ها وجود داشت که پژوهشگران علم رایانه انجام داده بودند؛ اما علم رایانه حوزه‌ای گسترده با ده‌ها زیرشاخه است. یافته‌های این حوزه پژوهشی در مجلات و همایش‌های مربوط به رشته‌های گوناگون ارائه می‌شد و

1. The "Cloud" of Pages

2. Visual Originality

3. Quantify

4. Visual Styles

نام منحصر به فردی نداشت. در علوم انسانی، اصطلاح «علوم انسانی دیجیتال»^۱ اولین بار در سال ۲۰۰۳ تعریف و از آن زمان به بعد کم‌کم شناخته شد. اما در اینجا نیز تحلیل محاسباتی مجموعه داده‌های فرهنگی، فقط بخشی از آن حوزه بزرگ‌تر و در حال ظهور بود که شامل دیجیتالی‌کردن و انتشار مجموعه‌های تاریخی و استفاده از ابزار دیجیتال در آموزش و دیگر فعالیت‌ها نیز می‌شد. اصطلاح «علوم انسانی دیجیتال» برای من محدودیت دیگری نیز در بر داشت: پژوهشگران علوم انسانی دیجیتال در مقایسه با انواع دیگر رسانه‌ها یا فرهنگ دیجیتال معاصر، تقریباً به‌طور انحصاری با مجموعه داده‌های متون ادبی تاریخی^۲ کار می‌کردند؛ بنابراین، این اصطلاح هم بسیار گسترده و هم بسیار محدود بود. من از قبل، با چند پروژه هنری و طراحی بسیار تأثیرگذار که مجموعه داده‌های فرهنگی زیادی را مصورسازی می‌کردند، آشنایی داشتم؛ اما این کار خارج از گروه‌های دانشگاهی انجام می‌شد و اصطلاح واحدی نیز برای توصیف آن وجود نداشت.

به نظر من، کارهایی که در زمینه تحلیل و مصورسازی داده‌های فرهنگی در مقیاس بزرگ و در بعضی حوزه‌ها انجام می‌شد، زمینه‌ساز ایجاد یک پارادایم تحقیقاتی جدید بود که البته هنوز نام خاصی نداشت. من احساس کردم به اصطلاحی نیاز داریم که بتواند به تحلیل محاسباتی الگوها و روندهای «فرهنگ دیجیتال معاصر» اشاره کند (در مقابل فرهنگ صرفاً تاریخی) و تحلیل انواع رسانه‌ها را پوشش دهد (در مقابل متن صرفاً نوشتاری). هنگامی که در بهار ۲۰۰۷ آزمایشگاه خود را برای انجام این تحقیق تأسیس می‌کردیم، (۵) اصطلاح «تحلیل کلان‌داده‌های فرهنگی» به ذهنم خطور کرد.

آزمایشگاه ما دو هدف را دنبال می‌کرد. هدف اول عملی بود: استفاده از روش‌های علم رایانه و مصورسازی داده‌ها و هنر رسانه برای کشف و تحلیل