

# Bibliotecas y archivos digitales con Greenstone



Las Humanidades Digitales  
**desde Argentina**

Tecnologías, Culturas, Saberes

Mariana Pichinini

[mariana.pichinini@gmail.com](mailto:mariana.pichinini@gmail.com)

Universidad Nacional de La Plata



ISBN, 978-987-3617-39-8

- ✗ Organización de **recursos y servicios**
- ✗ Una proporción significativa de los recursos de información se encuentran en formato **digital**
- ✗ Entre los servicios tenemos el harvesting, el uso de gestores de referencias, el uso social, información de valor agregado
- ✗ Involucra un **sistema de gestión** monolítico o por componentes

## X Público actual

### X Servicios

X Acceso primario por buscadores

X Navegación secundaria del Portal de

la bd

X Harvesters y servicios de búsqueda

## X Público futuro

X Preservación de los objetos digitales

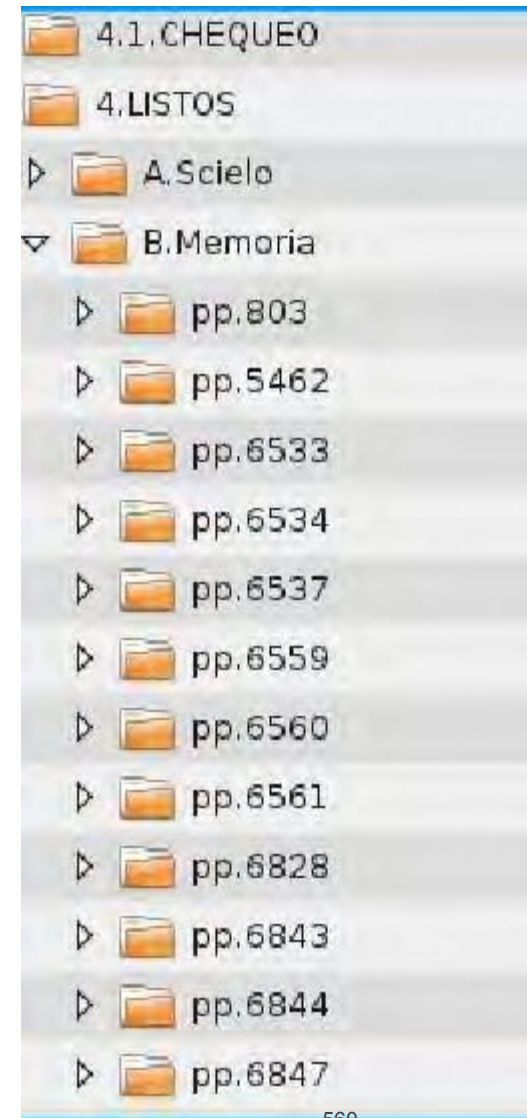
X Estandarización de los resultados

- Ingesta de recursos digitales
- Procesamiento y registro
- Almacenamiento y preservación
- Publicación y acceso

- Ingesta de recursos digitales
  - Organización de la tipología de recursos
  - Copyright
  - Digitalización / OCR
  - Autoarchivo / Gestión administrativa
  - Migraciones/emulaciones/encapsulado
- Procesamiento y registro
- Almacenamiento y preservación
- Publicación y acceso

- Ingesta de recursos digitales
- Procesamiento y registro
  - Unificado / distribuido / centralizado
  - Flujo de trabajo
  - Metadatos (DC, METS, MIX, PREMIS, EAD, MODS, y más y más)
  - PDF..... HTML, MP3, JPG, AVI
- Almacenamiento y preservación
- Publicación y acceso

- Ingesta de recursos digitales
- Procesamiento y registro
- Almacenamiento y preservación
  - Estructura de carpetas
  - Identificación unívoca
  - Copias de resguardo y refresco
  - Archivos master, divulgación,
    - originales
- Publicación y acceso



# Qué es Greenstone?

- Universidad de Waikato, Nueva Zelanda
- Software libre
- Multilingüe (+ de 60 lenguas) y multiplataforma
- Comunidad de usuarios en varios idiomas
- Versión estable 2.86 – Versión de desarrollo 3.05
- Lista de usuarios en español [greenstone\\_es@freelists.org](mailto:greenstone_es@freelists.org)
- Centro Nacional de Promoción de Greenstone, <http://cng.fahce.unlp.edu.ar>.  
Reuniones nacionales, Cursos introductorios y avanzados



- Servidor local que permite visualizar la biblioteca digital en un navegador web
- Interfaz del bibliotecario, opción gráfica para gestionar el sistema
- Gestor de esquemas de metadatos

- ✓ Agrupación de los objetos digitales en **colecciones**
- ✓ Para cada colección se define:
  - Conectores (**Plugins**) para documentos textuales, video, sonido e imagen
  - Esquema de **metadatos** para descripción de los objetos digitales
  - Navegación hipertextual del contenido de la colección (**clasificadores**)
  - Índices de **búsqueda**



Jornadas Nacionales de Humanidades Digitales  
Culturas, Tecnologías, Saberes



# Scrapping Visual

(Re)leyendo la imagen digital  
[tallerscrapping.tumblr.com](http://tallerscrapping.tumblr.com)

Gino Cingolani Trucco (UBA)  
[ginocingolani@gmail.com](mailto:ginocingolani@gmail.com)

Martín Rodríguez Kedikian  
(UBA)  
[martinrked@gmail.com](mailto:martinrked@gmail.com)

Gonzalo Vaccari (UBA)  
[vaccarigonzalo@gmail.com](mailto:vaccarigonzalo@gmail.com)  
ISBN 978-987-3617-89-8

Julio Alonso (UBA)  
[alonso.julito@gmail.com](mailto:alonso.julito@gmail.com)

# Objetivo del taller

En este taller trataremos diferentes herramientas para la extracción, procesamiento y visualización de grandes cantidades de archivos visuales.

Trabajaremos con fotos en Instagram, análisis sincrónicos de tapas de revistas, recolección y análisis del contenido de las grandes plataformas de video, y diferentes programas de visualización de información disponibles públicamente en la Web, desde el marco de las Humanidades Digitales.

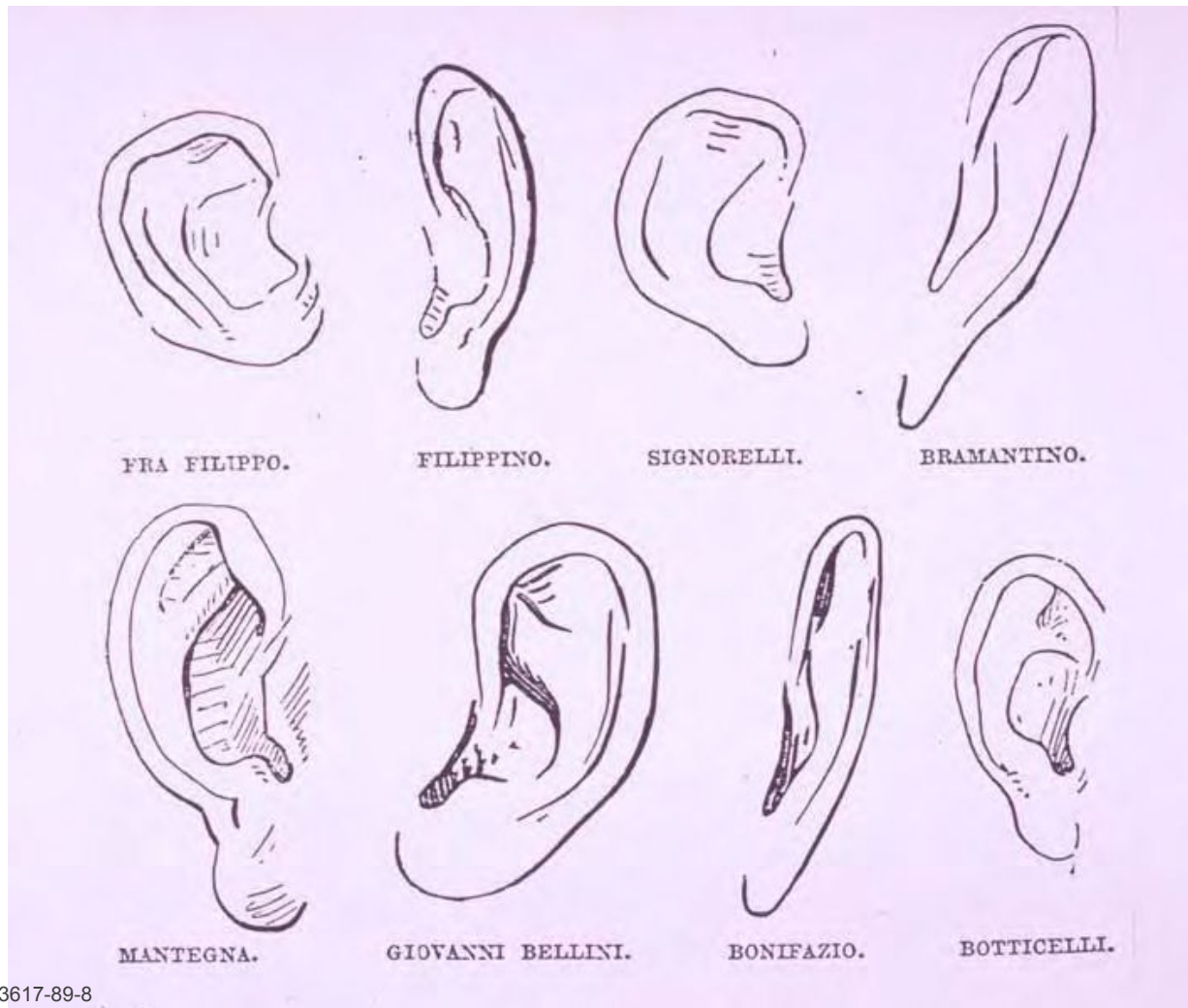


# Introducción al procesamiento digital de imágenes

Revistas de Humanidades Digitales | 2014

Construir un sistema formal de análisis alrededor de los elementos presentes y las posibilidades técnicas.

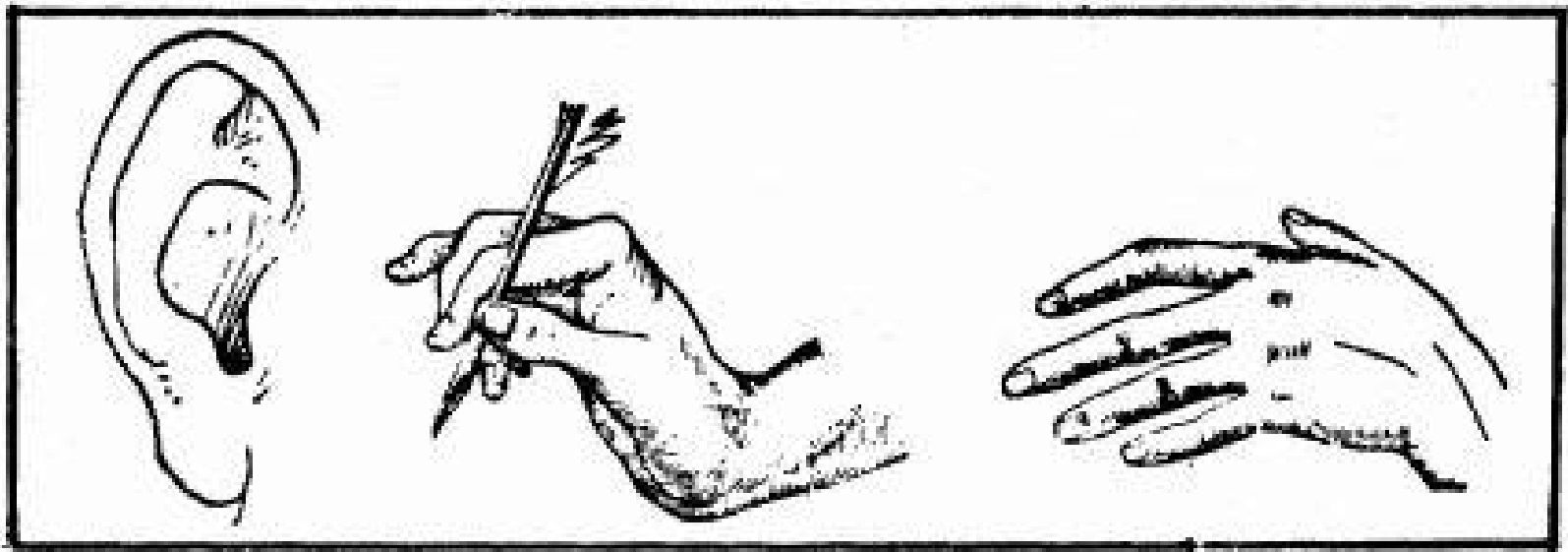
Cambiar la forma de mirar los mismos elementos de siempre. *Giovanni Morelli (1875)*



# Introducción al procesamiento digital de imágenes

La mimesis como espíritu de época más que obra aislada de un autor.

Morelli, Conan Doyle, Freud: de la medicina a las humanidades.



Ears and hands by Botticelli, from Morelli's *Italian painters* (1892).

# Aspectos técnicos de la conformación de la imagen digital

¿Qué es la imagen **digital**?

Representación numérica de una imagen analógica (continua)

Matriz de coordenadas en X e Y con valores asociados a cada uno de los puntos (Pixels. De 0 a 255)

Modelos de **Color**

Hay mucho, pero vamos a trabajar con los dos más comunes:

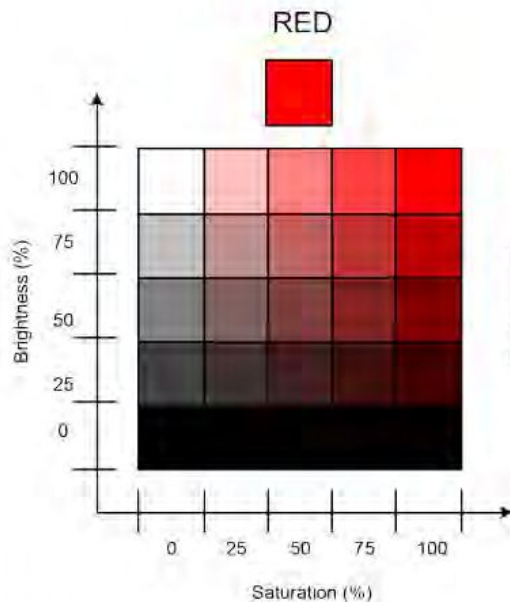
- Monocromo
- RGB (Con el sistema HSB + C)



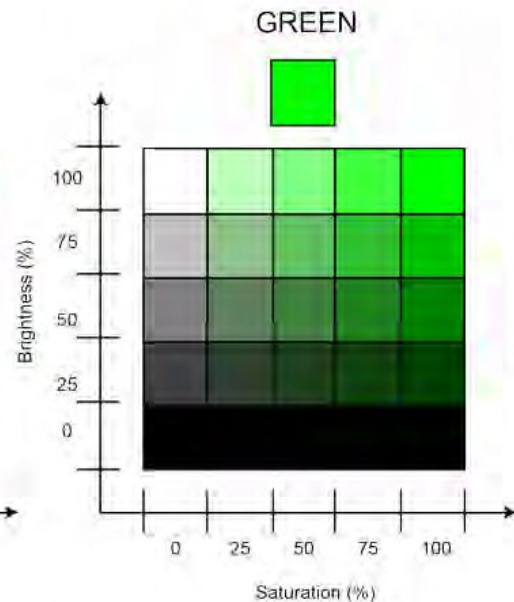
*Russel Kirsh, 1957*

# Aspectos técnicos de la conformación de la imagen digital

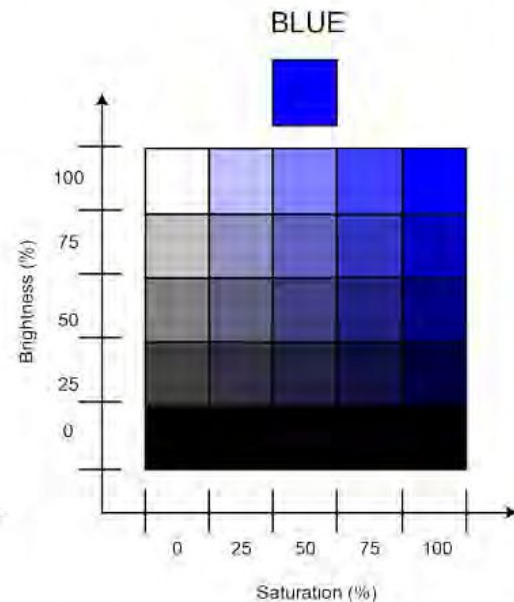
## The RGB Primary Colors - Hue, Saturation, and Brightness Variations



HSB(0,100,100)  
RGB (255,0,0)  
CMYK(0,99.4,100,0)  
HEX#FF0000



HSB(120,100,100)  
RGB (0,255,0)  
CMYK(62.76,0,100,0)  
HEX#00FF00

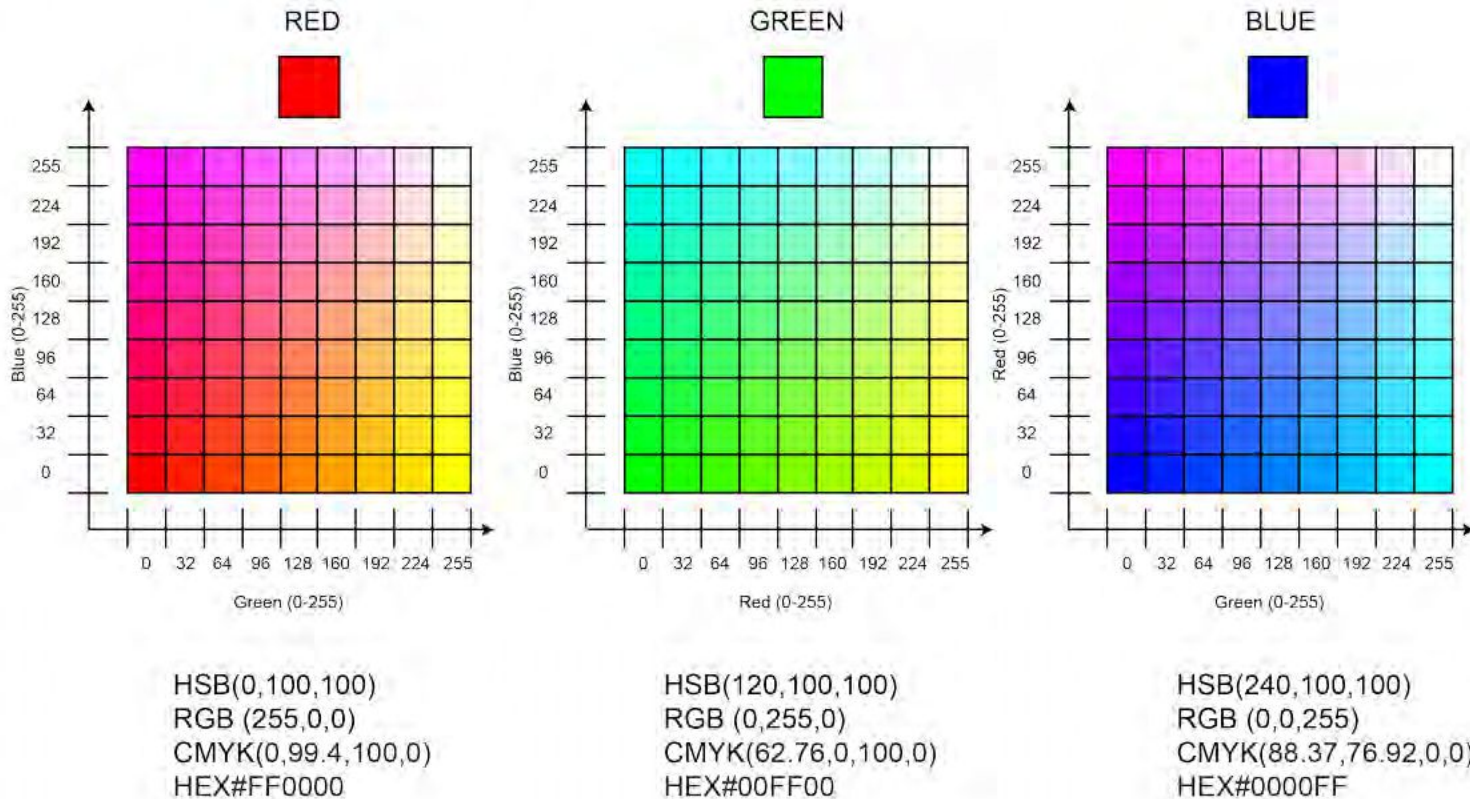


HSB(240,100,100)  
RGB (0,0,255)  
CMYK(88.37,76.92,0,0)  
HEX#0000FF



# Aspectos técnicos de la conformación de la imagen digital

## The RGB Primary Colors - Red, Green, and Blue Variations



# Herramientas y posibilidades

Journal of Digital Humanities 1 (2014)

Más allá del software que utilicemos el proceso tiene 5 pasos:

- Construcción de la base de datos
- Extracción de datos
- Procesamiento
- Visualización
- Interpretación

Construcción de la base de datos

Ingenio, trabajo manual y un poco de ayuda de código.

Scrapping



+



# Extracción de Datos

Convertir imágenes en tablas de valores

- QTip
- ImageMeasure

## Procesamiento

Agregar o modificar la estructura de las tablas

- Notepad++ / Excel (Windows)
- SublimeText o TextMate (Mac)

```
base64.cc
31 void base64_encode(const uint8_t * data, size_t len, char * dst)
32 {
33     size_t src_idx = 0;
34     size_t dst_idx = 0;
35     for (; (src_idx + 2) < len; src_idx += 3, dst_idx += 4)
36     {
37         uint8_t s0 = data[src_idx];
38         uint8_t s1 = data[src_idx + 1];
39         uint8_t s2 = data[src_idx + 2];
40
41         dst[dst_idx + 0] = charset[((s0 & 0xfc) >> 2)];
42         dst[dst_idx + 1] = charset[((s0 & 0x03) << 4) | ((s1 & 0xf0) >> 4)];
43         dst[dst_idx + 2] = charset[((s1 & 0x0f) << 2) | (s2 & 0xc0) >> 6];
44         dst[dst_idx + 3] = charset[(s2 & 0x3f)];
45     }
46
47     if (src_idx < len)
48     {
49         uint8_t s0 = data[src_idx];
50         uint8_t s1 = (src_idx + 1 < len) ? data[src_idx + 1] : 0;
51
52         dst[dst_idx++] = charset[(s0 & 0xfc) >> 2];
53         dst[dst_idx++] = charset[((s0 & 0x03) << 4) | ((s1 & 0xf0) >> 4)];
54         if (src_idx + 1 < len)
55             dst[dst_idx++] = charset[((s1 & 0x0f) << 2)];
56     }
57 }
filehippo.com
Line 31, Column 52
```

client	querytime	mark	deviceplatform	devicema	devicemodel	state	country	querydwelltime	sessionid
8	18:54:20	en-US	Android	Samsung	SCH-I500	California	United States	13.9204007	0
23	19:19:44	en-US	Android	HTC	Incredible	Pennsylvania	United States	0	0
23	19:19:46	en-US	Android	HTC	Incredible	Pennsylvania	United States	1.4757422	0
23	19:19:47	en-US	Android	HTC	Incredible	Pennsylvania	United States	0.245968	0
28	01:37:50	en-US	Android	Motorola	DroidX	Colorado	United States	20.3095339	1
28	00:53:31	en-US	Android	Motorola	DroidX	Colorado	United States	16.2981668	0
28	00:53:50	en-US	Android	Motorola	DroidX	Colorado	United States	1.7715228	0
28	16:44:21	en-US	Android	Motorola	DroidX	Utah	United States	11.6755987	2
28	16:43:41	en-US	Android	Motorola	DroidX	Utah	United States	36.9446892	2
28	01:37:19	en-US	Android	Motorola	DroidX	Colorado	United States	28.9811416	1
30	17:19:36	en-US	RIM OS	RIM	9650	Massachusetts	United States	3468.538966	0
30	17:17:18	en-US	RIM OS	RIM	9650	Massachusetts	United States	66.8533378	0
30	17:16:40	en-US	RIM OS	RIM	9650	Massachusetts	United States	0	0
43	00:44:46	en-US	RIM OS	RIM	9330	Massachusetts	United States	2.3198876	0
43	00:44:41	en-US	RIM OS	RIM	9330	Massachusetts	United States	0	0
45	21:24:03	en-US	Android	Samsung	SCH-I500	California	United States	1.7547729	1
45	21:09:43	en-US	Android	Samsung	SCH-I500	Illinois	United States	857.1453275	1
45	20:01:50	en-US	Android	Samsung	SCH-I500	New Jersey	United States	12.4195326	0
49	03:05:50	en-US	Android	LG	VS740	New York	United States	0	0
59	01:23:42	en-US	Android	LG	VS660	Nevada	United States	0.4996229	0
59	01:23:42	en-US	Android	LG	VS660	Nevada	United States	1.1773128	0
59	01:23:42	en-US	Android	LG	VS660	Nevada	United States	7.7791862	0
62	03:07:36	en-US	Android	LG	VS910	California	United States	39.4991038	0
62	03:08:15	en-US	Android	LG	VS910	California	United States	2.3677288	0
62	03:08:17	en-US	Android	LG	VS910	California	United States	2.5171807	0
62	03:08:20	en-US	Android	LG	VS910	California	United States	2.7081732	0

Detectar **patrones**.

Hacer emerger  
nuevas **preguntas**.

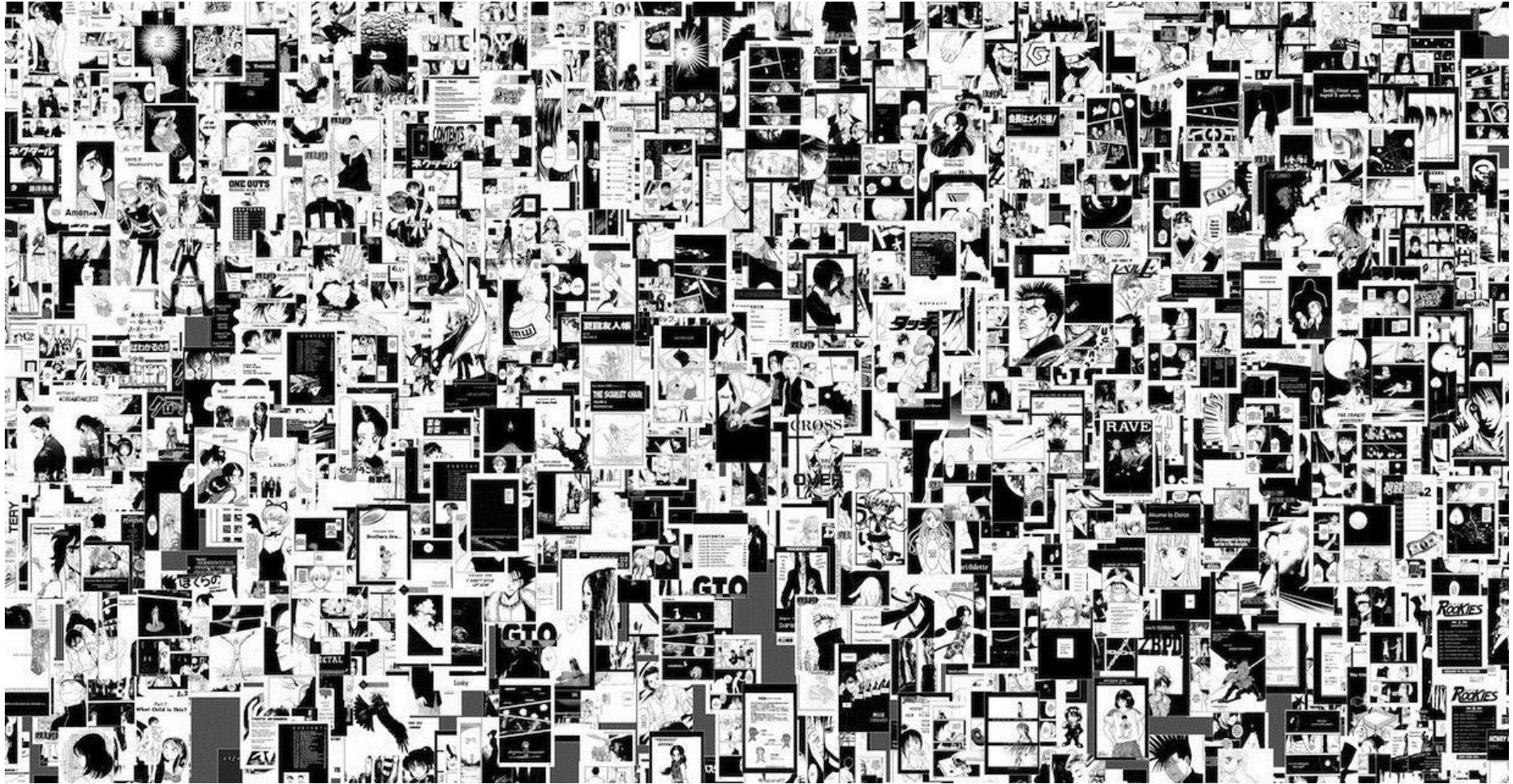
- *Viaje de egresados a Bariloche.*
- *6 cámaras distintas.*
- *Ordenadas por saturación (hacia el centro) y matiz (en vertical).*





# Investigaciones en el campo

Jornadas de Humanidades Digitales | 2014



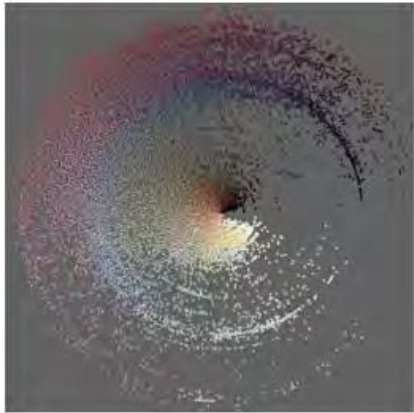
*Manovich, Lev. How to compare one million mangas pages?*

I Jornadas de Humanidades Digitales | 2014

# Proyecto Phototrails

## Manovich, la analítica cultural y las ciudades Instagram

New York



50,000 Instagram photos from NYC, organized by hue mean (radius) and brightness mean (perimeter).

San Francisco



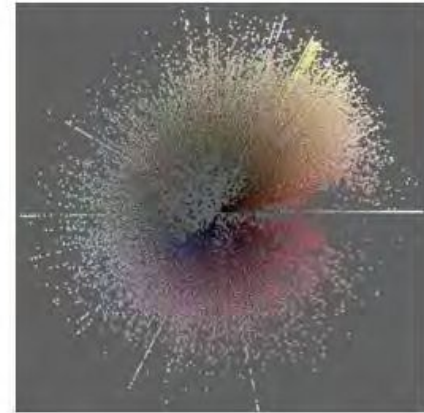
50,000 Instagram photos from San Francisco, organized by brightness median (radius) and hue median (perimeter).

Tokyo



50,000 Instagram photos from Tokyo, organized by brightness mean (radius) and hue mean (perimeter).

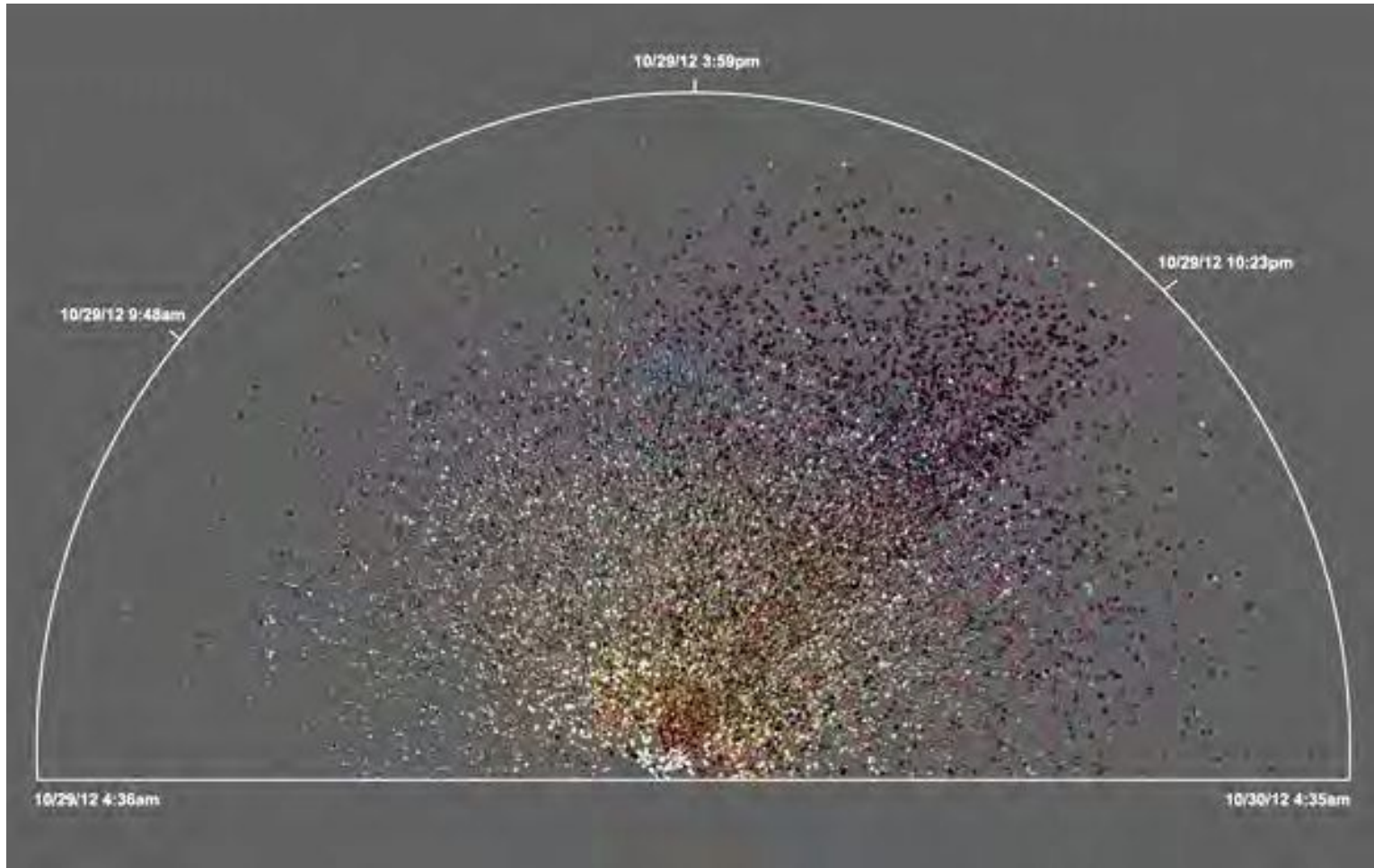
Bangkok



50,000 Instagram photos from Bangkok, organized by brightness mean (radius) and hue median (perimeter).

Las ciudades hablan, y los usuarios de Instagram reflejan sus actividades y su percepción de la ciudad. Ordenadas por algún criterio, **¿qué historias cuentan?**

## Manovich, la analítica cultural y las ciudades Instagram

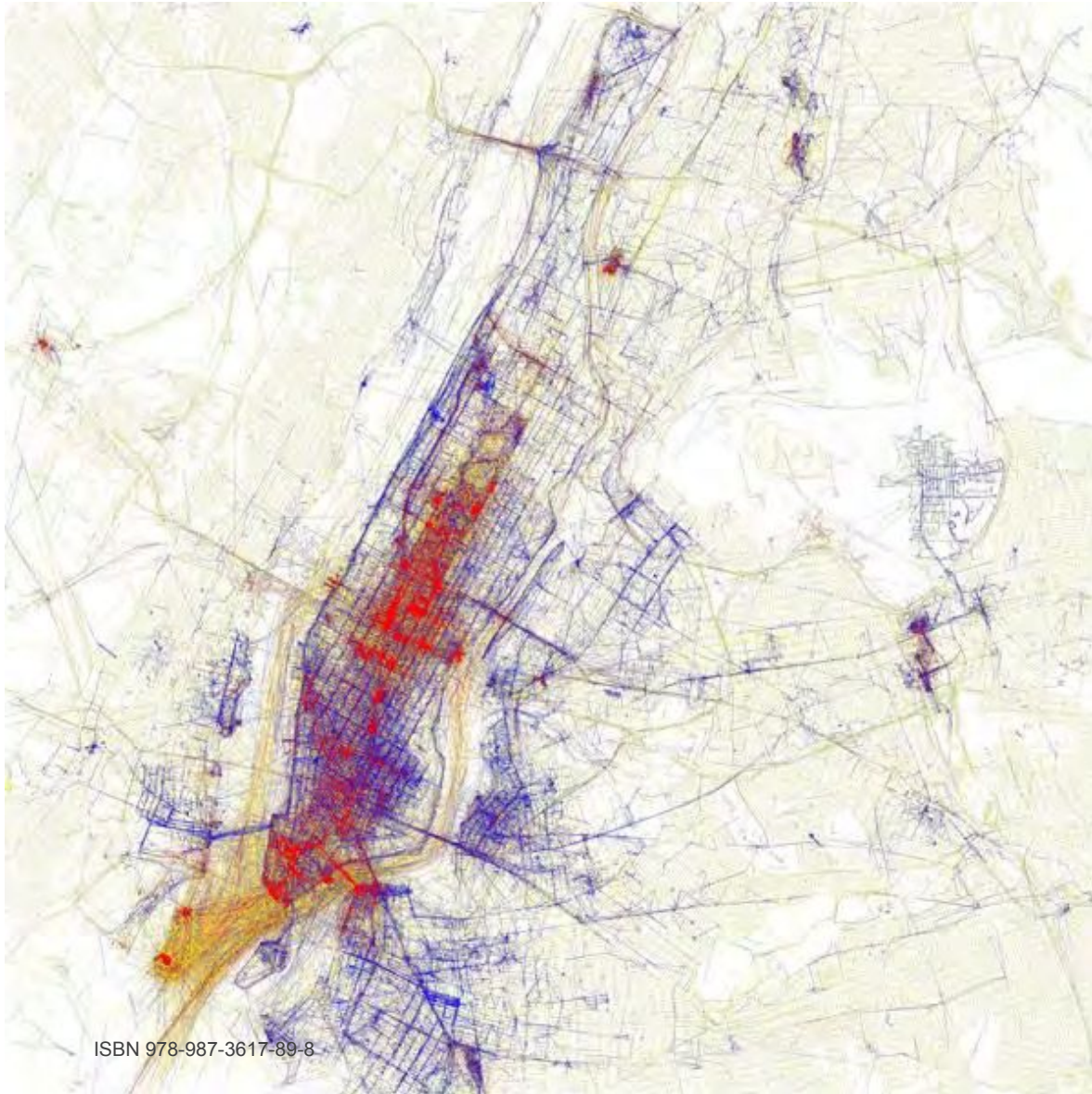


Mapeo de subidas de fotos a Instagram en Nueva York al momento del paso del huracán Sandy.



# Local and Tourists

Eric Fisher y un proyecto sobre quién saca fotos de/en las ciudades



ISBN 978-987-3617-89-8

Azul: locales  
Rojos: turistas  
Amarillos: indeterminado

[Eric Fisher en Flickr](#)