.UBAsociales
FACULTAD DE CIENCIAS SOCIALES

Repositorio Digital Institucional

Tipo de documento: Tesis de Grado de Ciencias de la Comunicación

Título del documento: Datos en comunicación: un abordaje desde las humanidades digitales

Autores (en el caso de tesistas y directores):

Julio Alberto Alonso

Estela María Dominguez Halpern, tutora

Datos de edición (fecha, editorial, lugar,

fecha de defensa para el caso de tesis): 2014

Documento disponible para su consulta y descarga en el Repositorio Digital Institucional de la Facultad de Ciencias Sociales de la Universidad de Buenos Aires.
Para más información consulte: http://repositorio.sociales.uba.ar/

Esta obra está bajo una licencia Creative Commons Argentina.
Atribución-No comercial-Sin obras derivadas 4.0 (CC BY 4.0 AR)
La imagen se puede sacar de aca: https://creativecommons.org/choose/?lang=es_AR

Repositorio Digital Institucional

.UBAsociales
FACULTAD DE CIENCIAS SOCIALES

UNIVERSIDAD DE BUENOS AIRES

FACULTAD DE CIENCIAS SOCIALES

CARRERA DE CIENCIAS DE LA COMUNICACIÓN

LICENCIATURA DE GRADO

DATOS EN COMUNICACIÓN:

UN ABORDAJE DESDE LAS HUMANIDADES DIGITALES

Mayo 2013

Nombre: ALONSO, Julio Alberto

DNI: 33.554.660

Correo: alonso.julito@gmail.com

Teléfono: 15 - 5864 - 9066

--

Tutora: DOMINGUEZ HALPERN, Estela María

Teléfono: 15 - 5325 – 7291

Correo: esteladomiha@gmail.com

DATOS EN COMUNICACIÓN:

UN ABORDAJE DESDE LAS HUMANIDADES DIGITALES

A mi familia, por estar siempre.

A la cátedra Datos, por el espacio.

A mi tutora, por la confianza y magia.

A la UBA y a Sociales, por el aprendizaje.

ÍNDICE

0. Abstract 6

1. Introducción 7

2. ¿Quiénes hablan?

2.1 Constelaciones 10

2.2 Ver pero no observar: lectura cercana y lectura distante 10

2.3 Avalancha de números: Big Data 11

2.4 Epistemología del software: social media y la producción cultural 13

3. ¿Que son las humanidades digitales?

3.1 Trazando límites 16

3.2 Las tres olas de las Humanidades Digitales 16

3.3 ¿Humanidades Digitales en un renglón? 18

3.4 Proyectos enmarcados en las Humanidades Digitales 19

3.4.1 ¿Cómo comparar un millón de páginas? 20

3.4.2 Inexplorado o no trazado: Big Data como un lente para la cultura humana 25

3.4.3 La forma de un discurso espectacular: Una infografía analítica sobre los elementos que hicieron grande al discurso “I Have a Dream” de Martin Luther King Jr. 27

3.5 Comentarios finales sobre el apartado 31

4. Herramientas de procesamiento de datos y visualización

4.1 N-gram 32

4.1.1 ¿Qué es un N-Gram 32

4.1.2 Funcionamiento 33

4.1.3 Advertencias sobre su utilización y críticas sobre accesibilidad 34

4.2 Gapminder 36

4.2.1 Guía para uso de la herramienta 37

4.2.2 Sobre la base de datos: ventajas y desventajas 39

4.3 Many Eyes 40

4.3.1 ¿Cómo ingresar información? 41

4.3.2 Opciones de visualización 42

4.4 Voyant Tools 43

4.5 Wordle 45

5. Programas de Comunicación #FSoc UBA

5.1 La imagen que comenzó todo 46

5.2 Armar la base de datos 47

5.3 Números iniciales: de Abal-Medina a Zweig 49

5.4 Ciudad #Fsoc 52

5.5 Marx hasta en la sopa. ¿Cuál es el autor más leído de la carrera? 53

5.6 Conclusiones provisorias sobre Programas de Comunicación #FSoc UBA 57

6. Conclusiones 59

7. Bibliografía 61

8. Anexo: Bases de datos Programas de Comunicación #FSoc UBA

0. Abstract

Las tecnologías de la comunicación e información (TIC) han cobrado relevancia en la mirada académica sobre la producción cultural. Día a día, usuarios que utilizan las redes sociales a partir de sus dispositivos móviles o computadoras portátiles, generan miles y miles de datos. Actualizaciones de estado, videos caseros, fotos, ubicaciones geolocalizadas, comentarios en sitios, artículos en la web y el desarrollo de plataformas son dinámicas que se van desplegando en el cotidiano. Todas estas palabras, sentidos, opiniones bajo el formato de publicaciones, dejan una huella en la red, generando datos, información.

La dinámica que transcurre en línea no tiene tiempo. Si bien puede leerse un continuo, no todo lo que sucede en la web, se conjuga en un presente. Muchos de estos usuarios digitalizan y cargan textos, libros, imágenes, fotos, mapas, videos en servidores, poniéndolos a disposición para su acceso. Todos estos formatos son transformados en bits, conformando granos de arena del vasto mar digital, constituyendo parte del inconmensurable proceso de digitalización.

Los gobiernos digitales y organizaciones no gubernamentales han abierto sus armarios, archivo, cajones (registros, investigaciones y base de datos), sumándose a la Big Data. De esta manera, han aportado a la comunidad web información valiosa para que los usuarios accedan a consultarlos.

Sin embargo, esta situación genera nuevos problemas y preguntas: ¿Quién y cómo se lee la información? ¿Cómo procesar y estructurar los datos a fin de facilitar su abordaje?

En la presente Tesina navegaremos las aguas de las Humanidades Digitales, el incipiente campo académico que intenta comprender cómo aprovechar los fenómenos de la digitalización y producción cultural masiva en la era de la información. Utilizaremos como ejemplo el proyecto http://programascomunicacion.tumblr.com, dónde, a partir de los conceptos y herramientas vinculadas a las Humanidades Digitales, se intenta indagar en la conformación de la propuesta académica de la Carrera de Ciencias de la Comunicación Social de la Facultad de Ciencias Sociales de la Universidad de Buenos Aires.

Iremos recorriendo y revistando las preguntas que se llevaron adelante, las decisiones alrededor del diseño de una base de datos, las dificultades que se encuentra el investigador al momento de recolectar información en la web, el trabajo sobre herramientas y software de procesamiento/visualización existente para abordar este tipo de datos, y finalmente los resultados provisorios obtenidos, como así también algunas advertencias sobre el trabajo en las Humanidades Digitales.

1. Introducción

La presente Tesina se encuentra contextualizada en la denominada era o sociedad de la información, en la que según Castells la generación, el proceso, y la transmisión de información se convierten en las principales fuentes de productividad y poder (Piscitelli, 2009). De esta forma, nos encontramos en un momento donde, por un lado, la producción anual de datos superan los 8 trillones de preguntas afirmativas/negativas (Aiden y Michel, 2013); y por el otro, comienzan a desarrollarse herramientas y software que pueden procesar, analizar y visualizar todos estos datos (Manovich, 2011).

En esta convergencia, entre producción masiva de información y herramientas que permiten procesarlos, los investigadores se encuentran ante la posibilidad de expandir la escala de su investigación y disponer de facilidades que permitan acelerar procesos de abordaje al objeto de estudio, reduciendo tiempos y costos. Sin embargo, algunos cientistas sociales aún no se han acercado a intervenir en estas prácticas, viéndose desafiados por tres retos: acceso, comunicación y análisis (Suárez, 2013)

El catalizador de este trabajo es el proyecto http://programascomunicacion.tumblr.com dónde, a partir de los conceptos y herramientas vinculadas a las Humanidades Digitales, se intenta indagar cómo se construye el perfil profesional del Comunicador Social, a partir de la conformación de la propuesta[1] académica de la Carrera de Ciencias de la Comunicación Social de la Facultad de Ciencias Sociales en la Universidad de Buenos Aires. A su vez, deseamos observar qué líneas teóricas se van constituyendo en el proceso de formación a lo largo de la carrera.

Para comenzar a entender y entrever estos recorridos, nos propusimos relevar qué es lo que se lee a través de revistar la cantidad de autores que circulan por los programas de las distintas cátedras y cómo se relacionan en las distintas materias o años. En consecuencia, se procedió a armar una base de datos que contuviera el nombre de todos los autores de los textos, posibilitando el uso de herramientas de procesamientos y visualización que respondieron a nuestras preguntas y generaron nuevas. Este proyecto ha posibilitado explorar conceptos y herramientas que se inscriben dentro del plano de este incipiente campo. Por esta razón, entendemos que era necesario realizar un trabajo formal para presentar en el ámbito de la Universidad de Buenos Aires, innovando el campo metodológico, acercando una nueva herramienta de trabajo al Comunicador Social al tiempo de sumar complejidades de lectura y apertura a la Carrera que me nos ha formado, y que en este momento se encuentra revisando su

1 Respecto a propuesta entendemos a la totalidad de los autores, textos y posibles recorridos de lectura que ofrece la Carrera de Comunicación. Si tuviéramos que resumirlo en una frase sería: "esto es lo que hay para leer aquí". Si bien la Carrera da lugar a un recorrido y tipo de lectura a partir de la estructuración en tronco común, orientaciones y las distintas cátedras, no hay nada que imposibilite a un alumno recorrer todas las materias y leer todos los textos que componen los programas.

Plan de Estudios, debatiendo los alcances de sus miradas epistémicas.

Por otro lado, deseamos que el evaluador sea parte de nuestra propuesta; esta Tesina de grado está planteada como una constelación, de manera que **su lectura pueda optar por diferentes recorridos** entre los capítulos que la componen. La misma puede ser lineal, pero también puede saltarse capítulos y dirigir su mirada hacia el contenido que sea de su interés para luego, completar con el resto de los apartados.

En principio abordaremos **a los autores que nos hablan** e iluminan con sus conceptos gran parte del campo de las Humanidades Digitales. Walter Benjamin será el continente mediante el cual exploraremos el campo como **traductores**; trazando en el mismo caminos junto a las lentes de Franco Moretti, **lectura cercana** y **lectura distante.** Por otro lado, Chris Anderson dará cuenta del universo de la **Big Data** como concepto ordenador de la cantidad de información producida mundialmente; y por último Lev Manovich cocinará en nuestro campamento epistémico: nos introducirá en la **epistemología del software**, para entender cómo se produce conocimiento hoy en día a partir de las herramientas que procesan y visualizan estos datos.

Se realizará en un apartado especial, la **analítica del sentido de las Humanidades Digitales**, repasando su historia desde la aparición de las computadoras y los primeros proyectos de digitalización, haciendo foco en sus alcances y límites como marco de investigación y entendiendo que características presentan sus trabajos a partir de tres proyectos representativos:

a) el procesamiento, visualización y comparación de 1 millón de páginas de manga por la **Iniciativa de Estudios de Software de Lev Manovich**.

b) el proyecto **N-Gram Viewer de Erez Aiden y Jean-Baptiste Michel** que procesa y visualiza los 5 millones de libros digitalizados por Google.

c) en una escala menor, el procesamiento y visualización del **discurso "I have a dream" de Martin Luther King Jr. por Nancy Duarte**.

En otro capítulo, abordaremos las **herramientas de procesamiento y visualización**, que si bien son parte de las investigaciones como facilitadores en el trabajo, también lo son como objeto de estudio en sí. Entendemos que debemos incluir la observación de los distintos software, con el objeto de introducir al usuario que desee comenzar a trabajar en el marco de esta propuesta, y por otra parte, dar cuenta de los alcances que estas herramientas ofrecen como así también las advertencias y cuidados que hemos recorrido en la presente investigación, sobre sus usos. Entre los software que presentamos en este capítulo, se encuentran:

a) **N-Gram Viewer**, descrito como Proyecto en el apartado anterior, ahora en el presente, indagamos en él como herramienta de visualización y la posibilidad de acceder a sus archivos;

b) **Gapminder**, software diseñado para observar las bases de datos de organismos mundiales como la OMS o la OIT;

c) **Many Eyes**, **d)** **Voyant Tools** y **e)** **Wordle**, herramientas que nos permiten crear bases de datos personales para procesarlas, visualizarlas y compartirlas con la comunidad.

Por último nos adentraremos en el proyecto **Programas de Comunicación #FSoc UBA**. Allí nos adentraremos en el proceso de trabajo, las preguntas iniciales, el armado de la base de datos, las herramientas que hemos utilizado para abordar la información recolectada, y por último los primeros resultados y visualizaciones posibles a partir de las indagaciones sobre la propuesta académica de la Carrera de Ciencias de la Comunicación.

Por último, deseamos focalizar que nuestro **objetivo general**, será describir los alcances y limitaciones que ofrecen las Humanidades Digitales en un contexto de producción masiva de datos culturales.

En tal sentido, los **objetivos específicos** que proponemos son los siguientes:

a) caracterizar a los proyectos dentro de las Humanidades Digitales.

b) caracterizar las herramientas y software de procesamiento / visualización de datos.

c) indagar en distintos proyectos enmarcados en las Humanidades Digitales, sobre sus herramientas/software y su aprovechamiento en los estudios e investigaciones en las ciencias sociales.

d) ejemplificar la utilización de esta metodología innovadora a partir de un trabajo aplicado a la Carrera de Ciencias de la Comunicación en la Universidad de Buenos Aires, en este caso, http://programascomunicacion.tumblr.com/

e) contribuir y enriquecer a partir de una mirada analitica, sobre la formación académica en Ciencias de la Comunicación

2. ¿Quiénes hablan?

2.1 Constelaciones

Intentar establecer un marco teórico para las Humanidades Digitales no es tarea sencilla. Para poder explorar este objeto de estudio, nos apoyaremos en diversos conceptos que actuarán como compuertas . El primero a analizar será el que nos presenta Walter Benjamin: **constelaciones.** La idea que traza a través de una magnífica metáfora es analizar la complejidad y a la vez la simpleza de los sistemas; los múltiples sentidos, aquellas retículas de conexiones significativas entre elementos independientes y distantes. Al mirar el cielo, observamos que las estrellas son independientes, sin relación aparente, pero a partir de nuestro enfoque, se hace posible trazar recorridos y límites, uniendo y produciendo formas. Este modelo que Benjamin propone, nos va a permitir acercarnos a las Humanidades Digitales desde diversos puntos de vista: autores, conceptos y recorridos interactuando.

En este acercamiento, **nuestra presencia** es fundamental, dado que nosotros también formamos parte de la constelación y esto implica una relación con las distintas estrellas que nos rodean. Según donde nos situemos y pongamos la mirada, podemos ver o dar cuenta de distintos comportamientos; cada pregunta que realicemos implica un posible recorrido; cada interrogación que formulemos a los autores, abrirán distintas entradas al tema. Las herramientas de visualización, serán nuestros lentes. Su arquitectura, su software, constituirá un espacio para recopilar fenómenos.

> “Benjamin recopilaba fenómenos. Las constelaciones que se pueden generar a partir de ahí son innumerables, y tal vez cada uno de nosotros construya la suya (cada uno pone su biografía, sus intereses, sus proyectos, su vida, en una palabra, en hacer la constelación), pero lo único que Benjamin pide es que se salve el fenómeno, que siga siendo reconocible, que no quede perdido en la marea de lo ente.” (El Pez Martillo, 2007)

2.2 Ver pero no observar: Lectura cercana y lectura distante

En el texto *Conjeturas sobre la literatura mundial (2000)*, el crítico literario Franco Moretti desarrolla dos conceptos importantes para este trabajo: **lectura cercana y lectura distante.** Para el autor, la *lectura cercana* representa el método tradicional norteamericano de investigación sobre la literatura. Este enfoque modeliza regularidades, cánones, que en muchos casos ni siquiera representa el uno porciento de lo producido, dejando de lado una gran cantidad de texto. Esta mirada exclusiva, deriva en un pensamiento en el que los libros elegidos son relevantes para la literatura en general. Pero en un contexto de investigación sobre literatura mundial, Moretti

considera que como condición mínima, la idea de que el investigador, debería expandir y superar este cánon, y por ende, abandonar los métodos de la lectura cercana. En cambio, en la *lectura distante*, presenta fuertemente a **la distancia** como condición de conocimiento. Ella nos permite concentrarnos en las unidades más pequeñas o más grandes del texto; enfocamos dispositivos, soportes, temáticas, estilos, géneros y sistemas. En este proceso, el autor advierte que seguramente el texto desaparezca como unidad de análisis, pero este movimiento es justificable:

> "Si queremos entender al sistema por completo, debemos aceptar que algo debe perderse. El conocimiento teórico siempre exige algo a cambio: la realidad es enteramente rica mientras que los conceptos son abstractos, pobres. Pero precisamente esta pobreza es lo que nos permite controlarla, y por consiguiente conocer. Esta es la razón por la que *menos* es realmente *más*." (Moretti, 2000)

Moretti propone distintas maneras de abordar la literatura mundial desde una lectura distante; observa por ejemplo, los patrones de producción como así también la venta mundial de textos en determinado momento histórico. A partir de esta analítica, traza por primera vez, claros patrones del surgimiento de determinados géneros literarios.

Buceando lo señalado anteriormente, la aparición de la novela moderna en Japón, India y Brasil (pertenecientes a la periferia del sistema literario) no se da de manera autónoma, sino constituye una consecuencia del conjunto de materiales locales y la influencia occidental francesa/inglesa. Es aquí, donde se detecta el patrón:

> "Luego de revisar los registros literarios de cuatro continentes en un período de doscientos años, junto a veinte estudios sobre el tema, se concluye que el surgimiento de la novela moderna, es siempre un compromiso entre materiales extranjeros y locales." (Moretti, 2000)

Como dato final y para el asombro de este investigador, su trabajo arroja una nueva conclusión, una vez establecida la regularidad, los recorridos hasta ahora utilizados para dar esta explicación (Francia, Inglaterra, España) pasaron a ser las excepciones.

2.3 Avalancha de números: Big Data

En un artículo del 2008, Chris Anderson afirma que la inmensa cantidad de datos producida en internet dejaría al método científico obsoleto. Lo que hoy se conoce como *Big Data*, es un proceso que el autor relata a partir de los avances tecnológicos en el procesamiento de la información.

Anderson nos marca un camino de (co) evolución. Veamos; hace 60 años la información podía leerse de manera digital; hace 20 Internet permitió que esa información estuviera al alcance de los usuarios; hace 10 los primeros buscadores transformaron toda es información en una sola base de datos; y desde hace 5 años empresas como Google están trabajando sobre este corpus como condición de lo humano:

> "La edad del Petabyte es diferente porque *más* es diferente. Los Kilobytes eran almacenados en diskettes de 3 ½. Los Megabytes se almacenan en discos duros. Los terabytes en una matriz de discos. Los Petabytes en la nube[2]. A medida que nos movemos en esta progresión, fuimos de la analogía de la carpeta, pasando por el fichero hasta la gran biblioteca y luego… bueno, al llegar al petabyte nos quedamos sin analogías." (Anderson, 2008)

Para ilustrar las analogías de Anderson, los autores del libro "Uncharted: Big Data as a Lens on Human Culture" (2013), Erez Aiden y Jean-Baptiste Michel, aportan algunos números:

> "En este momento, el promedio de datos producidos por un usuario es casi de 1 terabyte. Esto equivale a casi 8 trillones de preguntas afirmativas/negativas. Como colectivo, eso significa que la población total produce casi 5 Zettabytes anuales: **40.000.000.000.000.000.000.000** (40 trillones) de bits." (Aiden y Michel, 2013:164)

Es evidente que resulta imposible poder leer o procesar toda la información que se produce diariamente. Precisamos de nuevas herramientas o lentes que nos faciliten la exploración de estas bases de datos. La Big Data es una buena alternativa, aún con sus dificultades:

> "Uno de los mayores desafíos respecto del uso de la Big Data es que está estructurado de manera diferente a lo que los científicos están típicamente acostumbrados. Ellos prefieren responder a preguntas cuidadosamente construidas usando experimentos elegantes que producen resultados precisos de manera constante. Pero en la Big Data, los datos son confusos" (Aiden y Michel, 2013:273)

Como señalábamos, el proceso está en construcción: uno de los principales problemas de la producción de información de manera masiva es su estructuración y categorización. Por eso, el uso de computadoras que permite procesar información a esta escala pueden crear grandes oportunidades para comprender nuestro objeto a la hora de investigar: a) una de ellas estaría

2 Se denomina nube al servicio de almacenamiento online o en servidores externos a la computadora personal, reemplazando de esta manera la necesidad de transportar los datos y poder acceder a ellos desde cualquier dispositivo o terminal.

dada en la posibilidad de recortar o seleccionar la información que sea pertinente para nuestro interés b) podemos dividir y ser más precisos en la segmentación o filtrado de esta información c) jugar con a) y b) … y así mixturamos y recombinamos moebianamente.

El artículo de Chris Anderson cierra precisando que no es necesario detenerse en la pregunta del ¿por qué? las personas actúan como actúan, asociado a la microsociología o, retomando a Moretti, la lectura cercana.

Nuestro análisis hará foco en entender que las personas accionan. A partir de la utilización de herramientas adecuadas, podemos rastrearlas, registrarlas y medirlas con una precisión sin precedentes: “Con la cantidad de datos suficientes, los números hablarán por sí solos” (2008)

Sin embargo, debemos recordar que el dato no es neutral, sino una construcción social que representa un nuevo lente sobre cómo abordamos un objeto, lo que implica un cuidado especial dado que podemos pasar fácilmente de lo que nuestro ojo puede ver, a lo que nuestra mente quiere que nuestro ojo vea.

> “En su núcleo, la revolución de la Big Data, trata sobre cómo los humanos crean y preservan la memoria histórica de sus actividades. Sus consecuencias transformarán la manera en que nos miramos. Permitirá la creación de nuevos ámbitos que harán posible a nuestra sociedad probar su propia naturaleza de manera más efectiva.” (Aiden y Michel, 2013:129)

2.4 Epistemología del software: social media y la producción cultural

En su último libro, Software takes command (2013), Manovich intentará trazar una epistemología del software. Allí nos presenta el cambio de paradigma en la producción de conocimiento en la web: “Si la electricidad y la combustión permitieron la aparición de la sociedad industrial, entonces de manera similar, el software dará lugar a la sociedad de la información”. (2013, 02). Convertir todo en datos y utilizar algoritmos para procesar su análisis, arroja una serie de consecuencias importantes en el campo del saber/conocer. En este devenir, se crean nuevas estrategias que en conjunto conforman la **“epistemología del software”**: El código digital, algoritmos, máquinas de aprendizaje, grandes conjuntos de datos (big data), medios sociales, procesadores con una mayor capacidad de almacenamiento y otras partes del moderno universo tecno-social, introducen *nuevas formas de adquirir conocimiento*.

Para Manovich, el estudio de lo social y cultural se basaba en dos tipos de datos:

superficiales, específicamente estadísticas de encuestas y trabajos sociológicos; y **profundos**, como la psicología, antropología o historia del arte, con métodos como la descripción densa o la lectura cercana (Moretti). Con el auge de las herramientas digitales que permiten procesar grandes cantidades de datos, el acercamiento a los procesos sociales y su estudio cambia de manera radical. Este autor considera que podemos *estudiar patrones culturales*, conformado por millones de objetos, con exactitud.

> "En otras palabras, el conocimiento y la comprensión detallada, que antes sólo se podía llegar a partir de unos pocos textos, ahora puede obtenerse sobre las colecciones masivas de estos" (Manovich, Reading Patterns, 2011)

Utilizando la web, es posible conocer a un individuo al combinar la información de sus diversos perfiles sociales. A su vez, podemos adjudicar significados adicionales a cada una de los mismos. Veamos un ejemplo: si consideramos la técnica de pegar un número de fotos independientes de manera panorámica, estrictamente hablando, los algoritmos que subyacen no agregan información nueva a cada una de las imágenes (sus píxeles no cambian). Pero cada imagen ahora pertenece a un sector de otra más grande, por ende, para una observador el significado cambia. Las habilidades para generar nueva información a partir de la vieja, fusionar distintos tipos de fuentes, y crear conocimientos de fuentes analógicas son sólo algunas de las técnicas de la "epistemología del software".

> "El registro de historias digitalizadas están haciendo posible la cuantificación del colectivo humano como nunca antes (...) la historia humana es mucho más de lo que las palabras pueden decirnos. La historia también puede ser encontrada en los mapas que dibujamos y en las esculturas que hemos creado. Está en las casas que construimos, en los campos que cuidamos y en la ropa que usamos. Está en la comida de cada día, la música que se escucha y en los dioses que creemos. Está pintada en las cuevas e incluso en los fósiles de criaturas que estuvieron antes que nosotros. Inevitablemente, la gran mayoría de este material se perderá: nuestra creatividad por lejos le gana a nuestra posibilidad de preservación. Pero hoy, podemos preservar como nunca antes." (Aiden y Michel, 2013)

En el 2005 como así también en el 2007, Manovich desarrolla e introduce el término "Cultural Analytics o Analítica Cultural". Este concepto funciona como una fuente en donde el autor podrá vitalizar y sostener su análisis sobre Big Data. A través del mismo, construirá diversos métodos que permitan visualizar bases de datos a partir de tecnología computarizada[3]. Deseando

3 En su sitio http://lab.softwarestudies.com/ se encuentran disponibles todos los proyectos, la teoría y la práctica sobre

precision:

> “Es la aplicación de una masiva escala de datos y su posterior análisis para el estudio de la cultura humana ”(Aiden y Michel, 2013)

El desarrollo de este tipo de software social, tiene como objetivo ser de uso simple para aquellos usuarios que deseen realizar análisis y luego aplicarlos a sus bases de datos, o incluso aportar a otras existentes.

El gran desafío final es el **“cambio de domicilio de la información”** (Aiden y Michel, 2013). Los cientistas sociales recuperan datos a partir de experimentos en laboratorios o salen al exterior a desarrollar observaciones. La obtención de los datos está dentro de una metodología de investigación que ha sido regular y sistemática en la tradición académica. En el mundo de la Big Data, las grandes corporaciones e incluso los gobiernos, son usualmente los guardianes (gatekeepers) de los mayores conjuntos de información. Nuestra propuesta es crear espacios y contextos donde esta información sea accesible, transparente, de fácil aplicación para los usuarios o investigadores. Exploramos la posibilidad de buscar y compartir herramientas que permitan rescatar esta información, trabajando en conjunto con diseñadores y programadores. Nos aventuramos a desarrollar un software que permita alcanzar estos objetivos.

analítica cultural en base al procesamiento de imágenes y su posterior visualización

3. ¿Qué son las Humanidades Digitales?

En este apartado haremos un recorrido para observar que características presentan las humanidades digitales. Su historia, alcances y límite, características de los trabajos como asi también, algunos ejemplos concretos de proyectos enmarcados en este campo.

3.1 Trazando límites

Intentar trazar los límite de las Humanidades Digitales (HD) exige tener en cuenta su carácter histórico. En un corte diacrónico, pareciera las HD son objeto de discusión desde los años ‘40; pero a la vez, en cada uno de esos papers, artículos o libros donde se discute el tema, se lo presenta como un fenómeno actual e innovador. La amplitud de opiniones y discusiones sobre sus alcances y limitaciones está dado justamente por la apertura que este incipiente campo viene generando, en los estudios de medios emergentes y en las Ciencias Sociales.

En el año 2009, la Universidad de California realizó un seminario sobre “Digital Humanities and Media Studies” que tuvo como resultado la producción de un paper colaborativo que da cuenta las promesas y/o beneficios de este campo. En el mismo, se realizó una aproximación que nos deja entrever la amplitud de prácticas que abarcan a las HD:

> “Las Humanidades Digitales son un “término paragua” de una amplia colección de prácticas para crear, aplicar e interpretar las nuevas tecnologías digitales de la información. Estas prácticas no están limitadas a los departamentos convencionales de humanidades, pero afectan cada campo humanístico en la universidad, incluyendo historia, antropología, artes y arquitectura, estudios sobre datos e información, estudios sobre cine y medios, arqueología, geografía y el resto de las ciencias sociales.” (Presner y Johanson, 2009)

Como indicamos al comienzo, el abanico es grande e intenta agrupar todo lo que toca. Nos encontramos en los albores de una nueva manera de interpretar las tecnologías ligadas a la comunicación e información (TIC), En esta manera de organizar la mirada, se disloca el mismo campo de las ciencias. Por este motivo, nos proponemos revisar los antecedentes de estas prácticas para observar y posteriormente analizar la transversalidad que se nos presenta.

3.2 Las tres olas de las Humanidades Digitales

Los inicios de las Humanidades Digitales tienen sus comienzos en los años 40 con los primeros proyectos de digitalización de textos en las bibliotecas, con la finalidad de facilitar a los investigadores la recolección de información y datos.

El común denominador era pensar a las nuevas tecnologías como herramientas subordinadas al trabajo del investigador, siendo facilitadoras en el proceso de producción, pero lentamente los avances a nivel de programación, fueron modificando el rol de estas herramientas, situándolas junto al mismo objeto de investigación, abriendo de esta forma, nuevos interrogantes a los mismos usuarios.

Burdick (2012) comenta en el paper Digital_Humanities[4], que durante décadas las HD continuaron imaginando lo digital como una vía para expandir las bases de datos que los investigadores usaban en sus trabajos o simplemente estaban direccionadas hacia audiencias ansiosas de conocimiento. Estas actividades se enfocaban en el ensanchamiento del corpus humanístico, desarrollando máquinas que leyeran archivos, facilitando el trabajo de bibliotecarios o especialistas en información.

En este sentido, David Berry (2011) señala que las humanidades digitales comienzan con el nombre de *informática* y eran vistas como soporte técnico respecto del trabajo académico. Su rol en el proceso de investigación o desarrollo, se sustentaba en optimizar el procesamiento de datos, sin llegar a considerar su participación en la crítica de los mismos.

> "Cambiar el término "Humanidades "Digitales" significaba que el campo había emergido de un bajo nivel de status como un servicio técnico, para pasar a ser un esfuerzo intelectual genuino con sus propias prácticas profesionales, estándares rigurosos e interesantes exploraciones teóricas (Hayles 2011)." (Berry, 2011)

El autor retoma en su artículo a Presner (2010), quien explica el panorama reciente de las Humanidades Digitales como "dos olas". La **primer ola** de HD se da en los finales de los 90's y comienzos del 2000. Se tendía a focalizar en las digitalizaciones de proyectos en gran escala y el establecimiento de infraestructura. La **segunda ola**, a la cual denomina "Humanidades Digitales 2.0", es más generativa, creando entornos y herramientas para producir, curar e interactuar con el conocimiento que nace de lo digital.

Respecto a estos dos campos, Berry señala que la información comienza a crear su propia información, es decir, la digitalización de textos permite que el campo comience a dar sus propios datos independientemente del papel. Por lo tanto, la problemática de la veracidad de la fuente entra como una dimensión que caracteriza a este movimiento.

4 BURDICK, Anne; DRUCKER Johanna; LUNENFELD, Peter; PRESNER Todd y SCHNAPP Jeffrey (2012) *Digital_Humanities*, Massachusetts Institute of Technology

A partir de esto, Berry expande la mirada proponiendo una **tercera ola** y, sobre ella, un campo tentativo sobre dónde se debería avanzar:

> "En efecto, podríamos decir que la tercer ola de HD marca el camino en el que la tecnología digital debería remarcar las anomalías en un proyecto de investigación humanista y que nos lleva a cuestionar los supuestos implícitos en tales investigaciones, por ejemplo la lectura cercana, formación canónica, periodización, humanismo liberal, etc." (Berry, 2011)

Estas tres olas dejan en claro la transformación sufrida por las herramientas de digitalización y procesamiento de datos, en principio consideradas importantes para el desarrollo del trabajo pero en segundo plano respecto a la investigación, donde prima la observación del cientista. En este movimiento, los software fueron tomando relevancia hasta lograr una paridad con el investigador. Ahora las herramientas son parte importante de la investigación, como también objeto de estudio de los investigadores.

3.3 ¿Humanidades Digitales en un renglón?

Rafael Alvarado (2011) afirma que no hay definición para las humanidades digitales si estamos buscando un conjunto de preocupaciones teóricas y métodos de investigación que puedan ser alineados con una disciplina determinada. Sólo asegura que los investigadores en humanidades digitales comparten su origen de humanistas y entusiastas por trabajos de literatura, arte, etc.

Similar a la explicación de Alvarado, es el camino que nos propone Alejandro Piscitelli. Para el filósofo, las HD son un conjunto interrelacionado de dominios y recursos que contribuyen a la construcción del conocimiento; pero, según el enfoque, ya sea un texto o una herramienta, el grado de variación en los resultados puede llegar a ser importante:

> "Quizás lo que mejor define a las humanidades digitales no sean tanto las tecnologías a las que recurre (innumerables y volátiles), ni los métodos (que provienen de los campos consagrados más diversos), sino cierto "encuentro" alegre y juguetón con la representación digital en sí misma." (Piscitelli, 2013)

El presente autor sostiene que los humanistas digitales actuales están enmarcados en el movimiento maker, personas que trabajan en laboratorios programando y desarrollando softwares, dispositivos y plataformas, combinando aptitudes tanto teóricas como prácticas que agilizan la creación de conocimiento. Es así, que presenta la idea de espacios, como los labs[5]. Los mismos,

5 Los labs son entornos de innovación tecnológica, donde se generan propuestas que mejoren los usos y apropiaciones

abren el camino del trabajo colaborativo al académico digital, teniendo como premisa fundamental en un proyecto, el "campamento" que se realiza en la web:

> "También queda claro que lo que las tecnologías están haciendo no es solo cambiar el modo de "hacer" humanidades y ciencias sociales sino al sistema educativo en su conjunto." (Piscitelli, 2012)

Según Piscitelli, esto se daría a partir de la tríada "Trabajo Colaborativo - Producción - Implementación", que se presenta como invisible para la academia tradicional, respecto sobre como se piensa/hace en la época de la web masiva: no sólo se acelera la manera en que se produce, sino que es posible la colaboración simultánea en tiempo y espacio, permitiendo a personas que en principio no estaban calificadas a formar parte de la conversación.

En este sentido, siguiendo a Presner y Johanson, podemos caracterizar a los trabajos dentro de las Humanidades Digitales como:

a) Interdisciplinarios: no podemos pretender que existen campos puros de investigación. Es necesario el acercamiento de los campos para que las producciones puedan cubrir esos espacios o preguntas que se dejaban a las demás disciplinas.

b) Colaborativas: el rol del autor queda diluido y se pasa a trabajar en grandes equipos interdisciplinarios donde los aportes hacen al conjunto, siendo el concepto de inteligencia colectiva de Levy (1994) clave para comprender la dinámica que en esta constelación se despliega.

c) Socialmente comprometidas, globales, oportunas y relevantes: las redes sociales han cambiado las construcciones de las subjetividades, y con ello se alteraron las maneras de producir y consumir sobre estas nuevas plataformas. La actualización a tiempo real de las actividades o sentimientos; la escala global de los acontecimientos; las tendencias generadas en plataformas como Twitter o Youtube donde no importa el tiempo o espacio. Siempre hay alguien produciendo datos e información, que es compartida en la web. Esta masiva cantidad de información, concebida con el nombre de "Big Data", da lugar a nuevos acercamientos académicos en la producción de conocimiento.

3.4 Proyectos enmarcados en las Humanidades Digitales

A continuación exploraremos 3 proyectos que abarcan diferentes aristas de las Humanidades Digitales, y en las cuales podremos ver en acción los conceptos que guían a las mismas.

---

de las TIC. Piscitelli llevó adelante uno con el programa Conectar Igualdad, se llamó Conectar Lab http://conectarlab.com.ar/que-es-el-lab/que-hacemos/

3.4.1 ¿Cómo comparar un millón de páginas?

Tan pronto como un nuevo libro de Manga es publicado en Japón, los fans lo compran, escanean las páginas, lo traducen a otros idiomas, y distribuyen las imágenes digitales de las páginas traducidas a través de sitios web. En el proceso, también insertan páginas adicionales (como los créditos de producción, comentarios y fan-art[6]). Este proceso es referido por ellos como "scanlation"[7]. Hasta Julio del 2010, el sitio más popular de archivos online de scanlations era OneManga.com (y también era el más popular en toda la web, llegando a rankear entre los 300 sitios más vistos de Estados Unidos)

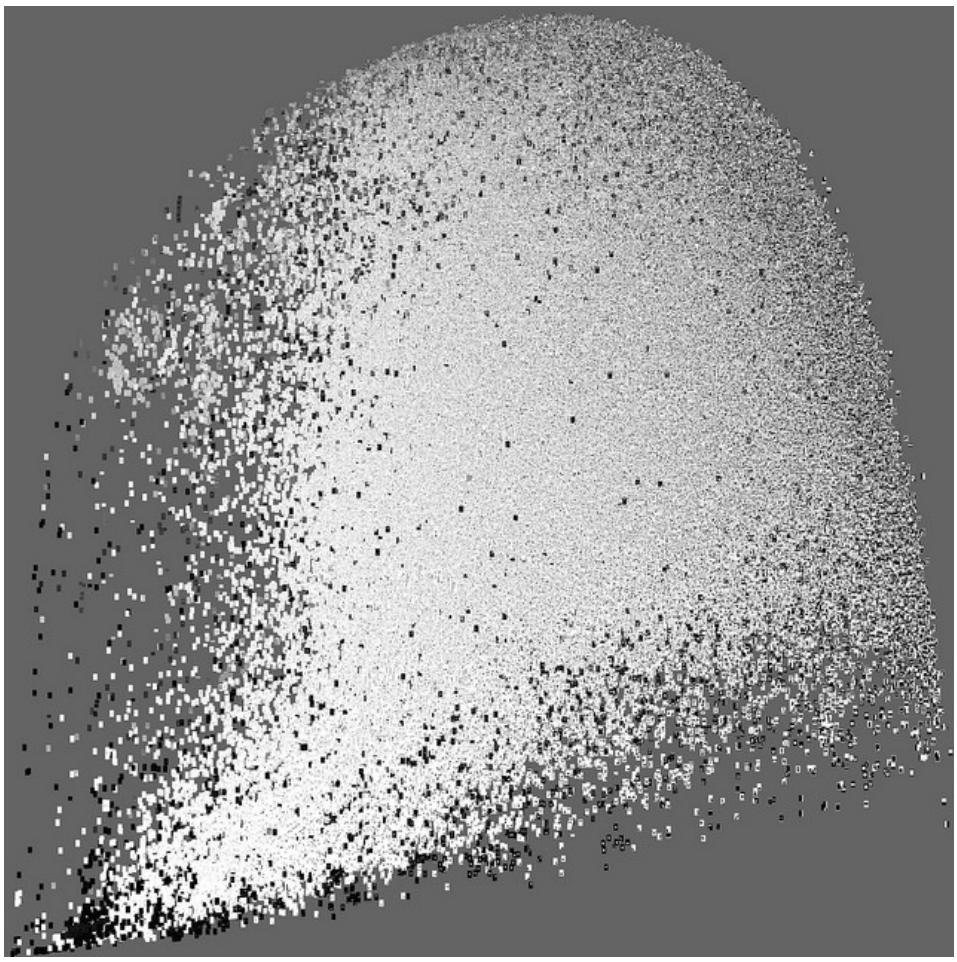
En el Otoño de 2009, el grupo de Estudios de Software[8] de Lev Manovich descargó del sitio 883 series de Manga, que contenían 1.074.790 páginas únicas. Utilizando sus programas de procesamiento de imágenes, a la que llaman "super-computadora", localizada en Departamento Nacional del Centro Energético de Investigación Científica Computarizada, analizan las características visuales de estas páginas.

La visualización en la imagen siguiente dispone la muestra completa de 1 millón de páginas de Manga organizadas en un espacio en 2 dimensiones de acuerdo a sus características visuales.

6 Producciones plásticas subordinadas a un elemento de un universo previo que el o la fan adopta y que le sirve como regla y punto de partida para proponer una nueva creación. Véase el Capítulo 2: Imágenes - Fan art, fanvids de BORDA, Libertad (2012)

7 Preferimos mantener el nombre dado que la denominación se ha adoptado entre todos los idiomas.

8 Iniciativa de Estudios de Software http://lab.softwarestudies.com

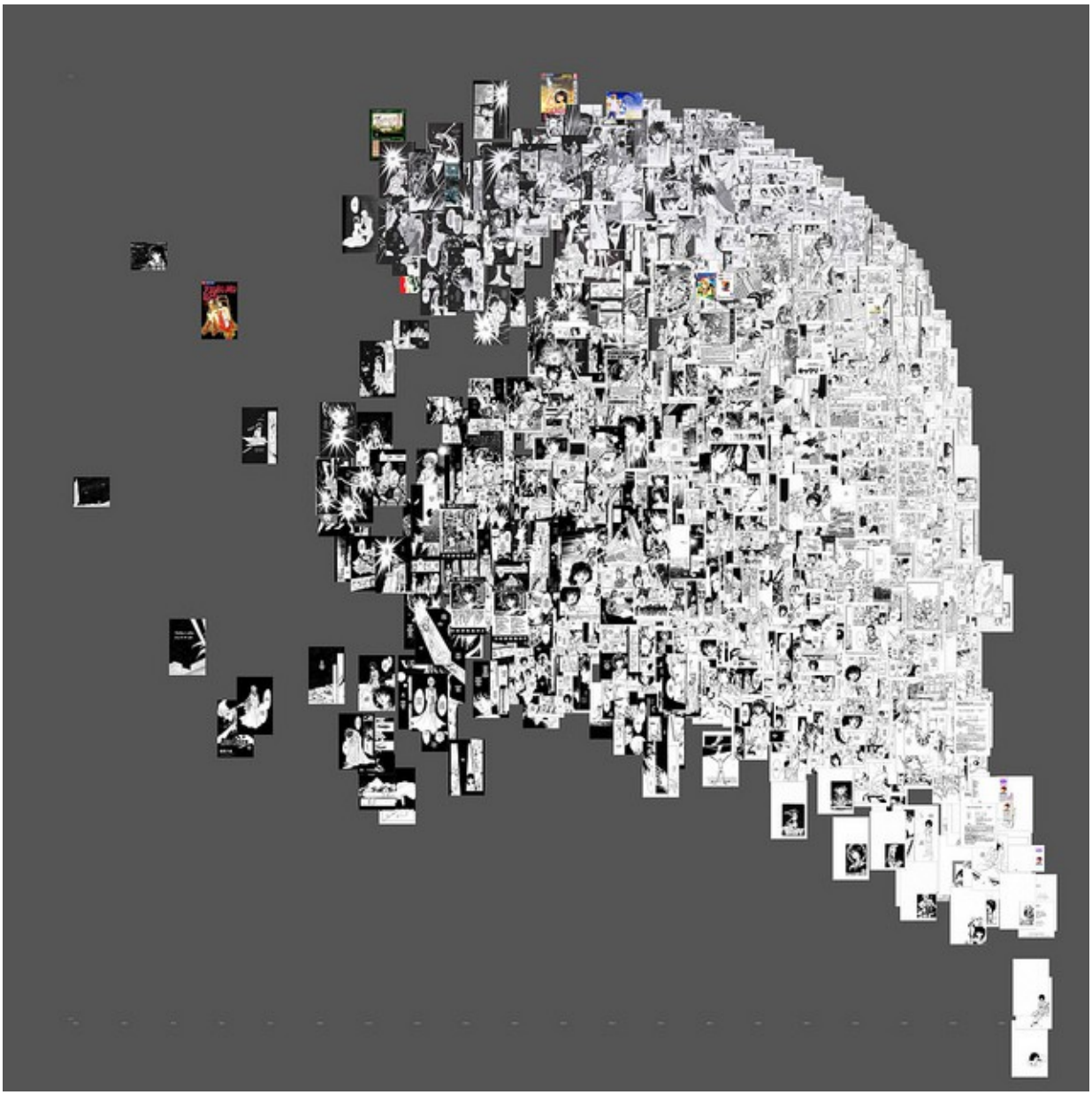


(nota: debido a que las páginas están procesadas una arriba de la otra, no se ven realmente 1 millón de páginas distintas, la visualización muestra una distribución de todas las páginas con los tipos de ejemplo apareciendo arriba)

Las páginas en la parte inferior son las más gráficas (significa que tiene la menor cantidad de detalles). Las páginas en la sección superior derecha tienen, por el contrario, muchos detalles y texturas. Las páginas con mayor contraste están a la derecha, y a medida que decrece el contraste se ubican en la izquierda. En el centro están todas las posibles variaciones estilísticas.

La idea principal que comienza a problematizar Manovich es la de **estilo**. Para el autor, el concepto asume la idea de que podemos dividir un conjunto de obras en un pequeño número de categorías discretas. Sin embargo, si encontramos un número muy grande de variaciones con muy pequeñas diferencias entre ellas, como sucede en este trabajo, entonces no podemos seguir hablando de “estilo” en el sentido estricto en un caso como este. Manovich sugiere, en cambio, que es mejor utilizar estas visualizaciones y/o modelos matemáticos para describir el espacio de las variaciones posibles. A partir de esto el autor se pregunta que sucedería si se tomara como caso una sola serie de Manga ¿Tendría sentido hablar de estilo?

Utilizando los mismos procedimientos de procesamiento de imágenes para el manga Anatalia Story, se encontró tal variabilidad gráfica que llegaron al mismo punto que Manovich. El concepto de estilo no era suficiente para dar cuenta del proceso que acontecía



Ejemplo: 879 páginas se organizan por medio de brillo (X) y la entropía[9] (Y):

En estos ejemplos, las páginas del manga están organizadas de acuerdo a características visuales particulares. Teniendo en cuenta otras características y también atributos de orden superior (contenido, composición, convenciones visuales del manga para presentar a los personajes, sus caras, fondos, etc.) podrían llegar a revelar la presencia de una coherencia estilística en títulos individuales. Sin embargo, es menester señalar que como cualquier análisis automático computarizado de grandes muestras de cultura humana[10], existen limitaciones y, por lo

9 La entropía describe en parte los datos de la imagen: si la imagen consiste de unas pocas áreas monocromáticas. su entropía será baja. En cambio, si la imagen tiene mucho detalle y texturas, y sus colores (o escala de grises en caso de ser en blanco y negro) varían de manera significativa, su entropía será alta. MANOVICH (2011)

10 La expresión cultura humana resulta engañosa, pero siguiendo al autor, podemos entender que se contempla una producción de cultura automatizada, por ej. la de robots programados en las distintas redes sociales. Si bien entendemos que en última instancia la programación está a cargo de un humano, este tema genera un punto de tensión

tanto, no se podrá reemplazar a la intuición y experiencia humana.

De todas maneras, aún teniendo en cuenta estas variadas limitaciones, las oportunidades que nos ofrece son inmensas: por ejemplo, tener acceso a un millón de páginas de manga nos permite, en principio, poder mapear de manera fiable el espectro total de las posibilidades gráficas usadas por los mangakas contemporáneos de la rama comercial. Tal mapeo nos permitirá entender qué series son estilísticamente más típicas y cuáles únicas; poder encontrar en cada serie los cambios significativos en el lenguaje gráfico utilizado a lo largo del tiempo (al día de hoy, las series más populares han estado siendo publicadas por una cantidad de años); investigar si tantas series cortas como largas tiene diferentes patrones; separar a los artistas que han cambiado significativamente su lenguaje gráfico de los que no lo han hecho, etc, etc. Como vemos, el campo se amplia, dado que la herramienta nos permite jugar con diversas lecturas (distantes y cercanas).

Para abordar estos desafíos, el grupo de investigación de Lev Manovich desarrolló una serie de métodos y técnicas denominadas Analíticas Culturales. Como explicamos en el apartado **2. "¿Quiénes hablan?"**, la idea central es el uso de software de visualización para explorar conjuntos de imágenes, videos o texto que conforman nuestra base de datos. Estas visualizaciones pueden usar los metadatos[11] existentes e incluso nuevos metadatos agregados por los investigadores, ya sea por anotación o codificación. Sin embargo, agregar etiquetas u otras anotaciones manualmente contrae serias dificultades: por un lado nuestro sistema visual natural no puede dar cuenta de diferencias visuales sutiles entre una gran cantidad de imágenes; y por otro, nuestro lenguaje natural textual no posee términos para describir todas las ínfimas características visuales, o nombrar las posibles variaciones.

Entonces, establecidas estas dificultades, el autor propone el siguiente acercamiento para poder superar estas barreras:

> "Nuestro acercamiento consiste en usar las visualizaciones como un nuevo sistema descriptivo. En otras palabras, describimos usando imágenes. Al hacer esto, tomamos la ventaja que nos proporcionan las mismas para registrar diferencias sutiles en cualquier dimensión visual." (Manovich, 2011)

Por último, deseamos destacar la idea que este proyecto (propuesta inicial del equipo de investigación de Manovich) utiliza imágenes como categorías de clasificación al tiempo que busca

que se aleja de nuestros propósitos.

11 Nos referimos a la información incorporada a las propiedades del archivo pero que no son visibles, por ejemplo: Fecha de publicación, modelo de cámara, velocidad de obturación, nombre de autor, etc.

ampliar la potencia de las herramientas disponibles para procesar datos. Manovich no sólo observa el campo de producción cultural masiva, está innovando e interviniendo en la manera en que se investigan las mismas.

3.4.2 Inexplorado o no trazado: Big Data como un lente para la cultura humana

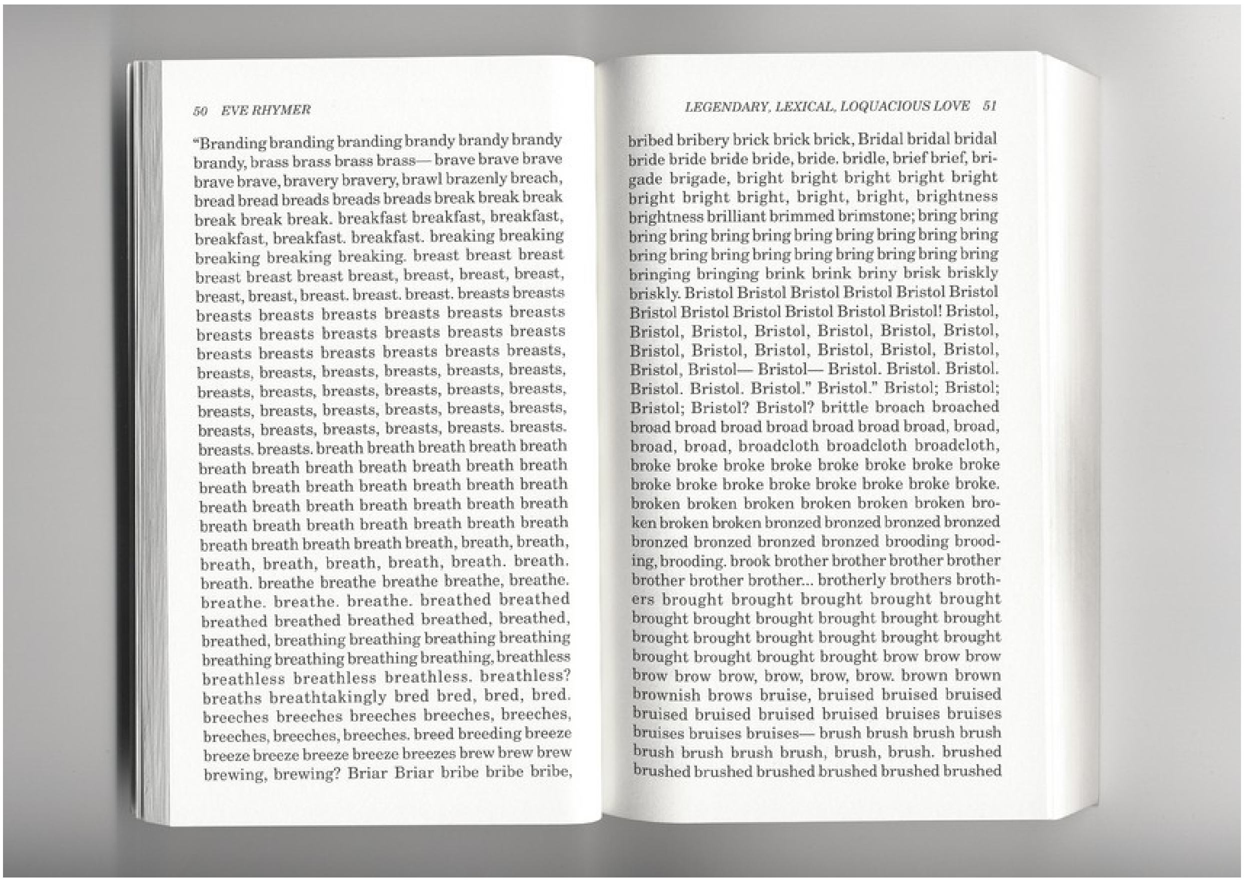
Erez Aiden & Jean-Baptiste Michel, recientes egresados apasionados por las letras, visitaban frecuentemente la biblioteca de la Universidad de Michigan buscando libros raros, con una peculiar característica adicional; revistaban aquellos que no eran retirados frecuentemente del estante. Su investigación estaba centrada en la historización de los verbos irregulares del idioma inglés, dado que habían detectado un patrón en el cual estos estaban siendo reemplazados por los regulare terminados en -ed)

Inaugurando con estos aspectos su campo de interés, decidieron realizar una investigación de forma más sistemática. Intentaron reunir información de todos los libros publicados existentes, es decir, unos 130 millones. Luego de varias reuniones con Google, la mayor empresa encargada de organizar la información mundial, Aiden y Michel terminaron creando *Google Books Ngram Viewer*, una poderosa herramienta que utiliza como base los 5 millones de libros digitalizados por Google hasta el momento, y que en su conjunto suman un aproximado de 500 billones de palabras.

Ngram Viewer funciona a partir de la indexicalización de las palabras registradas como datos duros (raw text - texto crudo) que forman parte de la base de datos de Google Books. La intención de crear una interfaz de estas características, tiene como objetivo poner a disposición del público una parte importante de estos textos, sin infringir las leyes de derecho de autor.

La idea de utilizar un buscador que tomara en cuenta la indexicalización de las palabras, provino del libro de Karen Reimer: *Legendary Lexical Loquacious Love*[12] . La autora ordenó toda una novela de manera alfabética, obteniendo 26 capítulos (uno por letra, exceptuando la letra X dado que no se contaba con la misma en la novela) en orden como se ve en la imagen:

12 Ver todo el texto en http://www.karenreimer.info/work/legendary-lexical-loquacious-love



50 EVE RHYMER

"Branding branding branding brandy brandy brandy brandy, brass brass brass brass— brave brave brave brave brave, bravery bravery, brawl brazenly breach, bread bread breads breads breads break break break break break break. breakfast breakfast, breakfast, breakfast, breakfast. breakfast. breaking breaking breaking breaking breaking. breast breast breast breast breast breast breast, breast, breast, breast, breast, breast, breast. breast. breast. breasts, breasts. breasts. breasts. breasts. breath, breath, breath, breath, breath, breath, breath, breath. breath. breath. breathe breathe breathe breathe, breathe. breathe. breathe. breathe. breathed breathed breathed breathed breathed breathed, breathed, breathed, breathing breathing breathing breathing breathing breathing breathing breathing, breathless breathless breathless breathless. breathless? breaths breathtakingly bred bred, bred, bred. breeches breeches breeches breeches, breeches, breeches, breeches, breeches. breed breeding breeze breeze breeze breeze breeze breezes brew brew brew brewing, brewing? Briar Briar bribe bribe bribe,

LEGENDARY, LEXICAL, LOQUACIOUS LOVE 51

bribed bribery brick brick brick, Bridal bridal bridal bride bride bride bride, bride. bridle, brief brief, brigade brigade, bright bright bright bright bright bright bright bright, bright, bright, brightness brightness brilliant brimmed brimstone; bringing bringing brink brink briny brisk briskly briskly. Bristol Bristol Bristol Bristol Bristol Bristol Bristol Bristol Bristol Bristol Bristol Bristol! Bristol, Bristol, Bristol, Bristol, Bristol, Bristol, Bristol, Bristol, Bristol, Bristol, Bristol, Bristol, Bristol, Bristol, Bristol— Bristol— Bristol. Bristol. Bristol. Bristol. Bristol. Bristol." Bristol." Bristol; Bristol; Bristol; Bristol? Bristol? brittle broach broached broad broad broad broad broad broad broad, broad, broad, broad, broadcloth broadcloth broadcloth, broke broke broke broke broke broke broke broke broke broke broke broke broke broke broke broke. broken broken broken broken broken broken broken broken broken bronzed bronzed bronzed bronzed bronzed bronzed bronzed bronzed brooding brooding, brooding. brook brother brother brother brother brother brother brother... brotherly brothers brothers brought brow brow brow brow brow brow, brow, brow, brow. brown brown brownish brows bruise, bruised bruised bruised bruised bruised bruised bruised bruises bruises bruises bruises bruises— brush brush brush brush brush brush brush brush, brush, brush. brushed brushed brushed brushed brushed brushed brushed

Utilizando esta función, NGram arroja resultados a partir de la frecuencia de aparición de la palabra en el canon de libros digitalizados por Google. Veamos lo expresado en el siguiente gráfico:

En el próximo capítulo retomaremos el funcionamiento específico de esta herramienta, pero deseábamos rescatar de la misma como fue el desarrollo del software. Esto posibilitó un lugar

nuevo a los usuarios de internet, dando acceso a una base de datos inmensa como la que indexa Google. Destacamos que su interfaz es sencilla, facilitando a los investigadores un acercamiento y posibilidad de lectura sobre los 5 millones de libros, generando así, un punto de partida o apoyo para sus trabajos.

3.4.3 La forma de un discurso espectacular: Una infografía analítica sobre los elementos que hicieron grande al discurso "I Have a Dream" de Martin Luther King Jr.[13]

Fuente: vimeo.com/18792376

Aquí tenemos un ejemplo que combina la visualización sobre un corpus literario que funciona como base de datos. Si bien no se compara con la magnitud de los trabajos de Manovich y Aiden - Michel, es importante tener en cuenta que una buena visualización de la información, también forma parte de los trabajos en las Humanidades Digitales. A su vez reconocemos que este es un caso especial, dado que ha sido analizado y estudiado su impacto como su potencia en el campo político - social y cultural; intentando comprender, desde la estructuración de sus componentes como así también los motivos que lo hicieron y hacen tan impactante. Este trabajo combina la voz del orador con la visualización del discurso, construyendo un nuevo significante, develando aspectos que se encontraban presentes pero que eran invisibles a nuestra mirada.

Nancy Duarte, especialista en oratoria y presentaciones, realiza un interesante análisis del discurso del Dr. Martin Luther King Jr.. En el mismo se da cuenta que Luther King utiliza una

13 Ver más en brainpickings.org/index.php/2013/08/28/nancy-duarte-mlk-speech/

manera más poética que los discursos políticos de largo aliento. La visualización nos permite ver las repeticiones, las metáforas, imágenes sensoriales, referencias políticas, citas bíblicas y canciones espirituales que conforman el cuerpo del discurso y desnaturalizando su lectura tradicional. Asomándonos a una nueva constelación.

En principio, la visualización dispone el modo de lectura en sentido horizontal (Fig. 1). Por otra parte, lo separa en momentos dónde el discurso asciende y desciende. Según la autora, esto corresponde a la forma persuasiva del discurso entre lo “que es” (barra inferior) y lo “que puede ser” (barra superior). Luego, agrega todo el discurso en manera vertical sobre cada momento en que los ha separado (Fig. 2). En estas instancias se hace muy dificultoso poder leerlo, pero *la búsqueda está más allá del contenido*, la idea central es poder ver desde lejos el despliegue del discurso y transformarlo en datos. Aquí es donde las palabras se transforman en barras (Fig. 3)

Fig. 1

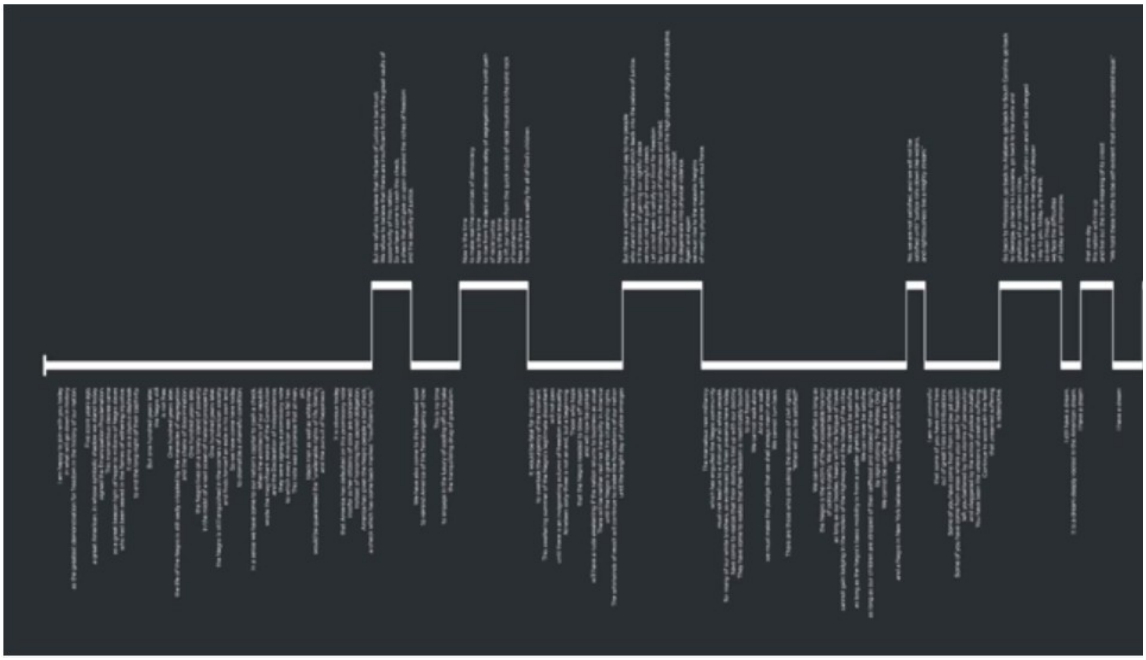


Fig. 2

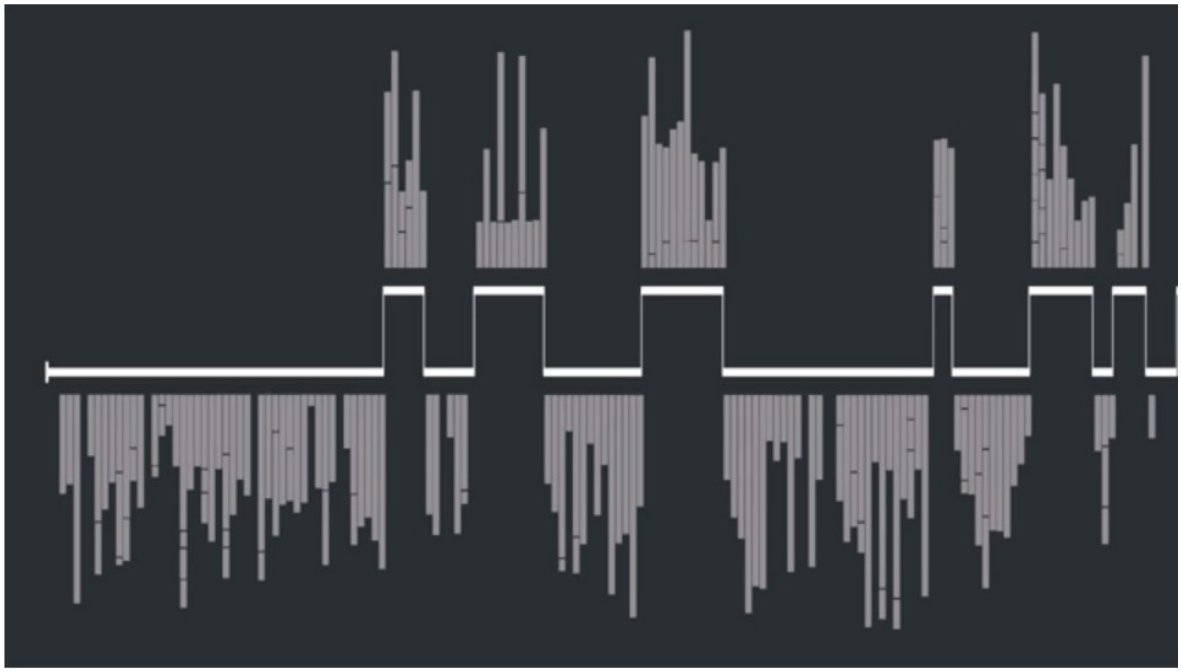


Fig. 3

El paso siguiente es identificar cada barra con un color según corresponda lo dicho por Luther King. Nancy Duarte encuentra en el discurso “repeticiones”, “metáforas”, “canciones, literatura o frases populares” y “referencias políticas”. A través de la textura que nos brindan los colores en las palabras, podemos dar cuenta de la predominancia de las metáforas. (Fig. 4)

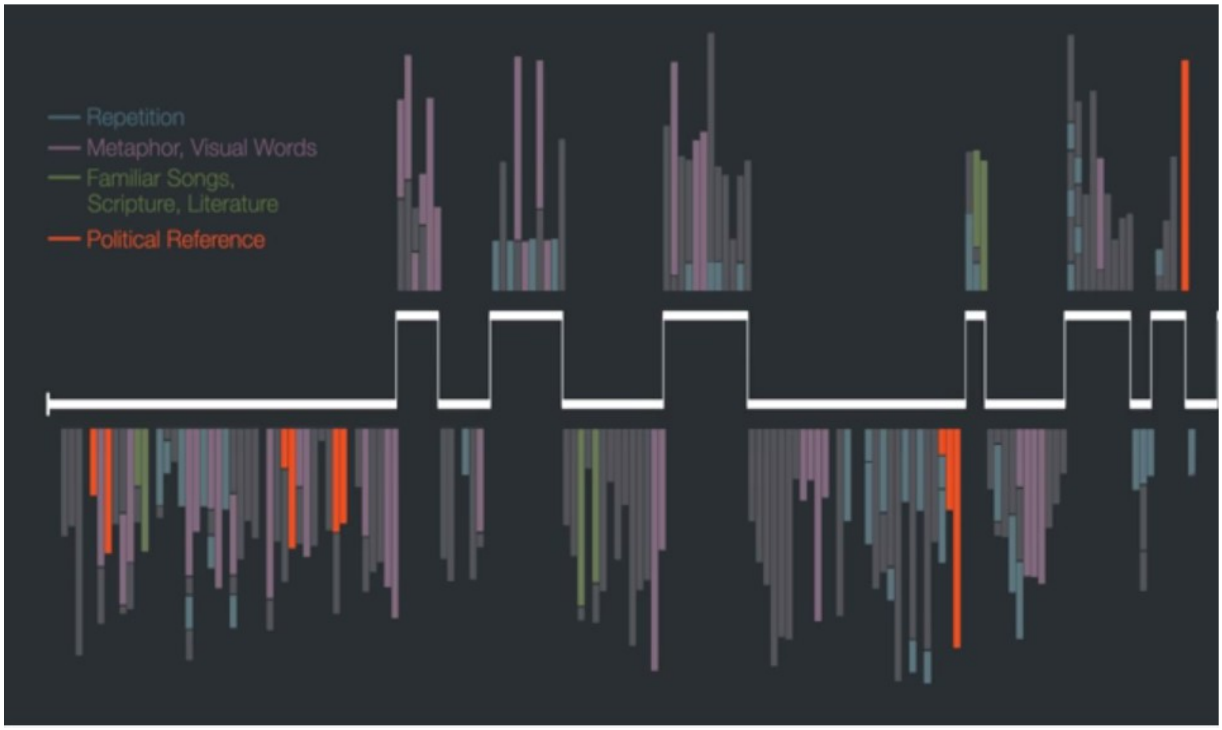


Fig. 4

Aún cuando el discurso fue una manifestación política concreta, con alto impacto en la sociedad y la historia, las referencias políticas concretas en el mismo (por ej. la Constitución) son menores. Llegando al final, podemos ver que directamente las referencias desaparecen. (Fig. 5)

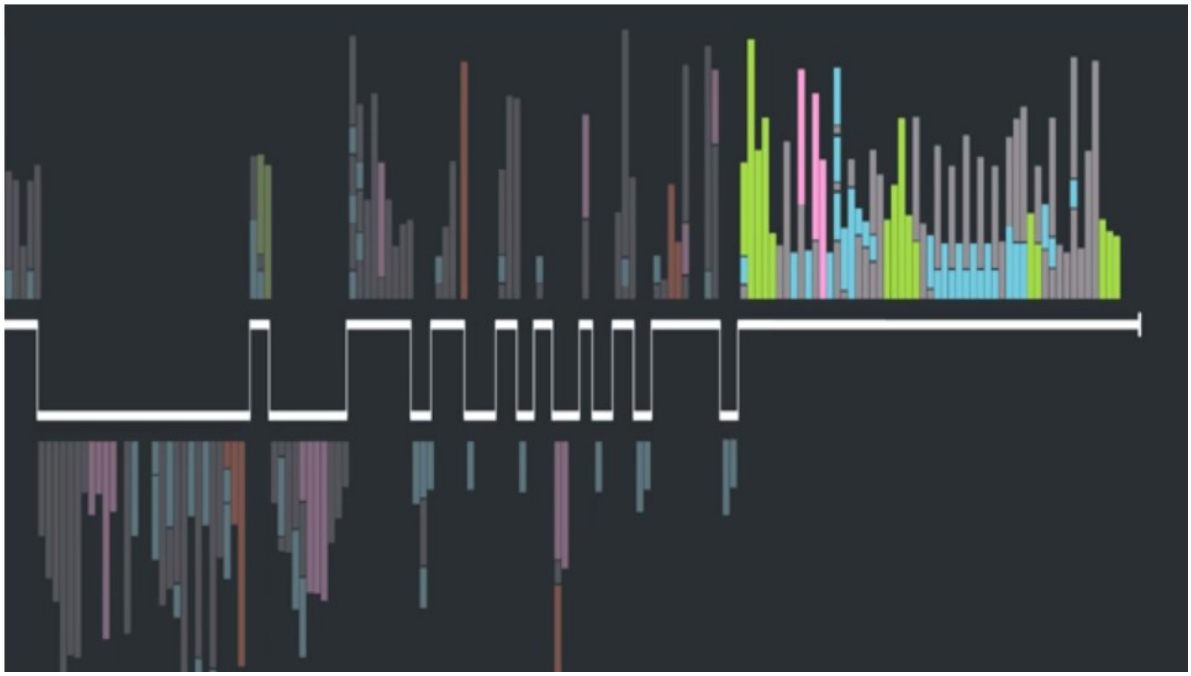


Fig. 5

Por otra parte, vemos cómo el discurso toma un ritmo vertiginoso en el medio antes del llegar al final. Aquí es donde las referencias a la Biblia o canciones populares son predominantes, siendo

la más importante “Free at last”.

Como mencionamos en el inicio de este sub-apartado, este trabajo no parte de la base de millones de unidades como lo hace el común de los proyecto enmarcados en la big data, pero si da cuenta de la forma en que se puede presentar la información a partir de la investigación sobre el contenido. Duarte, nos lleva de la mano de la lectura distante (si bien aquí está aplicada a un sólo discurso). Nos permite analizar y observar nueva información a la cual, si no es por estas herramientas, no se podía acceder, permitiendo hacernos nuevas preguntas sobre lo que vemos.

3.5 Comentarios finales sobre este apartado

A lo largo de este capítulo, hemos podido ver las distintas características y trabajos en el campo de las Humanidades Digitales: gran cantidad de datos, intervención de distintas disciplinas, colaboración de los usuarios, apertura a las grandes bases de datos, metodologías que involucran nuevos lenguajes, distintas maneras de categorizar la información, y la intervención de software y herramientas para procesar todos estos datos. En el siguiente capítulo, abordaremos y analizaremos el funcionamiento, posibilidades y limitaciones, de las distintas herramientas de procesamiento de datos y visualización.

4. Herramientas de procesamiento de datos y visualización

A continuación veremos una serie de herramientas que consideramos importantes para el trabajo de los que quieran experimentar en el campo de las Humanidades Digitales. Si bien el presente apartado, podríamos haberlo dejado como un anexo, creemos que es necesario realizar un análisis críticos de las HD, estableciendo límites y alcances, posibilidades de producción, como así también presentar algunas ideas sobre cómo colaborarán en el trabajo del cientista social.

En su conjunto y como observación previa, destacamos que las herramientas que se presentan a continuación, son aplicables solamente al trabajo con textos. Es por esta razón que tenemos una especial estima por los proyectos del laboratorio de Lev Manovich, en donde intenta analizar imágenes a partir de imágenes. Por otra parte, a cada herramienta le aplicaremos una categorización según el tipo de accesibilidad que le permita al usuario a la base de datos. Según pueda acceder y/o modificar diferenciamos entre: **a) Abiertas:** pueden aportar y/o modificar dentro del sitio; y **b) Semi-abiertas:** pueden ingresar y obtener pero no modificar la base dentro del sitio.

4.1 N-gram

Ficha Técnica:

Nombre completo: Google Books N-Gram Viewer

Creadores: Erez Lieberman Aiden and Jean-Baptiste Michel, en colaboración con Google Books Project.

Link / Enlace al sitio: https://books.google.com/ngrams

Tipo de herramienta: Análisis de Texto

Tipo de accesibilidad al usuario: Semi-Abierta

En el capítulo anterior mencionamos la creación y objetivos de esta herramienta, ahora nos focalizamos específicamente en su funcionamiento práctico.

4.1.1 ¿Qué es un “N-Gram”?

A partir de la definición que brindan sus creadores, Erez Aiden y Jean-Baptiste Michel, el **N-Gram Viewer** grafica la frecuencia de las palabras e ideas a través del tiempo, tomando en cuenta los libros digitalizados (y luego filtrados por ellos) en Google. Por esa razón, el término académico que establecieron para generar una unidad de búsqueda es “n-gram”. Como una gran ecuación matemática, “n” es la cantidad de “gramos” que tiene esa palabra o frase en el espacio digitalizado.

> "Estas palabras y frases - el término computarizado de fantasía es **n-gram**- incluye a *3.14159* (un 1-gramo), *banana split* (2-gramos), y *los Estados Unidos de América* (5-gramos). Por cada palabra y frase, el registro consiste de una larga lista de números, mostrando con que frecuencia ese n-gram particular aparece en los libros, año tras año, yendo hasta casi 5 siglos atrás."

4.1.2 Funcionamiento

Es muy común que en el juego scrabble, cuando se produce una disputa sobre si la palabra formada es o no una "palabra", se recurra al diccionario. Allí se encuentran ordenadas alfabéticamente palabras o términos, aprobados oficialmente, y con sus significados correspondientes.

Aiden y Michel cuentan con la base de datos de Google Books conformada por un total de 500 billones de palabras. Y teniendo en cuenta que el diccionario más grande es el Oxford English Dictionary, compuesto por 23 volúmenes y 446.000 palabras, los investigadores decidieron que debían buscar un método mediante el cual se pudiera seleccionar las palabras de la misma manera que los lexicógrafos en el diccionario.

Los autores señalan dos perspectivas al respecto: la *prescriptiva y la descriptiva*. La primera establece que los lexicógrafos están a cargo del establecimiento del lenguaje y su desarrollo de en el diccionario. Ellos legislan qué palabras usar y cuáles no. La dificultad del presente enfoque radica en la subjetividad del lexicógrafo (aquella persona "a cargo" del lenguaje), cuando para los autores, el lenguaje excede a los gobiernos y nacionalidades.

En cambio, la *perspectiva descriptiva* no presenta la idea del "monarca", sino la del "explorador"; reportando lo que hacemos con el lenguaje. En este caso, el diccionario constituye un mapa de lo que han encontrado. Aún así, esta perspectiva presenta otro problema; los lexicógrafos no pueden decidir si una palabra lo es por mandato, y también se presenta la variable del "tiempo" (cuánto se tarda en decidir si una palabra debería ser incluída). Estas dificultades nos arrojan a diversos interrogantes: si el explorador no ve el campo completo y sólo la porción que va descubriendo, es muy posible que su avanzada tenga errores. ¿Cómo subsanarlos? ¿Quién los ve? ¿Qué estrategia adoptar? Existe una "orbe" por la cuál cerrar el "campo". El explorador por tanto, cae en su propio frente.

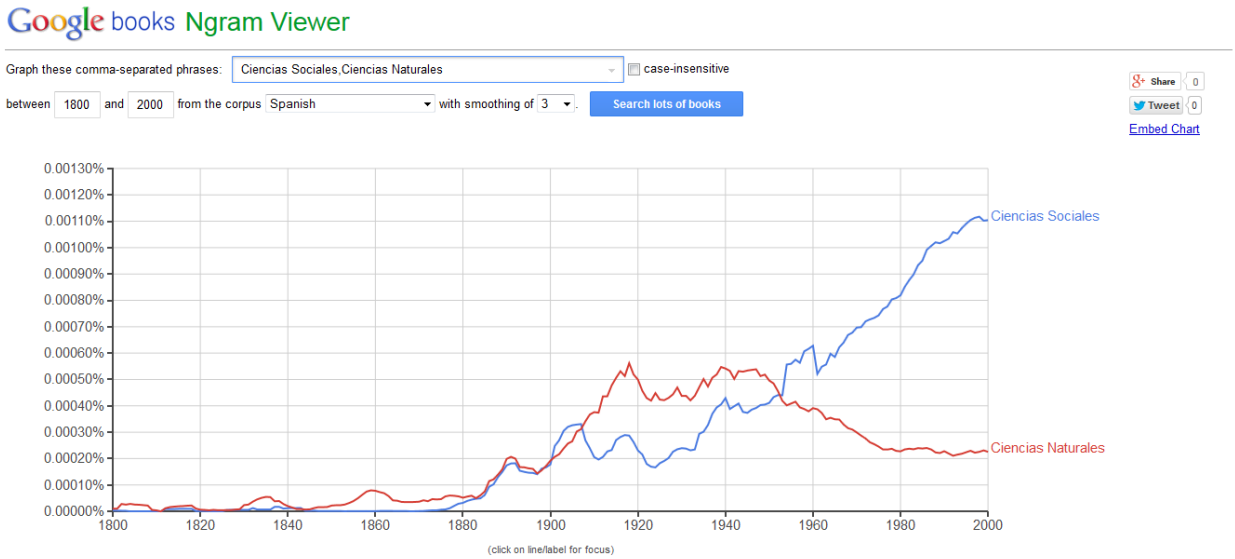
Por estas razones y a partir de las pruebas realizadas con N-Gram[14] los autores decidieron que debían crear su propio léxico descriptivo que contuviera toda las palabras contemporáneas

14 Véase AIDEN y MICHEL (2013:971), caso AHD (American Heritage Dictionary)

usadas en inglés:

> “Nuestra idea fue sencilla: si en un "cadena de caracteres" es suficientemente frecuente en los textos contemporáneos escritos en inglés, entonces es una palabra. ¿Cuán frecuente debe ser una palabra para ser considerada frecuente? El corte natural es usar la frecuencia de las palabras más raras en los diccionarios, que calculamos en 1 aparición cada 1 billón de palabras. Entonces nuestra respuesta para "¿Qué es una palabra?" fue la siguiente: *Una palabra en inglés es "1-gram" si aparece, en promedio, al menos una vez cada 1 billón de 1-grams de texto en inglés.”*

Bajo estas premisas, realizaron un recorte de tiempo dentro de su base de datos. Buscaron los libros publicados entre 1990 y 2000. Los mismos brindaron un dato asombroso: más de 50 billones de 1-grams. Para que un "1-gram" entrara dentro del corte, tenía que aparecer al menos 50 veces en esta colección. El resultado dio una lista de 1.489.337 palabras, entre las que encontraron unhealthiness, psychopathy y Augustean. Comparando el diccionario inglés con el contador de palabras analógico, este primero solo representa 1/3 del total de Google. Por fin los autores tenían su propio diccionario.



4.1.3 Advertencias sobre su utilización y críticas sobre accesibilidad

Los autores señalan que los usuarios de esta herramienta deben tener un gran cuidado: “si bien tenemos un lente potente para ver nuevos mundos, debemos recordar que aún la base de datos más potente yace bajo la soberanía de su interpretador. Cada lente es un nuevo espejo.”

En el mundo del NGram Viewer, los usuarios no podemos realizar aportes directos[15] a la base

15 La manera indirecta sería escribiendo un libro y que Google lo digitalice.

de datos de Google Books. Sin embargo, esta herramienta tiene liberado el texto crudo (raw text), bajo una licencia de creative commons 3.0[16]. Este detalle es altamente significativo: la base de datos puede bajarse y ser utilizada para realizar proyectos personales. Pero aunque sea significativo, lo es sólo enunciativamente, en la práctica todo es muy diferente. Para visualizar la magnitud de este proyecto, ingresamos a la sección[17] de descarga de su base de datos para tener una vista de cómo se muestra la información. Al ingresar, observamos una advertencia por parte de Google:

> Google Books NGram Viewer está optimizado para preguntas sencillas utilizando un pequeño conjunto de palabras. Si Ud. está interesado en realizar un análisis de mayor escala con grandes porciones de nuestra base de datos, seguramente deberá descargar una porción de nuestro corpus. *O toda ella, si tiene el ancho de banda y espacio necesarios*. Estamos felices de satisfacerlo.

Para ilustrar la potencia, observemos un instante la disposición de los enlaces de la base de datos:

A modo de ejemplo, procedimos a seleccionar de la sección “4-grams” el enlace que contiene las palabras con “bo”[18]. El peso del archivo comprimido es de 125 MB, pero el peso real es del archivo es de 1GB, lo que consideramos bastante grande para un solo enlace en toda esa lista que se ve en la imagen. Fue aquí que nos enfrentamos con un gran problema. Cuando intentamos abrir el archivo para ver cómo se estructura la información, la capacidad de procesamiento de

16 http://creativecommons.org/licenses/by/3.0/
17 http://storage.googleapis.com/books/ngrams/books/datasetsv2.html
18 storage.googleapis.com/books/ngrams/books/googlebooks-eng-fiction-all-4gram-20120701-bo.gz

nuestra máquina no se encontró habilitada para realizar tal tarea. Luego de la visita obligada a Google, encontramos dos sitios que explican como manipularla[19], pero no ahondaremos en esto, dado que no es nuestro objetivo de análisis. Solo podemos establecer que aquí hay un gran filtro de habilidades cognitivas: los que saben programar y los que no. Los que podrán utilizar toda su potencia, y los que solo la mirarán.

Aún con este problema, no impide que tengamos una gran apertura para potenciar el uso de esta herramienta. Aníbal Rossi (2013) nos ilumina sobre los usos de las variables que podemos introducir para refinar nuestra búsqueda: tags sintácticos, corpus y operadores matemáticos.[20]

Por último, a nuestro entender esta herramienta presenta una accesibilidad del tipo semi-abierta. En principio los mismos Aiden y Michel tuvieron dificultades para conseguir acceder a esta base de datos. Si bien se puede encontrar la "base de datos" cruda, las leyes del copyright no permiten darnos acceso completo, algo que estaría por fuera de nuestra perspectiva sobre el trabajo en las Humanidades Digitales. Sólo podemos consultar y utilizar los resultados. La mano del investigador es la que procede, la máquina hace lo que se ingresa. Sin embargo, es una herramienta potente y para nada desdeñable a la hora de trabajar en nuestras investigaciones.

4.2 Gapminder

Nombre completo: Gapminder for a fact-based world view
Creador: Hans Rosling / Hans Rosling Foundation
Link / Enlace al sitio: http://www.gapminder.org
Tipo de herramienta: Análisis de Texto / Bases de datos
Tipo de accesibilidad al usuario: Semi-abierta

En noviembre del 2010, la BBC presentaba el siguiente video www.youtube.com/watch?v=jbkSRLYSojo: 200 años, 200 países en 4 minutos. La alegría de las estadísticas. Un aplomado universitario animaba los datos en un espacio a través de burbujas; se atrevía a hablar y concluir sobre períodos históricos a través de círculos de colores. ¿Cómo lo hizo? ¿De dónde surgió la información? Sólo puso en relación bases de datos de organismos internacionales como la Organización Mundial de la Salud (OMS) o la del Trabajo (OIT)

19 http://homepages.inf.ed.ac.uk/lzhang10/ngram.html y http://homepages.inf.ed.ac.uk/lzhang10/ngram.html
20 ROSSI, Aníbal (2013) catedradatos.com.ar/2013/04/aprendiendo-a-usar-operadores-en-n-grams-viewer/

Para comprender su propuesta, tuvimos que viajar en el tiempo de la web. En el 2006, Hans Rosling[21] realizó su primer presentación, en las conocidas charlas TED, sobre su trabajo de investigación que llevaba más de una década. “Las estadísticas que reformulan el mundo que miramos” www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen

En dicha oportunidad, nos presentó un software que desarrollaron a partir de las herramientas de diseño existentes en dicha época[22]. Cómo menciona en el posterior video del 2010, su principal foco estuvo centrado en mostrar que:

> “Los datos no son suficientes. Yo necesitaba mostrarlos de alguna manera para que las personas los pudieran disfrutar y comprender”

Pero Rosling detecta un problema, y este tiene que ver con las pocas propuestas o iniciativas, similares a Gapminder, para analizar los datos. Si existen grandes bases de datos en organismos internacionales que destinan fondos a realizar estudios y encuestas, ¿por qué nadie opera sobre/con ellos?

> “Porque la información está escondida en las bases de datos. Existen algunos sitios que ofrecen información pero le ponen precios, contraseñas inútiles y estadísticas aburridas”

21 Hans Rosling es un médico sueco, académico, estadístico y orador. Es profesor de Salud Internacional en el Instituto Karolinska y es el co-fundador y director ejecutivo de la Fundación Gapminder, dónde han desarrollado el software detrás de las burbujas: el Trendalyzer.

22 Cabe señalar que los años a nivel de programación no se miden temporalmente sino en versiones y (re) versiones

Por esta razón, Rosling comenzó con la fundación Gapminder, entendiendo que no se necesitan bases de datos, sino herramientas de diseño que hagan visible esta información. Con las mismas, con el lente de la lectura distante, se podría liberar información de las bases de datos de organizaciones como las Naciones Unidas o como la OIT.

4.2.1 Guía para uso de la herramienta[23]

Gapminder es una herramienta es muy intuitiva y dispone de dos modalidades: trabajar en línea o descargar el programa para trabajar sin la necesidad de una conexión a internet. La ventaja de esta última, no sólo está dada por la disponibilidad o no de la conexión, sino que al descargar el programa, también se descarga toda la base de datos disponible.. Una vez elegida la modalidad, la interfaz es la misma para ambas. A continuación, se debe elegir el indicador (según la base de datos) para cada uno de los ejes x o y, y luego se observará el desplazamiento de las burbujas a lo largo del tiempo. A continuación brindamos una guía sencilla sobre los usos posibles de las herramienta:

23 Más información en www.gapminder.org/GapminderMedia/wp-uploads/tutorial/Gapminder_World_Guide.pdf

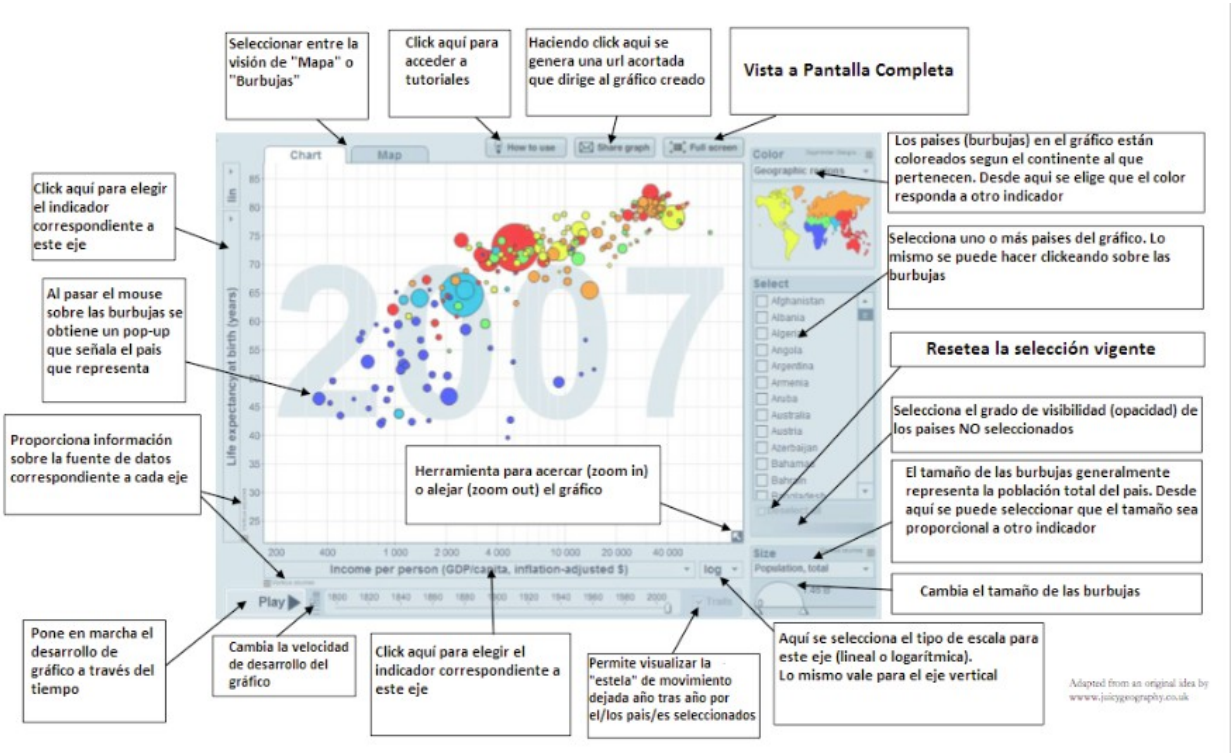
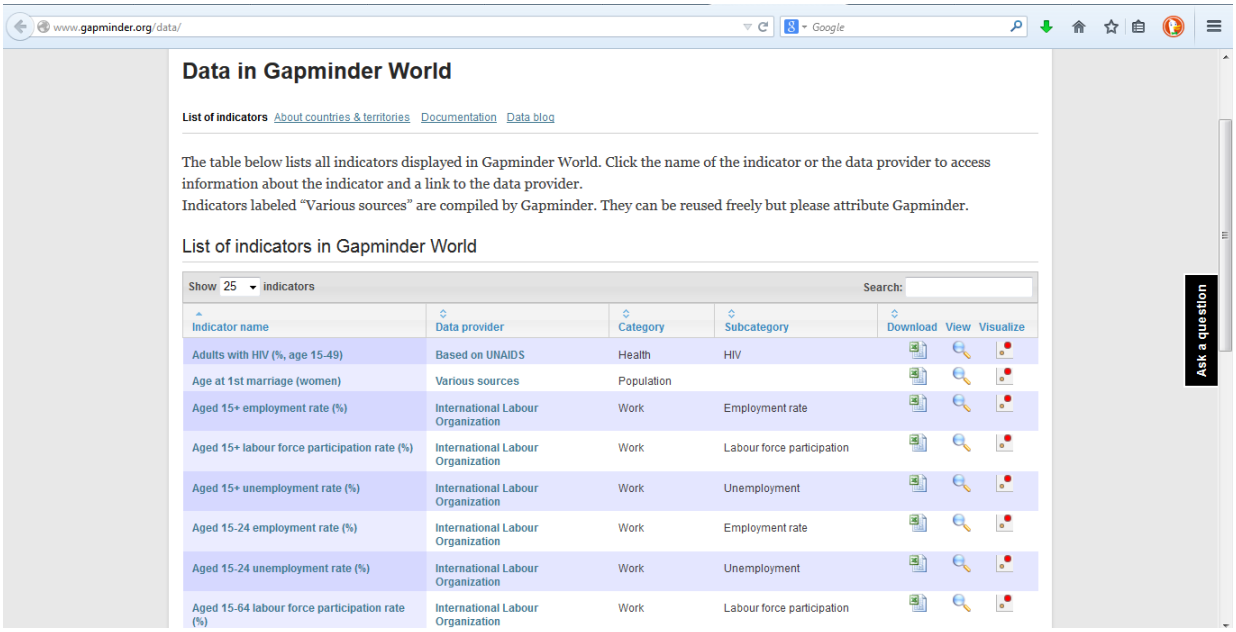


Gráfico traducido por el colectivo Digicom Siyp de la Universidad Nacional de Rosario

4.2.2 Sobre la base de datos: ventajas y desventajas

Al igual que con N-Gram, entendemos que Gapminder tiene un **tipo de accesibilidad del tipo semi-abierto**, dado que podemos entrar a ver la información e incluso descargarla. Comparativamente, resaltamos que la estructuración de los datos en Gapminder puede ser utilizada por cualquier usuario dado que utiliza tablas de hojas de cálculo. Sin embargo, no podemos participar del sitio subiendo datos, estando sujetos a la información del sitio.



Finalmente, Rosling advierte sobre el uso de estos datos *de manera genérica*. Entre sus conclusiones a partir del uso de Gapminder, destaca que es muy dificultoso hablar de brecha económica entre países cuando *dentro de los países podemos encontrar todos los estadíos de riqueza*. El autor enuncia que ya no puede hacerse la pregunta "¿Qué podemos hacer para sanear económicamente a África?" El objetivo es poder *ser lo más detallistas posibles*. En Gapminder, esa posibilidad está al alcance del usuario.

4.3 Many Eyes

Nombre completo: IBM Many Eyes Project

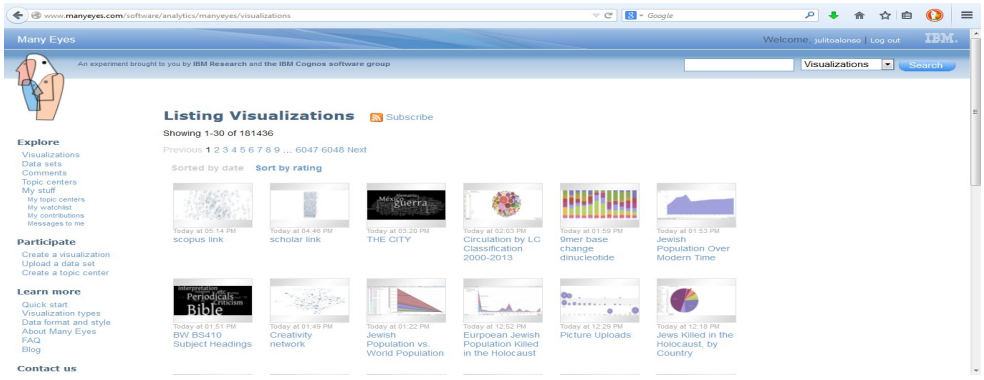
Creador: IBM Corporation

Link / Enlace al sitio: http://www.manyeyes.com/

Tipo de herramienta: Análisis de Texto / Tablas / Diagramas

Tipo de accesibilidad al usuario: Abierta

Many Eyes es uno de los softwares desarrollados por IBM para el análisis y visualización de datos. Esta herramienta presenta un variado abanico de opciones para trabajar sobre: **a)** la base de datos de IBM; **b)** base de datos de otros usuarios; **c)** base de datos propia.



Este es, a nuestro parecer, el punto fuerte de Many Eyes. Nos encontramos con una comunidad web donde la información es aportada y modificada por los usuarios. Hasta el momento solo podíamos depender de grandes bases de información de organismos internacionales o de una gran corporación como la de Google. Ahora el poder de armar bases de datos está del lado de los usuarios. En este sentido, esta herramienta es facilitadora: no es necesario poseer grandes conocimientos técnicos para poder usar las opciones de visualización.

4.3.1 ¿Cómo ingresar información?

Para realizar esta acción, Many Eyes pide la creación de un usuario en el sistema. De esta manera, todo lo que hagamos dentro de la página quedará registrado y será más sencillo buscarlo entre todos los datos que se suben por día. Por otra parte, la plataforma recuerda que toda la información que carguemos será visible, implicando que datos sensibles deberían tener un control especial. Por último, el límite de peso para cada archivo a subir es de 5 Mb.

Existen dos modos de ingreso de datos: texto libre o tablas. Respecto a las tablas, se pide que se coloque los datos en columnas, y en caso de tener que usar distintas unidades de medidas, especificarlo en las cabeceras. Se recomienda que se copie y pegue desde un programa de hoja de cálculo. Respecto al texto libre, se pide que primero se abra el texto con un procesador de texto, donde el formato desaparece, y luego copiarlo para posteriormente pegarlo.

Una vez ingresada la información, Many Eyes presentará una vista previa para evitar errores. Por último, tendremos que agregar información sobre la base de datos, es decir, los metadatos. Este momento es de suma importancia, dado que facilitamos su uso, a otros futuros usuarios que se encuentren interesados en usar nuestra información.

El uso de metadatos como las etiquetas, son de vital importancia en la organización de bases de datos. Veamos un ejemplo: si ingresamos información sobre la Carrera de Comunicación, entre las posibles etiquetas que deberían aparecer se encuentran: “Facultad de Ciencias

Sociales", "Universidad de Buenos Aires" y "UBA". Esto no sólo permite que más adelante podamos integrar trabajos sobre otras carreras/facultades de la UBA, sino también integrar distintos trabajos de otras Universidades o Facultades de Ciencias Sociales, por ejemplo, la Universidad Nacional de Rosario. El trabajo en comunidad (potencia de la red) está en poder integrar todos los datos que se suban. Por esa razón es tan importante el uso de los metadatos.

4.3.2 Opciones de visualización

Respecto a las opciones de visualización, podemos optar por las siguientes:

RELACIONES ENTRE PUNTOS DE DATOS

1) Diagrama de dispersión (Scatterplot). Esta opción juega con los valores para los ejes "x e y" a partir de valores de un cuadro de doble entrada. Tiene un funcionamiento parecido a Gapminder.

2) Gráfico de Matriz (Matrix Chart) Visualización multidimensional de datos a partir de la representación de datos en barras o tortas en varias filas.

3) Diagrama de redes (Network diagram) Las relaciones entre personas o entidades, la conexión a partir de la repetición entre variables, por ej. personas que se conocen y sus conexiones. Esta visualización permite visualizar los nodos centrales y como está conformada la red de relaciones. node and link diagram.

COMPARAR UN CONJUNTO DE VALORES

1) Gráfico de Barras (Bar Chart) Clásico gráfico con el agregado que permite intercambiar entre conjuntos de datos al mismo tiempo.

2) Histograma de bloque (Block Histogram) Esta opción permite ver la distribución de valores numéricos a partir de un conjunto de datos. El eje X está dividido en contenedores que corresponden a los intervalos de valor. Cada dato del conjunto se representa como un bloque y son apilados en cada contenedor mostrando el valor para cada intervalo.

3) Gráfico de burbujas (Bubble Chart) Este gráfico facilita la visualización de grandes conjuntos de datos que difieren en gran cantidad respecto a las magnitudes de valor ingresadas.

SEGUIMIENTO DE ALZAS Y CAÍDAS EN EL TIEMPO

1) Gráfico de seguimiento (Line Graph) Método clásico para observar cambios a lo largo del tiempo. Similar funcionamiento al de N-Gram Viewer.

2) Gráfico de pila (Stack Graph) Al igual que el anterior, con el agregado que la línea abarca toda la zona de valor y permite apilar o situarse por encima del otro valor ingresado.

3) Gráfico de pila para categorías (Stack Graph for Categories) Función compleja del

anterior, utilizado para cuando en el conjunto tenemos sub-categorías.

VISUALIZAR PARTES DE UN TODO

1) Gráfico de torta (Pie Chart) Clásica visualización para porcentajes.

2) Mapa de árbol (Treemap) Visualización funcional para estructuras jerárquicas.

ANÁLISIS DE TEXTO

1) Árbol de palabras (Word Tree) Esta herramienta nos permite visualizar y realizar búsquedas en textos sin estructuras[24], como un libro, artículos, discursos o poemas. Tenemos la posibilidad de elegir una palabra o frase y dar cuenta de los diferentes contextos en que esta aparece. La disposición se da como el tronco de un árbol (la palabra/frase) y todas las ramas son el contexto en que aparece. A partir de esta herramienta tenemos la posibilidad de ver la recurrencia y relevancia de una palabra, como también un gran método de resumen para la lectura de un texto.

2) Nube de etiquetas (Tag Cloud) Esta herramienta permite la visualización de la frecuencia de las palabras en un texto determinado. El tamaño de la palabra corresponde a la cantidad de veces que esta palabra aparece mencionada. Funciona idénticamente al software Wordle, pero con menos funciones para el estilo o diseño en que se muestra el resultado final.

3) Red de frases (Phrase Net) Esta visualización permite diagramar relaciones entre las distintas palabras usadas en un texto. El software utiliza un conjunto de variables predeterminadas para comenzar y dar cuenta del funcionamiento, sobre todo con los conectores más comunes, por ej. “y” o “en”. La visualización final es una red de palabras conectadas a un nodo central que es la palabra que se pidió.

4) Generador de nube de palabras (Word Cloud Generator) Funciona de manera idéntica a la **Nube de etiquetas**, pero esta opción si contiene las opciones que se encuentran en Wordle.

Por último, deseamos destacar que Many Eyes es una gran herramienta en su conjunto, con la posibilidad de poder trabajar sobre datos que tenemos disponibles: ingresamos o subimos nuestra información como así también utilizamos las bases que han usado otros. De esta manera podemos aportar a la comunidad y tener más posibilidades de trabajar con datos cruzados o de otros usuarios interesados en el tema de investigación.

4.4 Voyant Tools

Nombre completo: Voyant Tools

Creador: Voyant Tools, Stéfan Sinclair & Geoffrey Rockwell

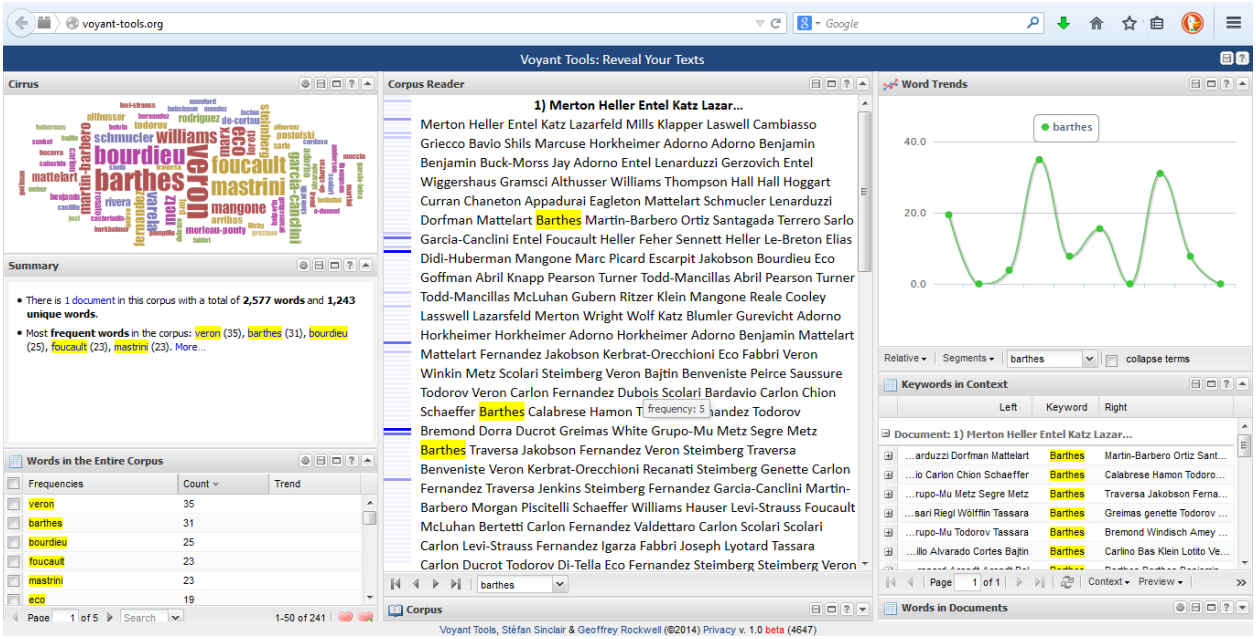
24 Nos referimos al modo de ingreso, se debe copiar y pegar directamente el texto crudo, no hay necesidad de tablas como en los anteriores. Haremos referencia al modo de ingreso en el apartado siguiente.

Link / Enlace al sitio: http://voyant-tools.org/

Tipo de herramienta: Análisis de Texto

Tipo de accesibilidad al usuario: Cerrada.

Este software es simple en su utilización pero complejo en posibilidades de uso. Se define por ser una herramienta basada en la web para leer y analizar textos digitales. Su manipulación es muy sencilla: se copia y pega una cantidad de texto o que se carga un documento. El programa realiza su procesamiento estándar como se ve en la siguiente imagen:



A partir de ahí, podemos complejizar el proceso de ingreso dado que se ofrecen distintos modos de organizar esta información. De uso similar a las posibilidades que ofrece Many Eyes, su diferencia radica en que no existe una base de datos visible para los usuarios.

Su gran ventaja, es el rápido contador de palabras del corpus ingresado y la separación de palabras únicas que aparecen en él. A modo de resúmen, esta herramienta se focaliza en poder mostrar las frecuencias con las que aparecen las distintas palabras.

En el link http://docs.voyant-tools.org/tools/ se podrán encontrar todas las opciones que permite Voyant Tools, pero al repetirse con Many Eyes, haremos una selección de las que nos parecen relevantes:

Líneas de Burbujas (Bubblelines) http://docs.voyant-tools.org/tools/bubblelines/

Visualiza la frecuencia y repetición de una palabra en el corpus ingresados. Cada documento en el corpus se representa en una línea horizontal y es dividida en segmentos de igual longitud.

Cada palabra seleccionada se representa con una burbuja, su tamaño corresponde a la frecuencia con que es mencionada y su ubicación en el segmento está dado por el lugar en dónde aparece en el texto.

Cuadrícula del corpus (Corpus Grid) http://docs.voyant-tools.org/tools/corpussummary/

Esta opción provee una visión general de una gran cantidad de textos a la vez. Se concentra en separar los nombres de cada archivo, el número total de palabras que la componen, la cantidad de palabras únicas que aparecen en el mismo.

Mandala http://docs.voyant-tools.org/tools/mandala

Permite importar archivos de texto para realizar un análisis en la frecuencia y conexión entre los términos. Si cargamos el archivo de un guión de teatro o película, se visualiza la relación y frecuencia entre una palabra y su enunciatario.

4.5 Wordle

Nombre completo: Wordle
Creador: Jonathan Feinberg
Link / Enlace al sitio: http://www.wordle.net/
Tipo de herramienta: Visualizador de texto / Nube de imágenes
Tipo de accesibilidad al usuario: Semi-Abierta

Wordle es la herramienta más simple que hemos investigado. La incluimos en este apartado dado que es una buena entrada para realizar una primer mirada en los procesos de visualización de la información. Consideramos que dado su bajo nivel de complejidad en programación, puede constituirse en el primer paso a explorar para aquel cientista social que comienza a elaborar/construir y trabajar con este tipo de software.

El modo de ingreso de la información es a través de texto libre. El proceso que se desarrolla es el de abrir primero un texto con un procesador de texto, y luego copiarlo para posteriormente pegarlo. La visualización que obtenemos en pantalla está dada por palabras con distintos tamaños que corresponden a la frecuencia en que son mencionadas en lo que hemos seleccionado.

El usuario puede editar la imagen, cambiando colores, ubicación de las palabras e incluso quitar aquellas que no contribuyan al propósito de la visualización. Respecto a los trabajos de otros usuarios, podemos ingresar para ver su trabajo final, pero no editarlo. Por lo que tenemos restricciones a la hora de conocer qué base de datos usaron.

5. Programas de Comunicación #FSoc UBA

5.1 La imagen que comenzó todo

La presente imagen representa la propuesta de **autores** que componen el tronco común de la Carrera de Ciencias de la Comunicación, Facultad de Ciencias Sociales de la Universidad de Buenos Aires. La visualización se realizó a partir de una base de datos construida con los programas de las distintas cátedras disponibles en la web de la facultad, en este caso comunicacion.fsoc.uba.ar y sociales.uba.ar. Los nombres están dispuestos a modo de una nube de tags, donde la relación nombre/tamaño está dada por la cantidad de repeticiones en los programas. El programa utilizado para realizar esta imagen fue Wordle.

En función de los conceptos y herramientas vinculados a las Humanidades Digitales, se intenta indagar cómo se construye el perfil profesional del Comunicador Social, a partir de la conformación de la propuesta académica de la Carrera de Ciencias de la Comunicación Social de la Facultad de Ciencias Sociales en la Universidad de Buenos Aires. A su vez, deseamos observar qué líneas teóricas se van constituyendo en el proceso de formación a lo largo de la carrera. Para comenzar a entender y entrever estos recorridos, nos propusimos relevar qué es lo que se lee a través de revistar la cantidad de autores que circulan por los programas de las distintas cátedras y cómo se relacionan en las distintas materias o años.

Como señalamos, el marco apropiado para abordar estas preguntas es el de las Humanidades Digitales. La cantidad de datos que debemos reunir para poder comenzar a contestar es grande. A

comparación de los trabajos realizados por Manovich o Rosling estamos a una escala mucho menor, pero aún así, para procesar todos los datos reunidos, debimos recurrir a los software de procesamiento mencionados en el apartado 4 “Herramientas”.

Como explican Presner y Johansson en “The Promise of Digital Humanities” estas herramientas y software, han creado nuevos mecanismos que involucran directamente a los estudiantes en investigación de vanguardia. A partir de estas nuevas metodologías en investigación, los mismos se unen a proyectos colaborativos, ansiosos por usar estas herramientas, aplicando conocimientos y habilidades.

Para estos autores, las Humanidades Digitales han redefinido la excelencia de la investigación para los no graduados, permitiendo que jóvenes universitarios hagan contribuciones significativas en campos de estudios avanzados, yendo desde la arqueología y la arquitectura hasta la literatura o historia.

> “Las Humanidades Digitales enseña a los estudiantes a desarrollar las habilidades necesarias para el pensamiento crítico en la era de la información digital, preparándolos para ser participantes activos en la producción de conocimiento y en la construcción de este nuevo mundo” (Presner y Johanson 2009)

Por esta razón, entendemos que es necesario realizar un trabajo formal para presentar en el ámbito de la Universidad de Buenos Aires, adhiriendo nuevas metodologías, acercando recientes desarrollos y herramientas de trabajo al Comunicador Social. Deseamos sumar complejidades de lectura y apertura a la Carrera que me nos ha formado, y que en este momento se encuentra revisando su Plan de Estudios, debatiendo los alcances de sus miradas epistémicas.

5. 2 Armar la base de datos

En nuestra consideración, nos encontramos en el corazón de nuestro núcleo del trabajo. Hemos dedicado gran parte del desarrollo de la presente Tesina en, buscar, reunir, y en muchos casos construir la información necesaria. Estos Datos han sido la base fundamental para poder interpelar al software, logrando posteriormente, diversas y complejas visualizaciones.

El procedimiento realizado fue el siguiente:

A) Se buscaron los programas disponibles en los sitios oficiales de la Carrera de Ciencias de la Comunicación de la UBA o de las distintas cátedras que conforman a la carrera.

La lista de sitios fue la siguiente:

A1) comunicacion.sociales.uba.ar/listadoprogramas.htm Este es el listado oficial de las

materias con los programas según el año. Nos encontramos con dificultades dado que existen muy pocos programas subidos.

A2) comunicacion.fsoc.uba.ar/programas.htm Durante el año 2013, se ha descargado y/o revisado los programas disponibles. Aquí nos encontramos con el problema de que luego de una actualización, el enlace no está disponible.

A3) sociales.uba.ar/?page_id=1409 Buscador de cátedras / profesores de la Facultad de Ciencias Sociales. Se encuentra desactualizado respecto a los recientes cambios o nombramientos de nuevas cátedras.

A4) Para el caso especial del Seminario de Diseño Gráfico y Publicidad, recurrimos al grupo de facebook denominado Facultad[25]: www.facebook.com/groups/169730909716616 donde un usuario cargó el documento del programa actual en el año 2013.

Regresando a los registros encontrados, su criterio de selección ha sido en función del último programa actualizado. En el anexo se puede encontrar la lista completa de las cátedras con su enlace al respectivo programa, detallando el año disponible.

B) Se utilizaron los programas correspondientes al Tronco Común de la carrera.

C) Se utilizaron tanto lecturas obligatorias como sugeridas de cada programa.

D) Se utilizaron solamente los Apellidos de los autores de cada texto que componen los programas. Se evitaron las tildes. En caso de encontrar un apellido compuesto o doble se unieron con un guión "-" Ejemplo: Jesús Martín Barbero es: **Martin-Barbero**

E) No se encontraron los siguiente programas:
- Taller I Cátedra Cortés.
- Comunicación III Cátedra Gassman.

Una vez reunidos todos los programas disponibles se procedió a transcribir en una planilla de cálculo todos los autores divididos por cátedras. Un autor o nombre de cátedra por casillero. Para el orden de ingreso de datos se tomó como guía el plan de estudios que presenta la Carrera de Ciencias de la Comunicación Social de la Universidad de Buenos Aires, de manera tal que la primer materia para la carrera es "Teorías y Prácticas de la Comunicación I" Cátedra Alicia Entel, mientras que la última materia corresponde al "Taller de Comunicación Periodística" Cátedra Rocco Cuzzi.

25 En este momento este grupo se encuentra en modalidad "Secreto" y solo se puede ingresar con aprobación del moderador.

Finalizado el presente proceso, se llegó a consolidar la primer base de datos sobre los programas de las cátedras que conforman la Carrera de Ciencias de la Comunicación Social. El paso siguiente fue comenzar a probar las herramientas disponibles para procesar estos datos.

5.3 Números iniciales: de Abal-Medina a Zweig

Utilizando la herramienta Voyant Tools, podemos discriminar que entre las **45 cátedras** que componen nuestro corpus sobre el tronco común, se reúnen una cantidad total aproximada de **2500 textos**. De ellos, filtramos los autores de textos repetidos, obteniendo un resultado de **1210 autores únicos.**

Los 25 autores más utilizados en los programas son los siguientes:

1. Verón (35)	14. Marx (14)
2. Barthes (32-31)	15. Loreti (13)
3. Bourdieu (25)	16. Mattelart (13)
4. Foucault (23)	17. Schmucler (13)
5. Mastrini (23)	18. Steimberg (13)
6. Eco (19)	19. Rivera (12)
7. Williams (17)	20. Rodriguez (12)[27]
8. Varela (16)[26]	21. Adorno (11)
9. García-Canclini (15)	22. Arribas (11)
10. Martín-Barbero (15)	23. Ford (11)
11. Metz (15)	24. Todorov (11)
12. Fernández (14)	25. Althusser (10)
13. Mangone (14)	

Estos **25 autores** representan un **2% del total de autores (1210 conforman el total)**, y entre ellos suman **407 textos**, es decir, un **16.28%** del total de textos que conforman la propuesta del tronco común. Respecto a la nacionalidad, 13 de ellos son **argentinos** (aún con los falsos positivos), 7 **franceses**, 2 **alemanes**, y finalizando la lista un **español-colombiano**, un **italiano** y

26 Error metodológico (falso positivo) a corregir, dado que “Varela” corresponde tanto a Mirta Varela (7), como a Francisco Varela (2), como a Graciela Varela (7), lo que dejaría a la primera por fuera de los 25 primeros.
27 Error metodológico (falso positivo) a corregir, dado que “Rodriguez” corresponde tanto a María Graciela Rodriguez (8), como a Pablo Rodríguez (1), a Rodríguez F. (1), a María del Carmen Rodríguez (2), a Rodriguez-Esperon (1) y Rodriguez Use (1), lo que dejaría a la primera por fuera de los 25 primeros.

un **galés-británico**.

A partir de 23 de los 25 autores (descartamos a Varela y Rodriguez por errores metodológicos a corregir) realizamos una búsqueda en internet[28], tomando como fuentes sitios oficiales, Wikipedia, artículos o blogs para recuperar información de los mismos. Esto forma parte de un trabajo a futuro[29], donde se buscará armar una base de datos online abierta de autores académicos, de manera tal que cualquiera pueda utilizar esa información en trabajos de esta índole. Las categorías provisorias son las siguientes: Nacionalidad / Edad / Área de Estudios / Corriente de Pensamiento / Título / País donde realizó trabajos / Era de influencia / Región de Pensamiento.[30]

De estos datos, recuperamos información que nos ayuda a completar la imagen construida en Wordle. Es interesante dar cuenta que 11 de los autores se **encuentran fallecidos**: Verón, Barthes, Bourdieu, Foucault, Williams, Metz, Marx, Rivera, Adorno, Ford y Althusser. Por tanto, en el proceso de selección de lecturas por parte de las Cátedras, se leen a protagonistas contemporáneos.

Por otra parte, no es menor observar que entre los 23 autores más leídos, la división regional se da entre latinoamericanos y europeos. No observamos el trazo de estudios de autores norteamericanos.

Si tomamos como categoría el siglo en dónde los autores realizaron su mayor parte de trabajo académico- intelectual observamos:

- Siglo XIX: Marx (1)

- Siglo XX: Verón, Barthes, Bourdieu, Foucault, Williams, Metz, Rivera Ford y Althusser (10)

- Siglo XXI: Mastrini, Fernández, Mangone, Loreti y Arribas (5)

- Siglo XX-XXI: Eco, García Canclini, Martín Barbero, Mattelart, Schmucler, Steimberg y Todorov (7)

Respecto a las áreas de estudios nos pareció apropiado armar una nueva nube de palabras:

28 Este trabajo tiene como premisa tratar de utilizar todas las fuentes disponibles en la web como parte del análisis de los contenidos analizables y que dificultades presenta la utilización de este método.

29 Una de las dificultades encontradas a la hora de armar este trabajo, solamente pudimos dar con este sitio que reune un index de algunos autores: http://thinkexist.com/nationality/

30 En el anexo se podrá ver los datos reunidos hasta el momento.

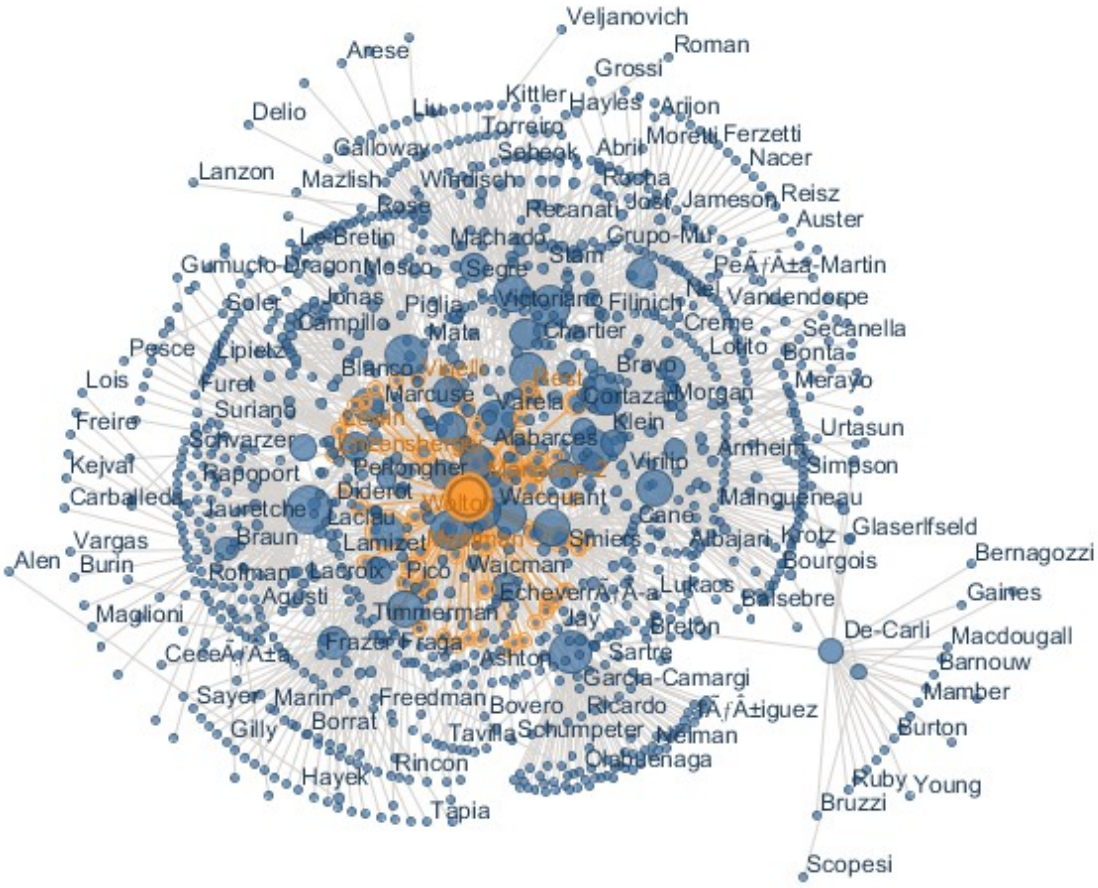
Esta nube se realizó a partir de esta lista[31]:

comunicación 12	derecho-a-la-información 2
filosofía 8	política 2
literatura 8	cine 1
cultura 7	cultura-popular 1
medios 7	derecho 1
semiótica 7	economía 1
sociología 6	música 1
antropología 5	psicología 1
historia 3	radio 1
semiología 3	

De esta lista podemos analizar que, más allá que la materia Comunicación es la principal área de interés en los autores, solamente representa, la mitad del total. Este cuadro da cuenta la presencia de las áreas de las ciencias sociales que estuvieron en el inicio de la carrera. Por último, también queremos dejar establecido que con las 77 palabras que conformar esta lista, las posibilidades de visualización son muy escasas, por lo que esperamos ansiosamente engrosar esta parte del proyecto al resto de los 1210 autores.

31 En el anexo se puede ver la lista completa de autores e intereses académicos.

5.4 Ciudad #FSoc[32]



Esta imagen fue creada con el programa Many Eyes, utilizando su función de **Diagrama de redes (Network diagram)**[33]. La imagen es el producto de nuestra base de datos de autores y cátedras, dispuesta en una tabla de excel en dos columnas en donde en la tabla de la izquierda están los nombres correspondientes a cada cátedra y en la derecha el nombre de uno de los textos que conforman el programa de su materia.

Entel	Adorno
Entel	Adorno
Entel	Adorno
Entel	Althusser
Entel	Appadurai
Entel	Barthes
Entel	…

32 #FSoc es la etiqueta utilizada en Twitter y Facebook para referirse a la Carrera de Comunicación (en ocasiones a las demás carreras) de la Facultad de Ciencias Sociales de la UBA.

33 Se puede consultar aquí http://www-958.ibm.com/software/analytics/manyeyes/datasets/programas-de-comunicacion-catedras/versions/1

Entonces vemos que para la cátedra Entel de la materia Comunicación I, nos encontramos con tres textos de Adorno, uno de Althusser, otro de Appadurai, uno de Barthes y así.

Volviendo sobre la imagen, el tamaño de los círculos es proporcional a la cantidad de menciones en la base de datos, por lo siguiente, los grandes corresponden a los jefes de cátedra, mientras que los más pequeños son los autores de cada programa. En la distribución de esta visualización, en el centro podemos ver la relaciones entre los textos que son más utilizados por las distintas cátedras, mientras que en la periferia los que no tienen relación. Los más evidentes son las cátedras De Carli (Taller II) y Rosenthal (Taller llave Publicidad) Al tener una gran cantidad de nodos con relaciones, el programa nos permite buscar e iluminar los puntos de nuestro interés. En la imagen el punto naranja son las conexiones que tiene la cátedra Mangone de la materia Comunicación II con los distintos autores de su materia.

Esta imagen creada con Many Eyes puede ser el disparador de muchos interrogantes a partir de la observación de las relaciones que se ven en la visualización, pero a la vez nos permite resolver algunas preguntas sobre los autores como haremos a continuación.

5. 5 Marx hasta en la sopa. ¿Cuál es el autor más leído de la carrera?

La pregunta principal que ha sido inmanente en el desarrollo de la presente Tesina ha sido si realmente se leía a Marx cómo se decía en las charlas de los pasillos. Seguramente este es uno de los tantos imaginarios que circulan por nuestra carrera, y nos dispusimos a poner en consideración esta premisa y observar qué nos podían relatar las herramientas digitales.

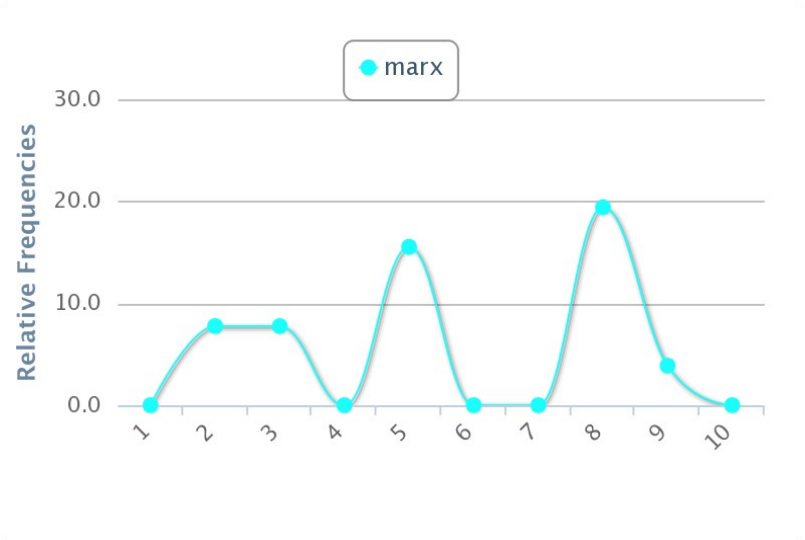
En principio, como mencionamos anteriormente, procedimos a utilizar la herramienta Voyant Tools para contabilizar los autores, y en la lista quedó muy claro que Marx no es el autor más utilizado o propuesto. Se nos hacía difícil comprobar este imaginario:

1. Verón (35)	14. Marx (14)
2. Barthes (32-31)	15. Loreti (13)
3. Bourdieu (25)	16. Mattelart (13)
4. Foucault (23)	17. Schmucler (13)
5. Mastrini (23)	18. Steimberg (13)
6. Eco (19)	19. Rivera (12)
7. Williams (17)	20. Rodriguez (12)
8. Varela (16)	21. Adorno (11)
9. García-Canclini (15)	22. Arribas (11)
10. Martín-Barbero (15)	23. Ford (11)

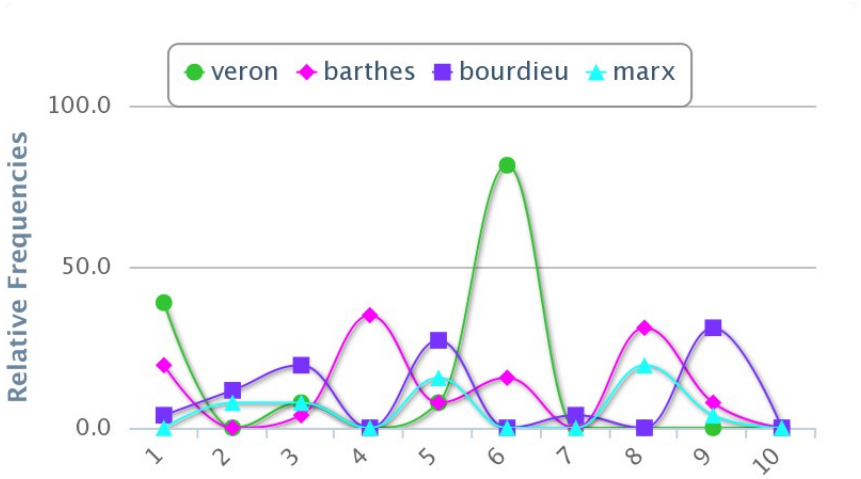
11. Metz (15) 12. Fernández (14) 13. Mangone (14)	24. Todorov (11) 25. Althusser (10)

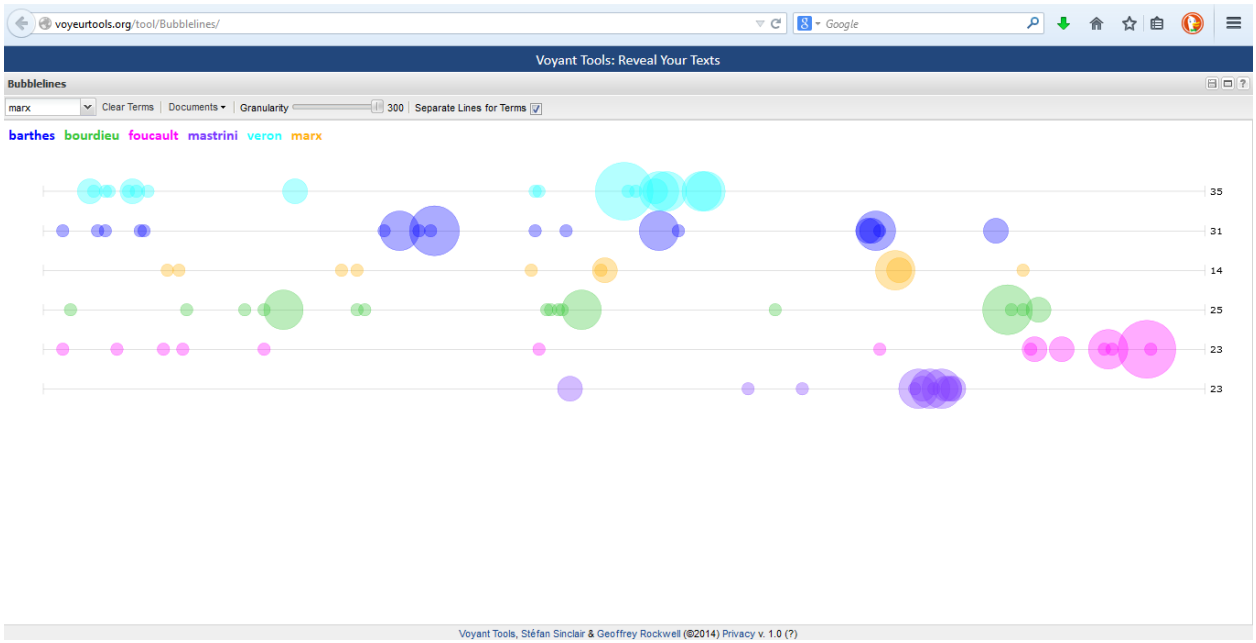
Como muestra la lista, Marx quedó por fuera de los 10 autores más leídos, en igualdad de propuesta con textos de 2 jefes de cátedra de la carrera: José Luis Fernández (Semiótica I) y Carlos Mangone (Comunicación I y II)

Utilizando otra variante del software Voyant Tools, que nos permite ver la frecuencia con la que aparece el autor a lo largo de los programas y materias. Si tomamos como inicio del tronco común a la materia Comunicación I (Cátedra Entel) y el final al Taller llave de Orientación en Periodismo (Cátedra Rocco Cuzzi) el gráfico que ofrece Voyant Tools es el siguiente:



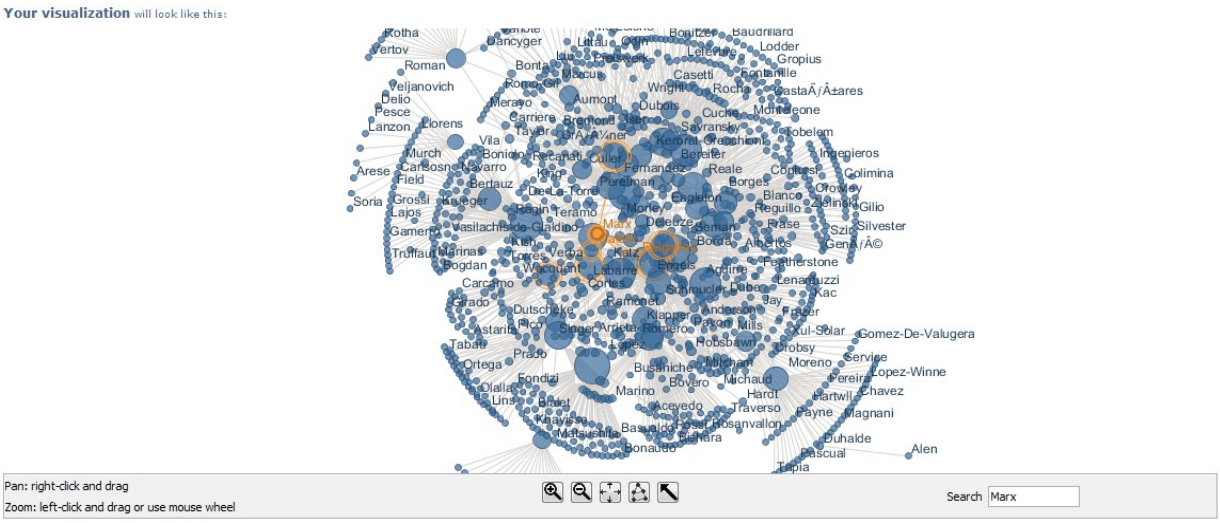
Se observa que la frecuencia del autor a lo largo de la carrera es relativamente alta. Decidimos compararlos con los 3 autores más propuestos de la lista (Eliseo Verón, Roland Barthes y Pierre Bourdieu) para observar que patrones presentan durante la carrera. Para esta comparativa, agregamos la opción de visualización en burbujas, que nos permite establecer puntos fijos de presencia.

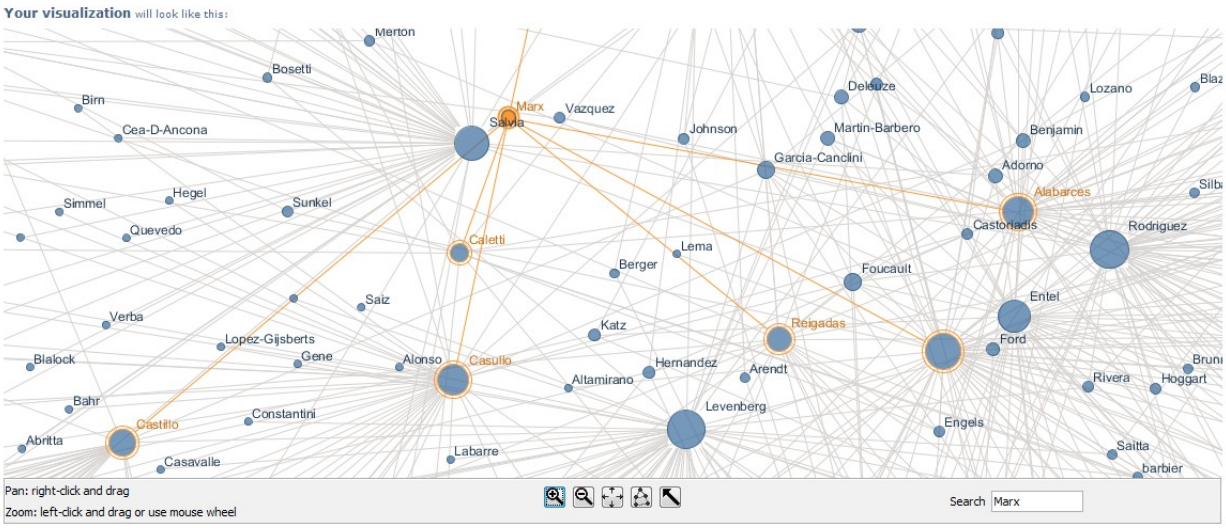




Si bien la cantidad de menciones de Marx es menor, podemos detectar la primera diferencia respecto al autor más propuesto. Eliseo Verón tiene 2 picos importantes, al comienzo y a la mitad de la carrera, mientras que al final desaparece por completo En cambio Marx, como corroboramos con el gráfico anterior, aparece en varios momentos a lo largo de toda la carrera, frecuencia que comparte con Barthes y Bourdieu.

Por último, vemos como los textos de Marx entran en relación con las distintas materias y sus respectivos jefes de cátedra a partir de la herramienta de visualización Many Eyes:





En principio podemos localizar a Marx en el centro de la ciudad, un poco apartado hacia la izquierda la gráfico. Por otra parte, vemos que es usado en 7 cátedras distintas, 6 de ellas muy cercanas dentro del grafo, una de ellas un poco más alejada. La lista es la siguiente:

Cátedra	Materia	Año de la carrera	Cantidad de textos
ROSATO	Antropología	1er Año	1
CASTILLO	Economía	1er Año	3
CASULLO	Principales Corrientes del Pensamiento Contemporáneo	1er Año	1
REIGADAS	Principales Corrientes del Pensamiento Contemporáneo	1er Año	1
MANGONE	Comunicación 2	3er año	1
ALABARCES	Seminario de Cultura Popular y Cultura Masiva	4to Año	1
CALETTI	Teorías y Prácticas de la Comunicación III	5to Año	5

A partir de este cuadro no debería sorprendernos entonces el gráfico de frecuencia presentado en primera instancia, efectivamente los textos de Marx aparecen a los largo de la carrera, en materias de 1er, 3er, 4to y 5to año, teniendo su pico más alto en Comunicación III de la cátedra Caletti quien utiliza 5 textos.

Podemos continuar trabajando y preguntando sobre estos datos, pero en esta instancia podemos fundamentar porqué efectivamente hay un imaginario en el cual Marx es uno de los autores más leídos: En principio dimos cuenta que en la totalidad de los programas de comunicación, Marx se ubica en el puesto 14 en una lista aproximada de 1200 autores. Al observar las relaciones entre autores y cátedras, vemos que Marx se encuentra en el núcleo de la red de relaciones, y de estas cátedras, vemos que las mismas están distribuidas a lo largo de toda la carrera, lo que aumenta las posibilidades de encontrarse con este autor en algún momento. Otro argumento para seguir trabajando sobre este imaginario, y que podremos comprobar una vez finalizada la lista de autores con sus intereses y corrientes de pensamientos, tiene que ver con las referencias e influencias, es decir, académicos como Bourdieu, Althusser o Gramsci que forman parte de la tradición marxista.

5.6 Conclusiones provisorias sobre Programas de Comunicación #FSoc UBA

A partir de lo expuesto podemos realizar las siguientes observaciones: dimos cuenta que las herramientas o software de procesamiento de datos facilitan el recorrido y abordaje de grandes cantidad de datos. No sólo nos permiten responder rápidamente preguntas, sino que también nos habilitan a realizarnos nuevas preguntas a partir de las visualizaciones realizadas. Es decir, las respuestas no son el punto final de la investigación, sino el comienzo de un nuevo recorrido por la base de datos, ingresando por otra puerta.

Por otra parte, como en todo proceso de investigación, el armado de la base de datos es vital para este tipo de trabajos. En este sentido, se realza la capacidad de las herramientas de poder rastrear y recuperar los datos (como Voyant Tools o Wordle) sin la necesidad de transcribir palabra a palabra, y que la misma herramienta de opciones de estructuración o visualización de los datos.

Por último, respecto al párrafo anterior, nos encontramos con una dificultad importante a la hora de reunir la información necesaria para nuestra base de datos. La gran mayoría de los programas no estaban actualizados o disponibles en los sitios oficiales de la carrera, muchos de

ellos tampoco se encontraban en los sitios de las cátedras, y algunos no disponen aún de espacios en internet para consulta de los alumnos-usuarios. Sin embargo, entendemos que esto no implica un problema particular con el objeto de estudio, sino que representa una problemática en general con el trabajo del cientista social que se introduce a las Humanidades Digitales. Se produce mucha información, pero gran cantidad de ella no se encuentra estructurada, lo que hace más relevante este tipo de trabajos que tienen como objetivo procesar la información existente y actualizar la que aún no está disponible.

6. Conclusiones

Las Humanidades Digitales tienen un vasto campo de aplicación dentro de las Ciencias Sociales. En la presente Tesina de trabajo hemos intentado trazar un recorrido dentro de esta gran constelación, observando desde nuestra posición las distintas características y proyectos que la componen.

Hemos observado cómo desde los comienzos de las Humanidades Digitales, los trabajos sobre digitalización de archivos tenían en cuenta a la tecnología como herramienta facilitadora de procesos, para luego comprender que la mismas herramientas se habían transformado en objeto de investigación. Estos softwares son una gran alternativa para abordar la cantidad de información generada que supera ampliamente nuestra capacidad de lectura.

Los proyectos o trabajos dentro de las Humanidades Digitales implican interdisciplinariedad, colaboración a escala global en tiempo real, gran cantidad de datos, apertura a las grandes bases de datos, metodologías que involucran nuevos lenguajes, distintas maneras de categorizar la información, y la intervención de software para procesar el gran entramado descrito. En su conjunto, volvemos a destacar que las herramientas seleccionadas en la presente Tesina, son aplicables solamente al trabajo con textos. Es por ello que toma particular interés, los trabajos que Lev Manovich realiza al analizar imágenes a partir de imágenes, creando así, un nuevo lenguaje de análisis superador.

Por otra parte, este trabajo considera que el nuevo perfil del cientista social debe incorporar, en algún grado, conocimientos mínimos de programación, estar dispuesto a experimentar con herramientas digitales, ser capaz de trabajar con datos y tener habilidades para recabar información cultural. (Hurst, 2012) La creación y producción de conocimiento de manera colectiva y en red permite una alternativa que se suma a la actual formación académica.

La apertura que implica este nuevo marco de investigación, tanto para estudiantes como graduados, facilita el análisis y producción universitaria. La Humanidad Digital, constituye una compuerta informacional abriendo un camino al futuro comunicador social. Un nuevo campo de producción se está desarrollando, sumado a los conocimientos y habilidades académicas presentes.

El trabajo realizado en http://programascomunicacion.tumblr.com/ obtuvo como resultados iniciales la visibilidad de los autores que tienen una relevancia especial en la composición de nuestra carrera. A partir de la misma, se puede visibilizar, las líneas teóricas presentes en el Perfil

Profesional del cursante de Comunicación, dando cuenta de una mayoría de autores europeos y latinoamericanos. La escuela francesa se encuentra presente entre los 5 autores más propuestos en los programas que componen el tronco común, aunque el autor más utilizado, Eliseo Verón, aparece al comienzo y al mitad de la carrera. El resto (Barthes, Bourdieu y Foucault) mantiene una regularidad a lo largo de los años.

Si bien, por problemas de acceso a la información (disponibilidad de programas en la web), no se pudo realizar un trabajo diacrónico sobre la composición de los autores, vemos que la "imagen" actual de nuestra carrera tiene como tema principal los estudios en **Comunicación**, aunque la presencia e influencia de las áreas fundadoras (Filosofía, Literatura, Sociología, etc.) todavía se encuentran presentes.

A partir del diagrama de redes realizado con el software Many Eyes, denominado **Ciudad #FSoc**, hemos podido visualizar la relación que se da entre las materias/cátedras y los autores. La disposición permite observar en principio la gran concentración de autores en el centro con el conjunto de las cátedras, pero a la vez, la periferia presenta una opción interesante de cátedras que no tendrían relación alguna con el común de los autores utilizados por el resto de las materias. Entendemos que este punto debe ser profundizado y considerar más lugares de abordaje sobre las distintas relaciones que se presentan y proyectan.

Las herramientas de procesamiento de información junto a el software permitió develar diversos aspectos del proceso de formación del estudiante en Ciencias de la Comunicación. Del análisis, surge no solo la fuerte presencia de la Escuela Francesa, sino también, la existencia en los Programas de Estudio de los Titulares de Cátedra en el mismo proceso de lectura.

En el actual proceso de reflexión sobre el nuevo Plan de Estudios, conocer y bucear en la Ciudad #Fsoc, puede permitir tomar algunas decisiones sobre las adecuaciones didácticas, como así también metodológicas que se llevan adelante. Por ejemplo, en las dinámicas que se organizan en el diseño de las correlaciones verticales como en los diálogos horizontales.

Esperamos que la presente Tesina, nutra el debate sobre estos procesos decisorios. No estamos realizando una crítica a la manera en que se produce conocimiento dentro de la Universidad. Al contrario; estamos tratando de dar cuenta del estado actual de una porción de las Humanidades Digitales y proponemos que debe buscarse en los próximos años una convivencia entre estas dos maneras de producir conocimiento. Las preguntas están abiertas y son muchas, pero estamos seguros que en las Humanidades Digitales se hace camino al producir, procesar, visualizar, preguntar y andar.

7. BIBLIOGRAFÍA

- ALVARADO, Rafael (2011) *The Digital Humanities Situation* Disponible en bit.ly/iyaOGW

- ANDERSEN Chris (2008) *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, Wired Magazine Disponible en bit.ly/LOU8

- BERRY, David (2011) *Humanidades Digitales: Primera, segunda y tercera ola*, en stunlaw: A critical review of politics, arts and technology. Traducción de Julio Alonso y Fernanda Ibañez. Disponible en bit.ly/17H6XtO

- BURDICK, Anne; DRUCKER Johanna; LUNENFELD, Peter; PRESNER Todd y SCHNAPP Jeffrey (2012) *Digital_Humanities*, Massachusetts Institute of Technology
Disponible en bit.ly/13x8AZC

- BORDA, Libertad (2012) *Bettymaniacos, Luzmarianas y Mompirris: El Fanastimo de los foros de telenovelas latinoamericanas,* Tesis Doctoral en Ciencias Sociales, Facultad de Ciencias Sociales, Universidad de Buenos Aires

- EL PEZ MARTILLO (2007) *Constelaciones,* Espejo de la realidad
Disponible en bitly.com/IyvLf6

- HURST, Marcus (2012) *Google contratará a más de 4.000 humanistas en los próximos años*, Yorokubu, España Disponible en bit.ly/x0Zs4U

- JENKINS, Henry (2006) *Convergence Culture: Where Old and New Media Collide*

- LATOUR Bruno (1991) *Nunca fuimos modernos: ensayos de antropología simétrica*

- LEVY, Pierre (1994) *Inteligencia colectiva, por una antropología del ciber-espacio*.

- MANOVICH, Lev (2011) *Un millón de páginas de manga,* en software studies iniciative. Disponible en bit.ly/1ojDrlg

- MANOVICH, Lev (2011) From *reading to patterns recognition* Disponible en bit.ly/RX5trm

- MANOVICH, Lev (2012) *Epistemología del Software*. Disponible en bit.ly/EpistemologiaSoftware

- MANOVICH, Lev (2012) *¿Cómo comparar un millón de imágenes de manga?*, en BERRY, David Understanding Digital Humanities, Palgrave. Disponible en bit.ly/1i87QOL

- MANOVICH, Lev (2013) *Software takes command* Disponible en bit.ly/175Fz4w

- MORETTI, Franco (2000) *Conjetures on World Literature,* New Left Review 1 ed. Enero - Febrero Disponible enbit.ly/Sqzfz7

- MORETTI, Franco (2007) *Graphs, Maps, Trees: Abstract Models for Literacy.*

- PISCITELLI, Alejandro (2012) *Las "Digital Humanities" y como pensamos en la era de la análitica cultural*, Filosofitis.com.ar Disponible en bit.ly/WPZ6TZ

- PISCITELLI, Alejandro (2013) *¿Cómo definir a las humanidades digitales? ¿O es mejor no definirlas?* Disponible en bit.ly/1e37rjz

- PRESNER, Todd y JOHANSON, Chris (2009) *The Promise of Digital Humanities. A whitepaper.* Universidad de California. Disponible en bit.ly/1fJRRF2

- ROSSI, Aníbal (2013) *Cultoronomía para todos. Aprendiendo a usar operadoradores en N-Grams Viewer*, en Cátedra Datos Web. Disponible en http://bit.ly/1lYyFaE

- SUÁREZ, Juan Luis Suárez (2013) *El Humanista Digital.* Disponible en bit.ly/14u1p7i