

Análisis de la producción institucional a partir del uso de herramientas de data visualización

^{1,2} Dra. Marisa R. De Giusti marisa.degiusti@sedici.unlp.edu.ar
¹ Lic. Ariel J. Lira alira@sedici.unlp.edu.ar
¹ Dr. Gonzalo L. Villarreal gonzalo@prebi.unlp.edu.ar
¹ DCV Lucas E. Folegatto lucas@sedici.unlp.edu.ar



¹ Proyecto de Enlace de Bibliotecas - Servicio de Difusión de la Creación Intelectual (PREBI-SEDICI), Universidad Nacional de La Plata, La Plata (1900), Argentina
² Investigador independiente, Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC), Proyecto de Enlace de Bibliotecas

MARCO

- La producción intelectual académica y científica de las instituciones universitarias se genera gracias al trabajo de los docentes, investigadores y becarios en sus correspondientes unidades académicas, laboratorios, centros e institutos de I+D. Esta producción se plasma principalmente en artículos científicos, tesis de grado y de posgrado, publicaciones en congresos, libros y material de aprendizaje.

Los **repositorios institucionales** tienen como fin primordial la preservación y divulgación del conocimiento generado en sus instituciones, y para ello alojan y exponen estos materiales, y brindan diversos servicios a partir de ellos. **SEDICI**, el repositorio institucional de la **Universidad Nacional de La Plata**, reúne toda la producción de la UNLP y no se limita exclusivamente a documentos científicos sino que incluye también trabajos de extensión, normativas y ordenanzas, audios y videos, entre otros. Todo esto es preservado y difundido bajo **licencias abiertas**, promoviendo así su uso en otros ámbitos y maximizando el impacto y la visibilidad de la producción de las unidades académicas. En otras palabras, SEDICI intenta ser una ventana al mundo desde donde puede accederse a todo lo que se genera desde la UNLP, y así devolver a la sociedad parte del esfuerzo invertido en la universidad pública y gratuita.

SEDICI POSEE MÁS DE
45mil
 RECURSOS EN ACCESO ABIERTO
 que pueden accederse desde
 web - OAI-PMH - RSS/Atom

Cada ítem en **SEDICI** cuenta una gran cantidad de metadatos (en total hay más de 60, aunque no todos se aplican a todos los ítems, ya que esto depende de su tipo, formato, etc). Entre estos metadatos, se considerarán para este trabajo aquellos que brindan información útil para establecer relaciones plausibles de ser analizadas: año de publicación, área temática, unidad académica, tipo de material, etc; como además los metadatos normalizados que pueden brindar datos adicionales a la hora de mostrar la información al usuario: "idioma" y "texto completo" (fulltext, boolean).

OBJETIVOS

A partir del contenido del repositorio **SEDICI** se propone generar una visualización que permita analizar el **volumen, alcance e impacto** de la producción científica de la universidad. Entre otras cosas se buscará encontrar la mejor manera de evidenciar las cantidades de material por unidad académica, destacando aquellas que aporten mayor volumen de material en acceso abierto y a texto completo, y la cantidad de visitas por unidad académica, para evaluar **cuáles tienen mayor impacto en la visibilidad de la UNLP**.

Es por esto que se consideró como objetivo interpretar cuál es el alcance y crecimiento de la producción de ese conocimiento a los fines de entender más profundamente la variedad de contenidos. Por otro lado, se proveerá a los visitantes del repositorio una **herramienta de visualización** que les permita lograr una idea del volumen, tipología y heterogeneidad de contenidos albergados en el mismo.

Una vez puestos en evidencia, estos datos pueden disparar diversos análisis sobre la información obtenida. Resulta en particular relevante la posibilidad de visualizar las relaciones entre las diversas unidades académicas con sus contenidos en el repositorio (áreas temáticas que abarcan, tipos de ítems), así como también evidenciar la evolución en el tiempo de estas interrelaciones.

Como valor agregado, la visualización de los datos de un repositorio ayudará al usuario a tener una noción más clara y mensurada de la variedad de contenidos que alberga.



El cerebro procesa una imagen **60.000** veces más rápido que las palabras



El **90%** de la información procesada por el cerebro es **visual**



Recordamos mejor imágenes y gráficos que textos

“

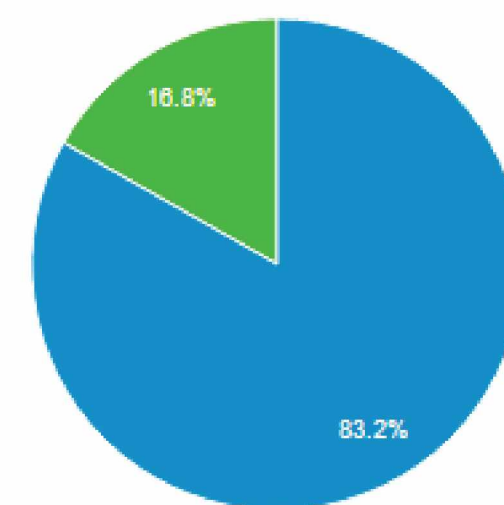
Hoy generamos cada 2 días tanta información como habíamos generado hasta 2003

Eric Schmidt



hace 20 años sólo teníamos 130 websites, Google no existía y teníamos que pagar por una cuenta de e-mail
onesend.designly.com

■ New Visitor ■ Returning Visitor



Otro de los objetivos es lograr que se incremente la frecuencia de reingreso en el repositorio.

Porcentajes históricos hasta octubre 2015

METODOLOGÍA

En esta instancia es cuando se decide

- Qué **estructuras visuales** resultan adecuadas
- Qué **atributos** se codificarán espacialmente, de qué manera.
- Qué **atributos gráficos** se utilizarán para codificar los atributos de las Tablas

“

El valor más importante de una imagen es que deja en evidencia lo que no esperábamos ver

John Turkey



Filtrado

Los datos crudos deben ser filtrados a partir de:

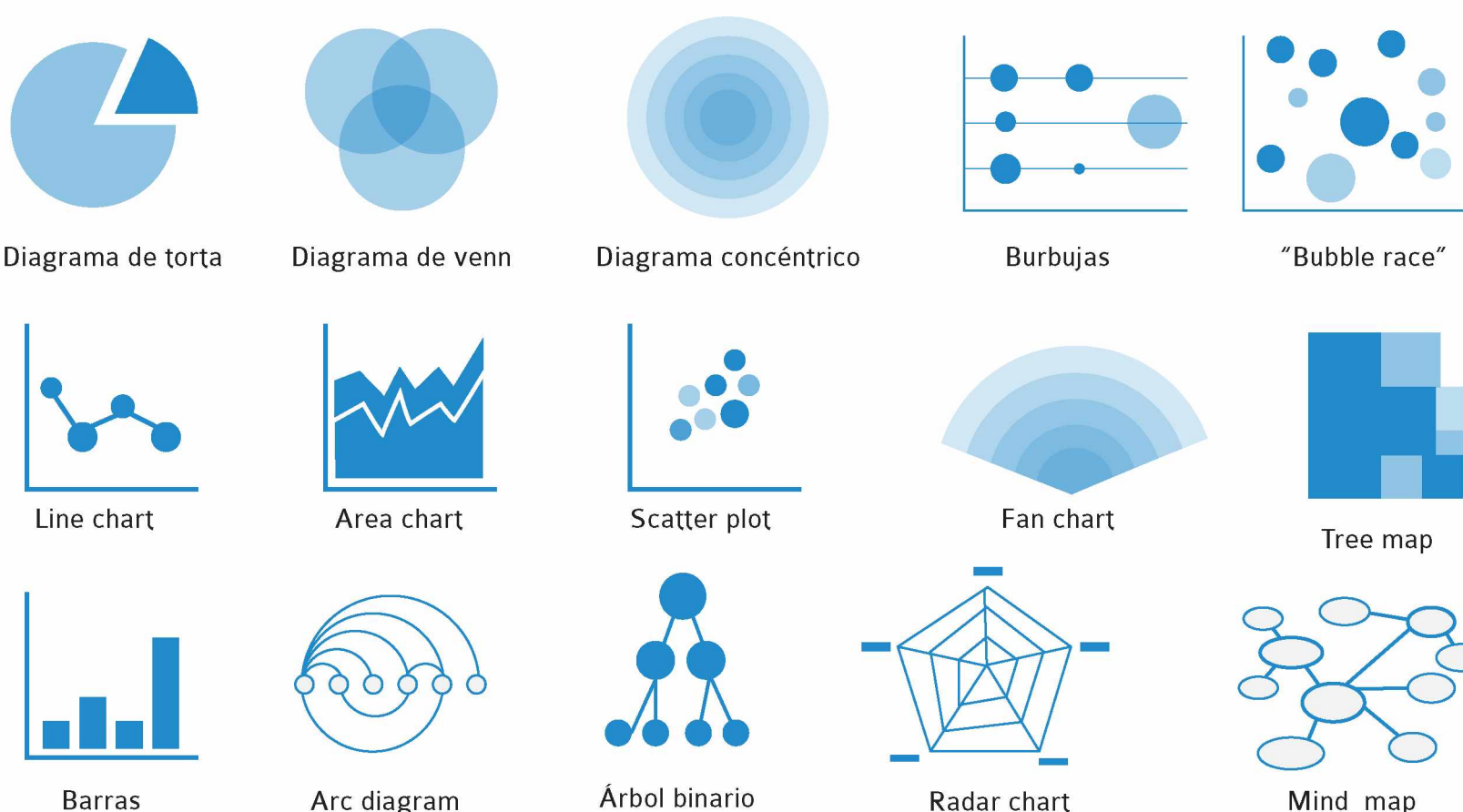
- Origen:** para limitar el análisis a las unidades académicas y al rectorado. En consecuencia, se excluirán colegios, así como también instituciones externas vinculadas a la UNLP
- Tipo:** para acotar el análisis producción científico-académica textual: libros, tesis y tesinas, artículos, trabajos en congresos. Se excluirán, audios, objetos de museos, etc.

Normalización

Será necesario aplicar transformaciones o adicionar información a fin de:

- Determinar la pertenencia de los autores a una UID
- Asignar UA a los trabajos, a partir de las UID (en muchos casos este dato ya está, pero no en todos)
- Determinar cuándo un trabajo pertenece a más de una UA (en principio, a partir del origen de las UIDs)

EVALUACIÓN DE MODOS DE REPRESENTACIÓN VISUAL



IMPLEMENTACIÓN

Herramientas en evaluación

- Visual.ly** <http://visual.ly/>
- D3 - Data Driven Documents** <http://d3js.org/>
- Tableau** <http://www.tableau.com/>

“

Overview first, zoom and filter, and details on demand

Ben Shneiderman



Se analizará la teoría en lo relativo a criterios de diseño aplicado a visualización de datos complejos, así como también las posibilidades que ofrecen las nuevas tecnologías en esta materia. Se seleccionarán y adaptarán los recursos y modalidades más adecuadas y finalmente se buscará una implementación de los resultados de este relevamiento con tecnologías web HTML5, CSS3 y Javascript.