

# Fuentes primarias: Grandes datos, IA contra la enseñanza humana

## Los portales de transparencia

Víctor Manuel Martínez Martínez  
Universidad Pedagógica Nacional  
vicm3@upn.mx

# Iniciemos por el principio

- Que hace que uno se plantee este problema
- La practica profesional
  - Docencia Licenciatura
    - Uso de hoja de calculo
    - Graficas sencillas
    - Fuentes de primera mano

# Docencia

- Licenciatura en Administración Educativa
  - Originalmente con datos creados ex profeso
    - Población, hombres, mujeres
    - Conteos, estatura, peso, edad
    - Medidas de tendencia central (media, moda, mediana)
- ¿Como obtener datos más representativos para los estudiantes?
  - En lugar de darles los datos
    - Que los construyan
    - Que los encuentren
- ¿Quien estudia esto, quien podría tener estos datos?
  - INEGI, CONAPO
  - SEP

# Una pregunta eje

- ¿Es cierto que hay más mujeres que hombres en nuestro país?
- Excusa para enseñar
  - De lo general,
    - Estados Unidos Mexicanos
    - 32 Entidades federativas
    - Estado de México, Ciudad de México
    - Universidad Pedagógica Nacional
    - Licenciatura
  - El problema los dos primeros en INEGI
  - ¿y los dos últimos?

# ¿Y los datos?

- Iniciar un nuevo conteo
  - Servicio social
  - Inventar la rueda
- Solicitar a INFOMEX
  - ¿No deberían ser públicos?
- !¿Que hacer?!
  - Sabático

# INFOMEX

- Se solicitó la información
  - Y se recibió en digital
- Pero de verdad la Universidad ¿no tiene estos datos?

# Año sabático 2019

- Probar y equivocarse
  - Big Data, Web Scraping, ML
  - Abordar de forma equivocada el tema
    - Errar es de humanos, herrar de herreros
      - Pero encontramos el camino
- Construir un material didáctico, que sea un REA
- !Terminamos pensando en una materia optativa!

# Que queremos

- Que los estudiantes y nosotros podamos obtener datos sobre la propia Universidad
  - Matricula, egresados, ingreso, reingreso, becas, etc.
- Tarea terriblemente tediosa y mecánica
- ¿Lo podrá hacer una maquina?



# Un paréntesis

- En 2017 para el EDUSOL
- Revisamos varios Portales de Transparencia
- Encontramos que esta información es accesible de mayor a menor en:
  - UNAM → UAM → UPN → UACM → UnADM
- Agenda estadística
- Transparencia
  - ¿Han intentado saber cuantos estudiantes hay en UACM?


# Portal de Obligaciones de Transparencia

- Son uniformes en cuanto a diseño
- No lo son en cuanto a que información proveen
- Son tan diferentes y complicados, que nos dio para planear un nuevo ejercicio
  - Cómo buscar información en los portales
    - Datos a buscar UNAM, UAM, UPN, UACM, UnADM
    - Profesores tiempo completo
    - Matricula
    - Presupuesto anual

Portal de Transparencia Unam x +

Not secure | transparencia.unam.mx

Incognito



Universidad Nacional Autónoma de México

Portal de Transparencia Universitaria

Estudiar en la UNAM

Personal Universitario

UNAM transparente y abierta

Normativa

Gestión UNAM

Normas en materia de transparencia, acceso a la información y archivos

Solicitudes de Información

SIPOT

Información pública prevista en el Reglamento de Transparencia de la UNAM

2986 de los cuales:

1531	640	870	5,766
Libros	Libros electrónicos	Publicaciones periódicas (fascículos)	Otras publicaciones


Fecha de actualización: abril 2019



UAM. Universidad Autónoma x +

Not secure | transparencia.uam.mx/index.html

Incognito



Casa abierta al tiempo

UNIVERSIDAD AUTÓNOMA METROPOLITANA

Inicio

Obligaciones de Transparencia

Solicitud de Información

Casa Abierta y Transparente


SIPOT

Transparencia Proactiva

Transparencia UAM

La Universidad Autónoma Metropolitana con objeto de garantizar el libre y pleno ejercicio del derecho de acceso a la información el cual se precisa, asegura y preserva a través de Ley General y Federal de Transparencia y Acceso a la Información Pública, y el Reglamento para la Transparencia de la Información Universitaria, habilita este Portal de Transparencia para proporcionar a la comunidad universitaria y a la sociedad en general, el acceso sencillo a la información que genera o posee la institución.

De esta forma la UAM refrenda su compromiso con la transparencia y la rendición de cuentas, valores que se consagran en su Ley Orgánica.




© Universidad Autónoma Metropolitana 2019


TRANSPARENCIA x +

upn.mx/index.php/transparencia

gob.mx Trámites Gobierno

UPN INICIO CONOCE LA UPN UNIDADES UPN AJUSCO ADMISIÓN COMUNIDAD TRANSPARENCIA CONGRESO UNIVERSITARIO

 **EDUCACIÓN**  
SECRETARÍA DE EDUCACIÓN PÚBLICA





 **UNIVERSIDAD PEDAGÓGICA NACIONAL**

Está aquí:

Transparencia

Google Búsqueda personalizada

Sitios y Enlaces de transparencia

-  Sistema de Portales de Obligaciones de Transparencia **SIPOT**
-  **Contraloría Social**  
Información sobre PFCE Y PRODEP en la UPN
-  **APF-UPN**  
Informe de Rendición de Cuentas de la Administración Pública Federal UPN
-  **SEPE**  
EVALUACIÓN EXTERNA **SEP**  
PROGRAMAS FEDERALES

Contenido:

[Normatividad en Transparencia](#)

Unidad de Transparencia > x +

transparencia.uacm.edu.mx

ARTÍCULO\_134\_abc

**UACM**  
Universidad Autónoma de la Ciudad de México  
*Nada humano me es ajeno*

Portal de Transparencia 

Login

**OBLIGACIONES DE TRANSPARENCIA**

- ARTÍCULO 121**
- ARTÍCULO 134**
- ARTÍCULO 143**
- ARTÍCULO 146**
- ARTÍCULO**

**UNIDAD DE TRANSPARENCIA UACM**

**RESPONSABLE DE LA UNIDAD DE TRANSPARENCIA:**  
Lic. María Alejandra Becerra Gutiérrez

**DOMICILIO:**  
Dr. García Diego No. 170, Planta Baja, Colonia Doctores, Alcaldía Cuauhtémoc, C.P. 06720, Ciudad de México

De lunes a viernes de 10:00 a 15:00 horas

**TELÉFONO:**  
1107-0280 Exts. 16410 y 16411

**CORREO ELECTRÓNICO:**  
oipuacm@uacm.edu.mx

DÍAS INHÁBILES UT-UACM



# Transparencia

“En el Gobierno de la República hemos adoptado un enfoque proactivo en materia de transparencia. En esta sección ponemos a tu disposición información que contribuye a una gestión pública abierta y que rinde cuentas.”

Última modificación:  
Jueves 15 de diciembre de 2016 a las 11:00 hrs

## Enlaces

- [Portal de Obligaciones de Transparencia](#)
- [Sistema Infomex](#)
- [INAÍ](#)

## ¿Qué es gob.mx?

Es el portal único de trámites, información y participación ciudadana. [Leer más](#)

- [English](#)
- [Temas](#)
- [Declaración de Accesibilidad](#)
- [Aviso de privacidad integral](#)
- [Aviso de privacidad simplificado](#)
- [Términos y Condiciones](#)
- [Política de seguridad](#)
- [Marco Jurídico](#)
- [Mapa de sitio](#)

## Contacto

- [Denuncia contra servidores públicos](#)

# Descubrimos

- La institución y sus portal(es) de Obligaciones de Transparencia
- El Portal de Obligaciones de Transparencia y el ahora Sistema de Portales de Obligaciones de Transparencia (SIPOT)
- Quien diseña el sistema en el Instituto Nacional de Transparencia (INAI) no es el mismo que termina llenando los reportes en el organismo obligado
- Peor aún la implementación del portal institucional normalmente la hace un tercero que no tiene relación con los dos de arriba
  - Que la dificultad de consulta no es debido a malicia, es prisa por cumplir con las obligaciones
  - Que mucha información esta disponible para cumplir las obligaciones, no porque sea útil

# Horrores técnicos

- De parte de UPN hay:
  - <http://transparencia.ajusco.upn.mx> PHPNuke
  - <http://sipot.upnvirtual.edu.mx> Joomla
- De parte de INAI
  - <http://portaltransparencia.gob.mx>
  - <https://consultapublicamx.inai.org.mx>
- !Todos son diferentes! Y en ocasiones un sistema te lleva a otro o te avisa que la información la puedes encontrar en el otro sistema, ejemplo: nuevo ingreso
- Pero no vayamos más lejos matrícula UACM...

# Web scraping

- Ojo en algunos países es ilegal
- Otro nombre para espejo (mirror)
- Se recomienda que haya alguna regularidad en las páginas
  - Python
- En nuestro caso
  - wget y httrack



# Horrores técnicos...

- UPN usa Joomla para sus portales y phocadownload, genera URL como este:
  - `http://sipot.upnvirtual.edu.mx/index.php?option=com_phocadownload&view=category&download=1940:se14&id=381:2018&Itemid=310`
- Todos los documentos son PDF
- Por eso Google y demás no nos ayuda, las páginas son generadas “dinamicamente” y no tienen un destino real, fuerzan la descarga, entonces no se encuentran en los índices de búsqueda
  - Solución
    - PDF → xpdf/pdftotext → Texto plano
    - PDF → xpdf/pdftoimages → imagen → gocr → Texto plano
    - En el escritorio Recall

# No se puede con ML/Big Data

- <<“Big data” often means “data produced by someone else” and there’s lots of it>> Tom Smith (2019)
- Sociólogo que puede hacer bastantes cosas con la computadora
  - “Cliff, he’s not much of an astronomer, but what a computer hacker [...] Cliff’s not much of a programmer, but what an astronomer!”,  
The Cuckoo’s Egg
- Trate con xpdf → Texto → Wordcloud, problema, “no se habla español”
  - Seminario de redes IIEc (gran hack)
- Sin usar herramientas de la nube, porque la nube, **es la computadora de alguien más**

# No se puede con ML/Big Data

- Mi trabajo es enseñar a futuros Administradores Educativos a obtener los datos y la información, el nivel de complejidad resulto muy alto
  - No es que no se pueda, yo no puedo en el estado actual y no encontré las herramientas para hacerlo posible para un estudiante de LAE UPN
  - Hackear la cosa de ML a Aprendizaje humano

# La solución

- Encontrar el sistema de gobierno de la IES
- Quien reporta, en este caso organismo desconcentrado, la Comisión Interna de Administración (CIA)
- Gobierno, ciclos trimestrales
  - Resultado, cuatro informes al año, de 2004 a 2019, es decir 60 informes...
- El objetivo era fortalecer contenidos de dos materias, se nos ocurrió una materia optativa

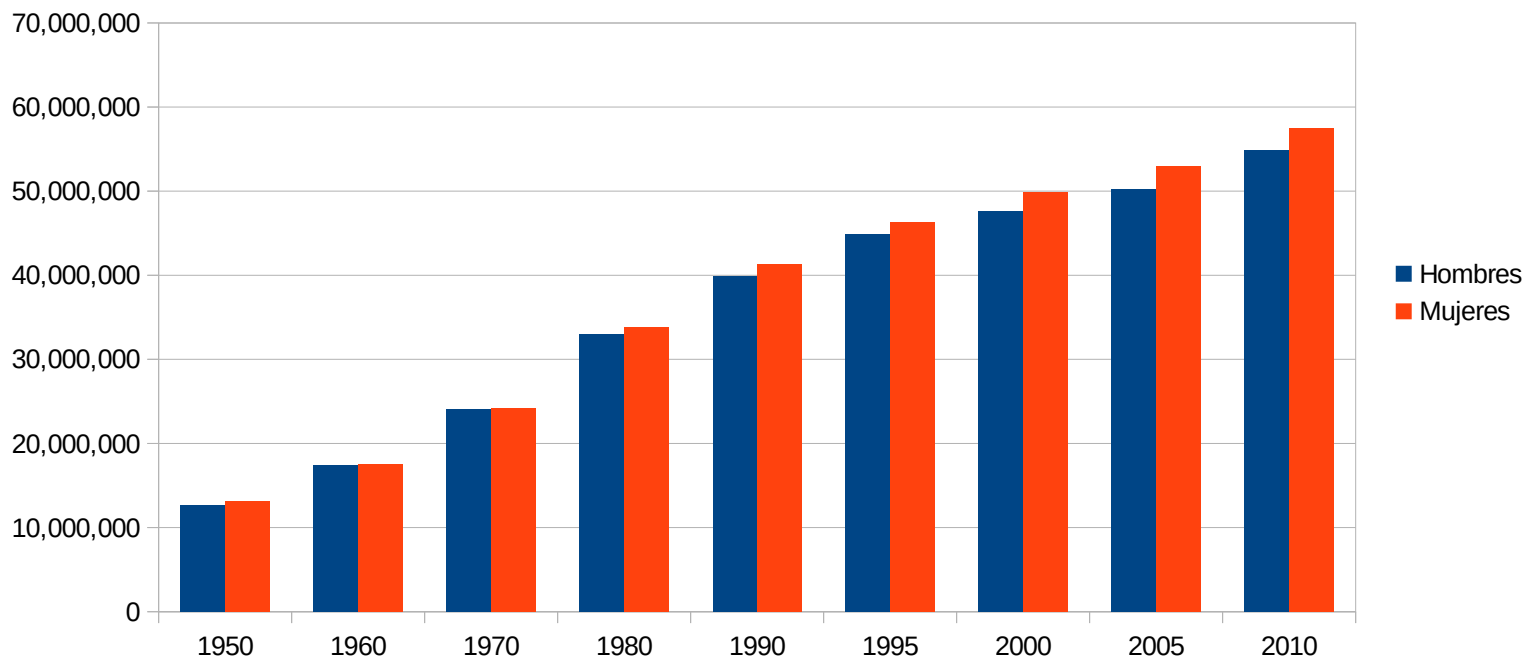
# La solución...

- Aprendizaje basado en problemas
- Queremos que el estudiante razone que busca, no darle una receta
- Ejercicios que den pistas y no la respuesta
- Que pueda aplicarla a otros datos otras instituciones
  - Con buenos resultados en IES

# ¿Y la pregunta eje?

Distribución por edad y sexo

Población total según sexo, 1950 a 2010



Fuente: INEGI <https://www.inegi.org.mx/app/tabulados/default.html?nc=mdemo01>

# Gracias por su atención

- ¿Preguntas, comentarios?

Víctor Manuel Martínez Martínez

[vicm3@upn.mx](mailto:vicm3@upn.mx)