

Herramientas digitales al servicio de tu tesis

Elaborado por: Silvia Gutiérrez De la Torre (Licencia: CC-BY-4.0)

Contacto: segutierrez@colmex.mx

Tipos de documentos primarios	Métodos digitales
a) Fuentes documentales (digitalizados)	a) Métodos para la búsqueda, extracción y guardado
b) Entrevistas	b) Herramientas de análisis cualitativo de datos
c) Noticias	c) Búsqueda avanzada de noticias
d) Twitter	d) Código para archivar tuits
e) Otros textos	e) Visualización de textos para el análisis exploratorio

Fuentes documentales - Búsqueda, extracción y guardado

Paso 1. Corroborar que no esté ya digitalizado y si hay versiones con OCR o en texto plano

Ejemplo de búsqueda:

Se buscan los discursos recogidos en:

Bailyn, Bernard. *Pamphlets of the American Revolution 1750-1776*. The John Harvard Library. Cambridge, Mass: Belknap Press of Harvard University Press, 1965.

El primero es:

"A Discourse Concerning Unlimited Submission and Non-Resistance to the Higher Powers" por Jonothan Mayhew.

Se observa que empieza así:

"let every soul be subject unto the higher powers" (una cita bíblica)

Y que en la parte media del texto contiene esta línea:

"it is evident that the affair of"

Se hace pues la búsqueda: [https://www.google.com.mx/search?](https://www.google.com.mx/search?ei=9kmpW6jPBJPV9AP03qr4CQ&q=%22let+every+soul+be+subject+unto+the+higher+powers%22+%22it+is+evident+that+the+affair+of%22&oq=%22let+every+soul+be+subject+unto+the+higher+powers%22+%22it+is+evident+that+the+affair+of%22)

[ei=9kmpW6jPBJPV9AP03qr4CQ&q=%22let+every+soul+be+subject+unto+the+higher+powers%22+%22it+is+evident+that+the+affair+of%22&oq=%22let+every+soul+be+subject+unto+the+higher+powers%22+%22it+is+evident+that+the+affair+of%22](https://www.google.com.mx/search?ei=9kmpW6jPBJPV9AP03qr4CQ&q=%22let+every+soul+be+subject+unto+the+higher+powers%22+%22it+is+evident+that+the+affair+of%22&oq=%22let+every+soul+be+subject+unto+the+higher+powers%22+%22it+is+evident+that+the+affair+of%22)

Paso 2. Si sí se encuentra; selecciona el recurso más adecuado.

Aspectos a considerar:

a) fuente confiable

b) elementos para citar la fuente

c) calidad de la digitalización

d.1) en caso de tener OCR (identificación de caracteres); corroborar calidad

d.2) en caso de estar en texto plano (es decir, que no sea una imagen sino texto "digital") escoger la versión con menos formato.

Ejemplo: <http://ota.ox.ac.uk/text/3173.html>

De PDF a TXT

<https://prezi.com/bqmnzyehjtul>

Extracción de texto: <https://prezi.com/bqmnzyehjtul/extraccion-de-pdf-a-txt/#>

1. Descargar: <https://www.xpdfreader.com/download.html>

2. Comprobar que se configuró el entorno
3. En tu terminal (cmd) abrir el directorio de archivos donde están tus pdf [cd + directorio]
4. Ingresar las palabras mágicas: FORFILES /M *.pdf /C "cmd /c pdftotext -enc UTF-8 @file" (para entender la sintaxis mirar: <https://ss64.com/nt/forfiles.html>)

De TXT a Corpus

Paso 1. Se copia el texto en un editor de texto plano

Copia el texto que te interesa desde la primera palabra dicha hasta la última y guardarla en un editor de texto sin formato. Por ejemplo:

- en Windows podría guardarse en [Bloc de Notas](#)
- en Mac, en [TextEdit](#);
- y en Linux, en [Gedit](#).

Paso 2. Guardar el archivo de acuerdo con ciertas convenciones

Pésima práctica:

"FINAL".doc



FINAL.doc!



FINAL_rev.2.doc



FINAL_rev.6.COMMENTS.doc



FINAL_rev.8.comments5.
CORRECTIONS.doc



FINAL_rev.18.comments7.
corrections9.MORE.30.doc



FINAL_rev.22.comments49.
corrections.10. #@\$%WHYDID
ICOMETOGRADSCHOOL?????.doc



JORGE CHAM © 2012

WWW.PHDCOMICS.COM

Tomado de www.phdcomics.com

Cuando guardes el texto debes considerar tres cosas esenciales:

La primera es que, si los textos de tu corpus están en español, deberás **guardarlos en UTF-8**, que es un formato de codificación de caracteres estándar para este idioma.

*¿Qué es utf-8? Si bien en nuestra pantalla vemos que al teclear una "É" aparece una "É"; para una computadora "É" es una serie de ceros y unos que son interpretados en imagen dependiendo del "traductor" o "codificador" que se esté usando. El codificador que contiene códigos binarios para todas los caracteres que se usan en el español es UTF-8. Siguiendo con el ejemplo "11000011", es una cadena de ocho bits --es decir, **ocho** espacios de información-- que en UTF-8 son interpretados como "É"*

La segunda es que **el nombre de tu archivo no debe contener acentos ni espacios**, esto asegurará que pueda ser abierto en otros sistemas operativos

¿Por qué evitar acentos y espacios en los nombres de archivo? Por razones similares a el inciso anterior, un archivo que se llame Ébano.txt no siempre será entendido de forma correcta por todos los sistemas operativos pues varios tienen otro codificador por defecto. Muchos usan ASCII, por ejemplo, que sólo tiene siete bits de manera que el último bit (1) de "11000011" es interpretado como el inicio del siguiente caracter y se descuadra la interpretación.

La tercera es que es una buena práctica integrar ciertos metadatos de contexto (v.g. fecha, género, autor, origen) en el nombre del archivo que te permitan partir tu corpus según diferentes criterios y también leer mejor los resultados. Para este tutorial hemos nombrado los archivos con el año del discurso presidencial, el código del país ([ISO 3166-1 alfa-2](#)) y el apellido de quien profirió el discurso.

[2007_mx_calderon.txt](#) tiene el año del discurso dividido con un guión bajo, el código de dos letras del país (México = mx) y el apellido del presidente que dictó el discurso, Calderón, (sin acentos ni eñes)

Entrevistas - Herramientas de QDA

Existen multiplicidad de herramientas para el Análisis Cualitativo de Datos (QDA por sus siglas en inglés).

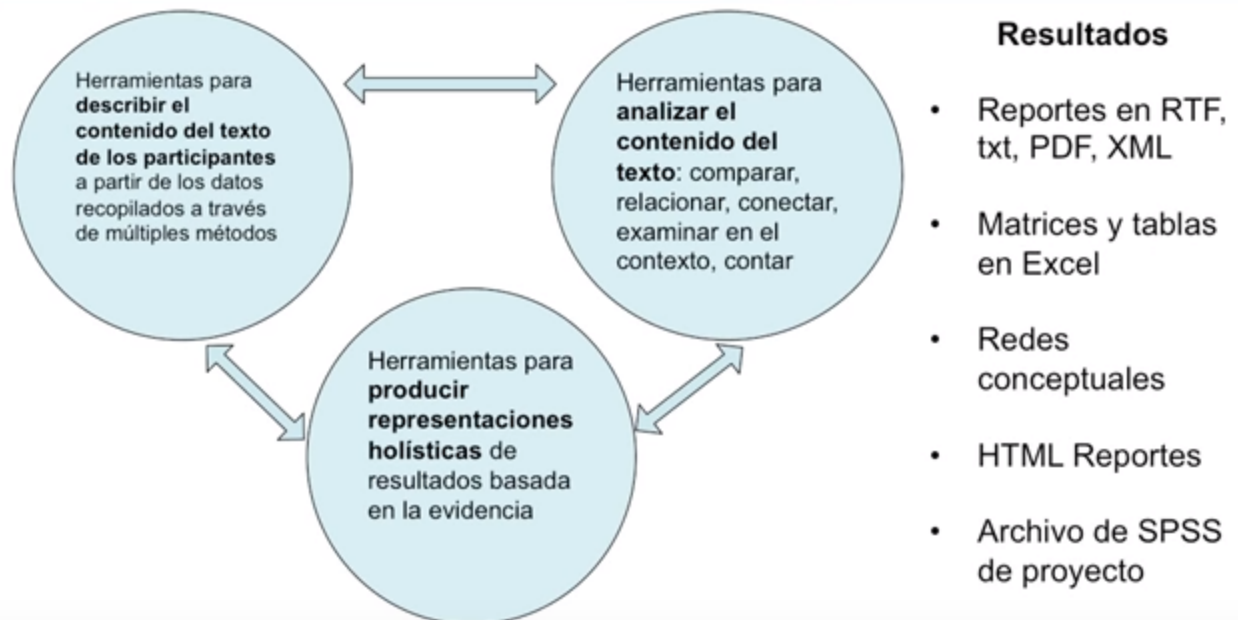
Licencias compradas por El Colegio de México: NVivo y ATLAS.ti (cf. <http://promepcm.colmex.mx/Formatos/TRANS2016/DrNieto3erTrim.pdf>)

Atlas.ti

Programa informático para el análisis cualitativo de datos no estructurados (textuales y multimedia) -- <https://atlasti.com/>

Análisis cualitativo cíclico

Ciclo en ATLAS.ti



Nota: Modelo inspirado en Harry Wolcott

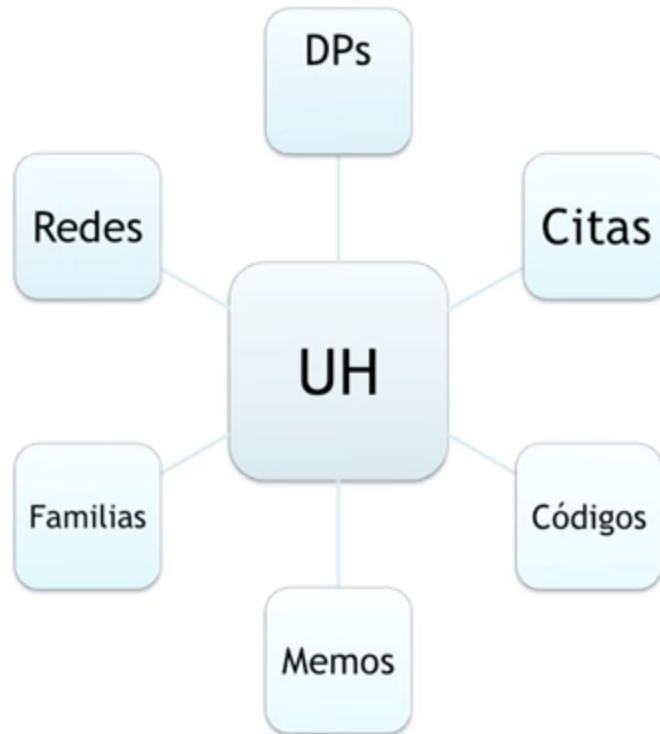
Software que facilita el análisis cualitativo de datos recogidos por medio de métodos no estructurados o semi-estructurados.

12:26 / 1:31:51

Tutorial completo en: <https://youtu.be/JprvFIBY5z4>

Unidad hermenéutica (UH) de un proyecto

Los objetos de la UH



Unidad Hermenéutica

DPs - documentos primarios (txt, videos, mapas)

Citas - Unidad básica de análisis (subrayados) -> pueden estar relacionados o no a **Códigos**

Memos - notas

Familias - agrupar los ejemplos (hombres y mujeres)

Redes - gráficos

Paso 1. Alimentar el software con el formato correcto de entrevistas

Introducción a ATLAS.ti 7 Windows-Taller Gratuito

Participante	Lugar entrevista	País: País en el que se realizó la entrevista	Genero: Genero del entrevistado	Población: Cómo definiría a la población cubierta por su servicio de salud?	Servicios: Qué servicios ofrece su centro de salud?
Participante 1	Centro de salud	Brasil	F	Servimos a individuos y familias que han migrado desde el noreste del país hacia la ciudad. Principalmente provienen de zonas rurales que han sido afectadas por sequía. La población infantil muestra signos de nutrición deficiente. Los hombres trabajan en la construcción y las mujeres en el servicio doméstico en hogares de clase media y clase alta.	Ofrecemos servicios de pediatría, incluyendo atención al niño sano y vacunaciones; planificación familiar, obstetricia y servicios ginecológicos; y servicios preventivos en general.
Participante 2	Centro de salud	Colombia	F	población originaria de zonas rural que se caracteriza por haber sido desplazada de zonas afectadas por el conflicto armado. Muchos de ellos son hablantes de idiomas de grupos originarios (pueblos indígenas). Éstos tiene concepciones culturales distintas a las de quienes trabajamos en el centro de salud. Por ejemplo, sus ideas de causas y tratamientos de enfermedades se basan en creencias muy relacionadas con espíritus de la naturaleza. Los	Enfocamos nuestros servicios en servicios pediátricos, planificación familiar, salud mental, el manejo de enfermedades crónicas. Hemos capacitado a un grupo de mujeres de la comunidad, todas pertenecientes al grupo indígena, quienes visitan los refugios donde están los desplazados para proveer información en salud y respecto a los servicios que ofrece el centro de salud.
Participante 3	Centro comunitario	Brasil	M	Nuestro centro de salud está localizado en una favela de la ciudad de Rio de Janeiro. Esta favela tiene aproximadamente 7.000 habitantes. Los hombres trabajan en la construcción, en servicios varios, y en el comercio informal. Las mujeres se desempeñan en el servicio doméstico, y también en servicios varios así como en el comercio informal. Servimos a una población urbana habitante de barrios periféricos. Somos el único centro de salud de	Ofrecemos servicios pediátricos, de planificación familiar, manejo de enfermedades crónicas, y atención de la mujer. Hemos establecido convenios con escuela para la realización de campañas educativas en prevención en salud, con niños pero también con padres.

Minuto 35: Cuestionarios

Paso 2. Crear códigos (categorías o unidades de sentido) que funcionen como observaciones controladas

Introducción a ATLAS.ti 7 Windows-Taller Gratuito

Población

Nuestro centro de salud está localizado en una favela de la ciudad de Rio de Janeiro. Esta favela tiene aproximadamente 7.000 habitantes. Los hombres trabajan en la construcción, en servicios varios, y en el comercio informal. Las mujeres se desempeñan en el servicio doméstico, y también en servicios varios así como en el comercio informal. Servimos a una población urbana habitante de barrios periféricos. Somos el único centro de salud de

Servicios

Ofrecemos servicios p...
enfermedades crónic...
convenios con escue...
prevención en salud,

Crear código(s) libre

Ingresar nombres de códigos. Presione fecha abajo para ingresar otro c...

Serv-sociales
Serv-salud
Serv-prevención
Serv-pediatría
Serv-
Agregar otro código...

Ingresar fecha

24/02/2014

Autor

Neringa

Creado

24/02/2014

Modificado

24/02/2014

OK Cancelar

¿Qué servicios ofrece su centro de salud?

2 Códigos 31 Servicios Todos Nombre Título

Minuto 44: Creación de códigos

Paso 3. Anotar la UH con basen en estas categorías

Población
Servimos a individuos y familias que han migrado desde el noreste del país hacia la ciudad. Principalmente provienen de zonas rurales que han sido afectadas por sequía. La población infantil muestra signos de nutrición deficiente. Los hombres trabajan en la construcción y las mujeres en el servicio doméstico en hogares de clase media y clase alta.

Servicios
Ofrecemos servicios de pediatría, incluyendo atención al niño sano y vacunaciones; planificación familiar, obstetricia y servicios ginecológicos; y servicios preventivos en general.

Familias	Nombre	Fundamenta...	Densidad	Autor	Creado	Modi
Unidad de Cód.	Población--	1	0	Neringa	24.02.2014 17:46	24.02.2014 17:46
Survey Bases (2)	Servicios--	1	0	Neringa	24.02.2014 17:46	24.02.2014 17:46
	Serv-pediatría	1	0	Neringa	24.02.2014 17:46	24.02.2014 17:46
	Serv-ginecología	0	0	Neringa	24.02.2014 17:46	24.02.2014 17:46
	Serv-salud	0	0	Neringa	24.02.2014 17:46	24.02.2014 17:46
	Serv-salud mujer	0	0	Neringa	24.02.2014 17:46	24.02.2014 17:46
	Serv-sociales	0	0	Neringa	24.02.2014 17:46	24.02.2014 17:46

Minuto 45: Asignación de códigos a secciones de la entrevista

Paso 4. Ordenar las categorías en relaciones semánticas

```
graph TD
    S72[Servicios 7-2] --> S21[Serv-sociales 2-1]
    S72 --> S24[Serv-salud 2-4]
    S72 --> S25[Serv-salud 2-5]
    S24 --> S21
    S24 --> S22[Serv-salud mujer 2-2]
    S24 --> S23[Serv-pediatría 2-3]
    S25 --> S21
```

Minuto 51: Ordenamiento semántico de los códigos

Paso 3. Visualizar los textos anotados para

NVivo

<https://youtu.be/SgY0MNdKVwM>

<https://youtu.be/SgY0MNdKVwM?t=647s>

RQDA

<https://youtu.be/TodbCCbV6Wc>

<https://youtu.be/TodbCCbV6Wc>

Marcado TEI

TEI (Text Encoding Initiative), estándar internacional para el marcado académico de textos.

Ejemplo de texto codificado: <https://hal.inria.fr/hal-01819505/document>

Ejemplo de aplicación: <http://www.bdpn.unam.mx/terms/name/planet>

Noticias - Búsqueda experta

Buscar por país o región

En la barra de búsqueda, introduce los términos seguidos de

`location:nombredepaís`. Por ejemplo: `televisión location:Chile`.

Buscar por periodo

Haz clic en Herramientas.

Junto a "Recientes", haz clic en la flecha hacia abajo

Selecciona un periodo o introduce uno (ejemplo: [enero 2010 - diciembre 2011](#)).

Buscar por palabra clave

- Para buscar palabras en el titular, utiliza `allintitle:palabraclave`. Por ejemplo: `allintitle:futbol`.
- Para buscar palabras en el texto del cuerpo, utiliza `allinbody:palabraclave`. Por ejemplo: `allinbody:televisión`.
- Para buscar palabras en la URL, utiliza `allinurl:palabraclave`. Por ejemplo: `allinurl:nacionalización`.

Buscar en publicación específica

- Para realizar una búsqueda en un sitio web de noticias específico, utiliza `site:nombredelsitioweb`. Por ejemplo, `site:reuters.com`.
- Para buscar una palabra clave específica, utiliza `site:nombredelsitioweb palabraclave`. Por ejemplo: `site:elmundo.es NBA`.
- Para buscar solo en fuentes específicas, utiliza `source:nombredepublicación`. Por ejemplo: `source:BBC`. Si la fuente tiene más de una palabra, utiliza comillas. Por ejemplo: `source:"The Guardian"`.

Twitter - Cosecha

DocNow responde al uso público de las redes sociales para el registro de eventos históricamente significativos, así como la demanda de académicos, estudiantes y

archivistas, entre otros, que buscan un medio fácil de usar para recopilar y preservar este tipo de contenido digital. DocNow tiene un fuerte compromiso con la priorización de prácticas éticas cuando se trabaja con contenido de redes sociales, especialmente en términos de recolección y preservación a largo plazo.

<https://www.docnow.io/>

Catálogos de Tweets:

1. <https://www.docnow.io/catalog/>
2. <https://zenodo.org/search?page=1&size=20&keywords=tweets&type=dataset>
3. <https://figshare.com/search?q=tweets&searchMode=1>

Rehidratación de Tweets: <https://github.com/DocNow/hydrator>

Cosecha individual:

<https://youtu.be/Cd8efC4fQZc>




<https://youtu.be/Cd8efC4fQZc>

Ejemplo: https://docs.google.com/spreadsheets/d/1M5hiVC-RKRyYj8GBNFv80YMR3O_PKPPAtc0MMsEn9Yo/edit?usp=sharing


Textos - Visualización para el análisis exploratorio



VOYANT

see through your text

Añadir textos   

Escriba una o más URL en líneas separadas o péguelas en un texto completo.



 Abrir  Cargar  Revelar

Cargar uno o mas documentos desde tu computadora

*Voyant Tools es un entorno de lectura y análisis basado en web para textos digitales .
Traducción al español por Manuel Vargas Alegría*

Tutorial: <https://programminghistorian.org/es/lecciones/analisis-voyant-tools>