

# Búsqueda Sistemática de Información

Fabio A. González Ph.D.

Depto. de Ing. de Sistemas e Industrial  
Universidad Nacional de Colombia, Bogotá

Seminario de Investigación I

# Contenido

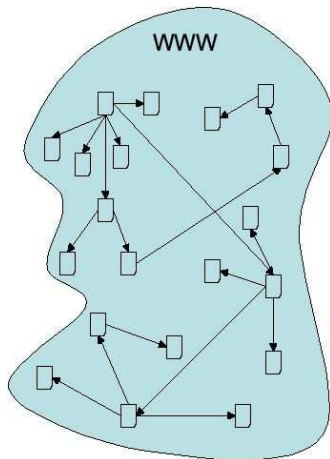
- 1 Introducción
- 2 Aproximación Sistemática
- 3 Bibliografía Anotada
- 4 Herramientas de Apoyo

# Preguntas Claves al Comenzar un Proyecto de Investigación

- Cómo encontrar un tema?
- **En dónde se encuentra información acerca de este tema?**
- Qué hacer con la información encontrada?

# Proceso típico de búsqueda de información en Internet

- 1 Buscar en Google usando términos generales que describan el tema
- 2 Visitar algunos de los links que parezcan relevantes (primeros 30-100 resultados).
  - 1 Leer contenido
  - 2 Si encuentra un link interesante, seguirlo (vaya a 2.a)
  - 3 Si encuentra nuevos términos relacionados, hacer una nueva búsqueda en Google (vaya a 1)
- 3 Piense en nuevos términos de búsqueda y vaya a 1

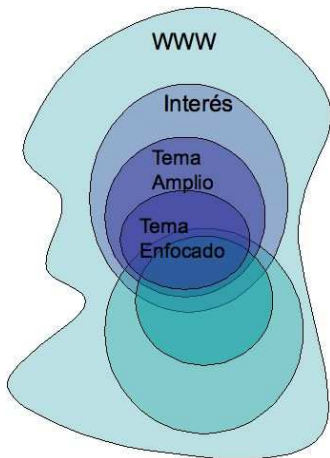


## Desventajas de este tipo de búsqueda

- Proceso tipo ensayo-error, muchos intentos antes de obtener resultados.
- El objetivo de la búsqueda no es claro.
- El proceso no produce resultados intermedios.
- El proceso no es fácilmente reproducible.
- No se puede percibir el avance.
- Es difícil retomar el trabajo, es como empezar de nuevo cada vez.

# Una aproximación sistemática a la búsqueda de información

- Recolección inicial de referencias (información relacionada → interés)
- Filtrado inicial de referencias (referencias más relevantes → tema amplio)
- Filtrado final de referencias (referencias claves → tema enfocado)
- Bibliografía anotada (lectura sistemática)
- Estado del arte



# Recolección inicial de referencias (1)

- Partiendo de un interés general se trata de recopilar un conjunto de referencias relacionadas (artículos, libros, páginas web, etc.)
- 200 a 300 referencias inicialmente.
- Se debe recopilar la siguiente información:
  - Título
  - Autor
  - Ubicación de la referencia (url, base de datos, texto fuente, etc.)
  - Deseable: referencia bibliográfica completa, resumen

## Recolección inicial de referencias (2)

- En donde empezar?
  - Libros de texto
  - Medios de difusión de la comunidad académica de la disciplina:
    - Identificar Journals más importantes
    - Identificar conferencias más importantes
    - Identificar grupos/instituciones de investigación más importantes
    - Recopilar tablas de contenidos de los últimos 2-5 años
  - Bases de datos bibliográficas
  - Motores de búsqueda especializados (citeseer, Google scholar, etc.)



## Filtrado Inicial de Referencias

- Partiendo de un tema(s) (amplio inicialmente, después enfocado) determinar las referencias más relevantes.
- 80 a 120 referencias inicialmente.
- En el caso de un artículo se debe leer el mínimo el resumen. Idealmente se debe leer introducción y conclusiones.
- En el caso de libros o tesis se debería leer la tabla de contenidos. Si es posible, leer alguna reseña (Amazon, Computing Reviews, etc.)
- Se debe consignar toda la información bibliográfica, el resumen y el texto completo si es posible.

## Filtrado Final de Referencias

- El propósito es determinar cuales son las referencias claves necesarias para ubicarse en el contexto del tema enfocado.
- 30 a 50 referencias inicialmente.
- Los artículos se deben leer totalmente, aunque esto no implica entenderlos hasta el último detalle.
- Se deberían identificar y leer las partes más relevantes de un libro o tesis. Si es posible, leer alguna reseña (Amazon, Computing Reviews, etc.)
- Estas referencias son la base para la bibliografía anotada.

# Qué es?

## Bibliografía Anotada

Una bibliografía anotada es una lista organizada de fuentes (referencias) tales como: libros, revistas científicas, periódicos, páginas web, etc., cada una de las cuales está acompañada de una anotación o descripción.

## Contenido de las Anotaciones

- Las anotaciones pueden incluir los siguientes tópicos:
  - Describir el contenido de la referencia
  - Describir la utilidad de la referencia
  - Discutir alguna limitación que la referencia pueda tener
  - Describir a qué audiencia está dirigido
  - Evaluar los métodos usados
  - Evaluar la confiabilidad de la referencia
  - Discutir el background del autor
  - Discutir las conclusiones presentadas por el autor
  - Describir sus reacciones a la lectura de la referencia

# Formato

- Las referencias se deben organizar de manera alfabética (apellido del autor). Cada referencia va acompañada de las respectivas anotaciones.
- La estructura de cada referencia debe ser la siguiente:
  - 1 Referencia completa
  - 2 Resumen en sus propias palabras (no una traducción literal del *abstract*)
  - 3 Anotaciones

## Ejemplo

**Naftali Tishby and Noam Slonim. Data clustering by markovian relaxation and the information bottleneck method.** *In Proceedings of the Neural Information Processing Systems Conference (NIPS), pages 640-646, 2000.*

El artículo presenta una técnica de agrupamiento que utiliza la matriz de distancias entre los objetos de la muestra, para definir una Cadena de Markov donde cada estado corresponde a un objeto. La Cadena de Markov definida resulta ergódica lo cual permite explorar el comportamiento de la misma en el camino hacia la distribución estacionaria, lo cual se refiere como relajación del proceso. En tal camino, emergen estructuras estables, donde la estabilidad se mide mediante una distancia entre distribuciones de probabilidad llamada pérdida de información mutua. Una vez encontradas esas estructuras estables, utilizan el llamado Bottleneck Method para extraer los grupos formados en esas estructuras. Se muestran resultados experimentales sobre un conjunto de datos sintéticos y uno real (Iris), reportando una efectividad del 97% en la clasificación para este último.

Es interesante el hecho que durante el proceso de agrupamiento, no se utiliza explícitamente la geometría del espacio que se está explorando, ya que el criterio de agrupamiento es similitud de distribuciones de probabilidad. Resulta desafortunado que no se presente el marco experimental y no se discuta la necesidad de algún tipo de preprocesamiento de datos para validar la técnica.

# Herramientas de Software para Apoyar el Proceso

- Es indispensable usar un sistema automatizado de manejo de referencias.
- Para el seminario se requiere manejarlas en formato bibtex.
- Sugerencia:
  - JabRef: <http://jabref.sourceforge.net/>