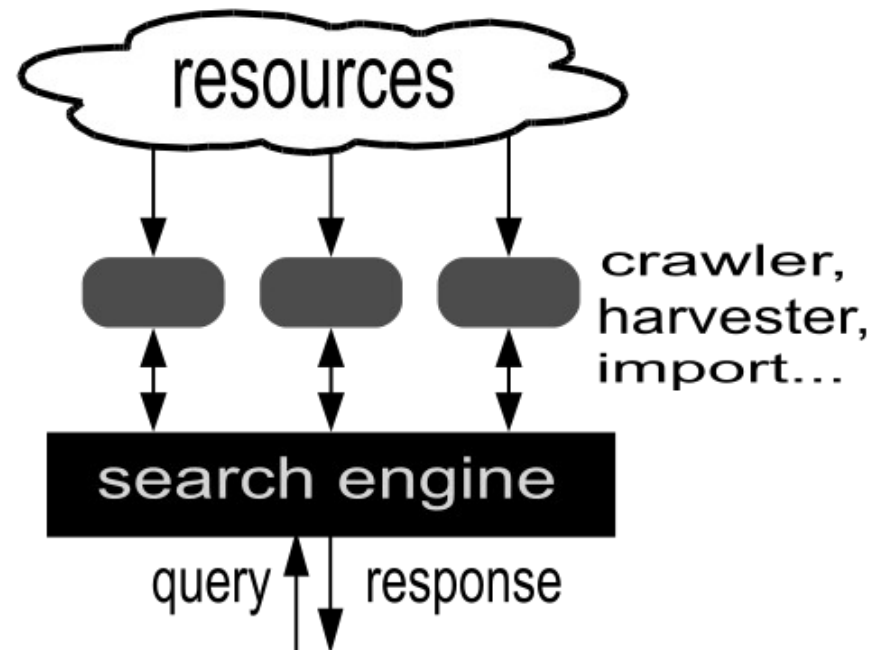


III Encuentro de Desarrolladores de las Redes BVS y Asociadas















Panel Engines de Busca

Conceptos Generales

- Operación básica de los engines de busca:
 - Web crawling (Bots or Spiders)
 - Indexing (read the text-based content)
 - Searching (index)



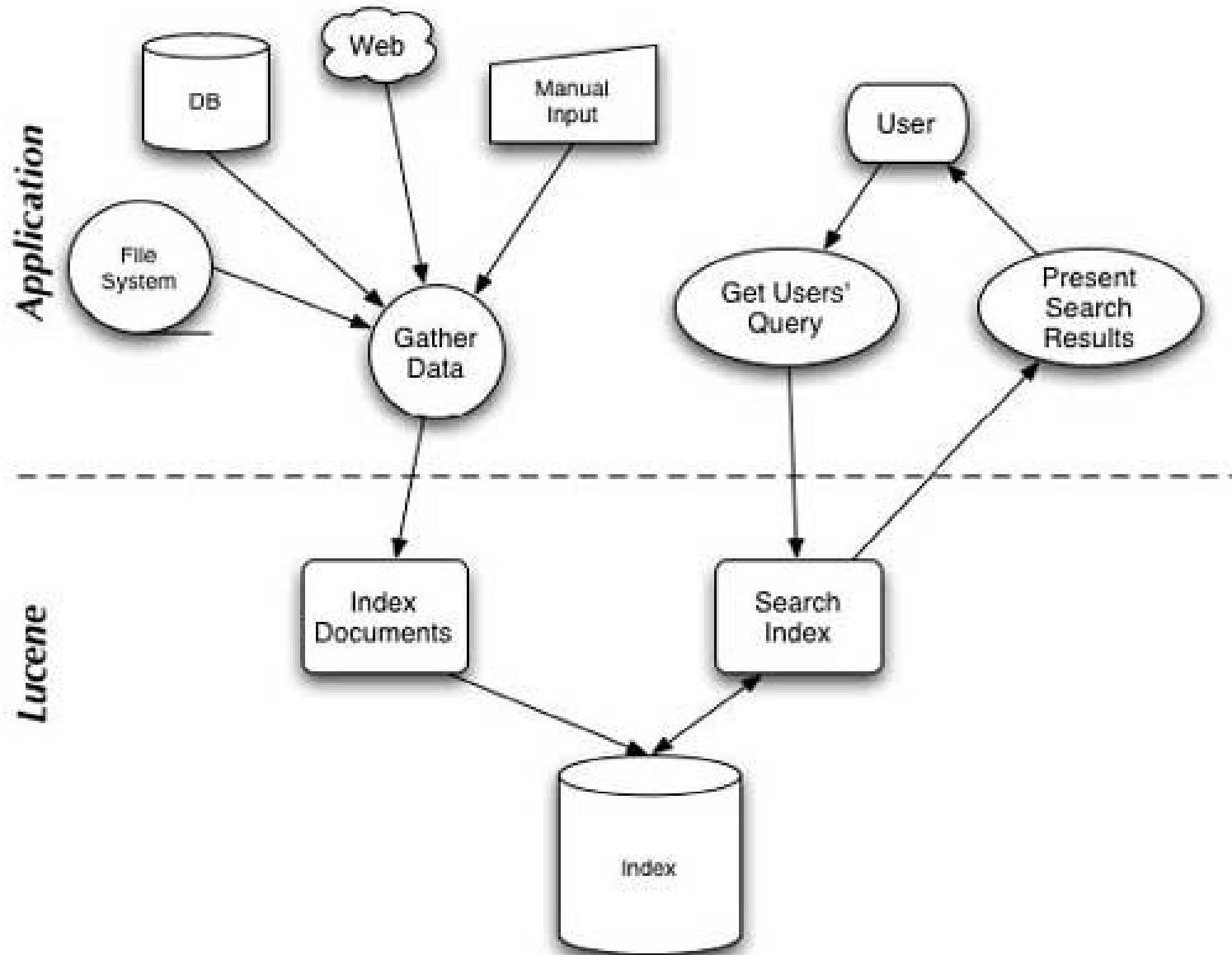
Participación de Mercado

Search engine ↕	Market share in May 2011 ↕		Market share in December 2010 ^[9] ↕	
Google	82.80%		84.65%	
Yahoo!	6.42%		6.69%	
Baidu	4.89%		3.39%	
Bing	3.91%		3.29%	
Yandex	1.7%		1.3%	
Ask	0.52%		0.56%	
AOL	0.3%		0.42%	

- Softwares para indización y recuperación de un dominio de documentos (intranet, biblioteca de la universidad, etc)
- Opciones libres
 - Lucene – Solr
 - Elastic Search
- Opciones propietarias
 - Fast Search (Microsoft)

- Solr/Lucene (Apache) es utilizado en la recuperación de la Biblioteca Virtual en Salud (BVS)
 - A full text search server based on Lucene
 - XML/HTTP, JSON Interfaces
 - Faceted Search (category counting)
 - Flexible data schema to define types and fields
 - Service oriented

Lucene Architecture





Google es una amenaza para el servicio de búsqueda de las bibliotecas virtuales?

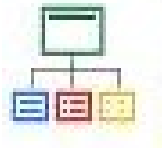
Visibilidad

- Diferentes dominios de fuentes información
 - Bibliotecas Virtuales = Bases de datos acerca de temas específicos.
 - Buscadores = Websites, Blogs, Listas de discusión, etc

Dificultad en los buscadores indizaren el contenido de las fuentes referenciales (paginas dinamicas) = **menor visibilidad de las bases de datos de las bibliotecas**

Promover la indización de los contenidos

- **Create a account in Google webmasters**
Crear una cuenta en Google webmasters
- **Create a sitemap**
Crear un mapa del sitio
- **Submit your content to Search Engines**
Enviar su contenido para los buscadores



- Dashboard
- Messages (2)
- Configuration
- Health
- Traffic
- ▾ Optimization
 - Sitemaps**
 - Remove URLs
 - HTML Improvements
 - Content Keywords
 - Structured Data
 - Other Resources
- Labs

Help with:
[Sitemaps for multiple sites](#)
[About Sitemaps](#)
[Additional support](#)
[Help center](#)

Sitemaps

[ADD/TEST SITEMAP](#)

By me (11) All (11)

Sitemaps content

All content types	Web pages
<ul style="list-style-type: none"> ■ Submitted ■ Indexed 	<p>313,431 Submitted</p> <p>187,702 Indexed</p>



Sitemaps (All content types)

Show 1-11 of 11

<input type="checkbox"/>	#	Sitemap ▲	Type	Processed	Issues	Items	Submitted	Indexed
<input type="checkbox"/>	1	/sitemaps/sitemap-artigos-1.xml	Sitemap	Jul 25, 2012	-	Web	30,000	14,403
<input type="checkbox"/>	2	/sitemaps/sitemap-artigos-2.xml	Sitemap	Jul 24, 2012	-	Web	30,000	12,613
<input type="checkbox"/>	3	/sitemaps/sitemap-artigos-3.xml	Sitemap	Jul 26, 2012	-	Web	30,000	14,957
<input type="checkbox"/>	4	/sitemaps/sitemap-artigos-4.xml	Sitemap	Jul 26, 2012	-	Web	30,000	11,753
<input type="checkbox"/>	5	/sitemaps/sitemap-artigos-5.xml	Sitemap	Jul 28, 2012	-	Web	30,000	12,169
<input type="checkbox"/>	6	/sitemaps/sitemap-artigos-6.xml	Sitemap	Jul 29, 2012	-	Web	13,976	6,108

Search Engine Optimization

- Mejoras en la estructura del sitio y de las paginas para aumentar la visibilidad y experiencia de uso del Portal.

- Elegir un título que comunica de manera eficaz el tema de los contenidos de la página
- Lo ideal es tener una etiqueta de título único, que ayuda a los buscadores a saber que la página es distinta de las demás de su sitio.

`<title>Consumo de alcohol entre estudiantes de medicina | LILACS </title>`

- La descripción es muchas veces presentada en los resultados de búsqueda de Google y otros buscadores y ayuda a los usuarios a elegir el documento.
- Tener una meta etiqueta de descripción diferente para cada página

```
<meta name="description" content="Este estudio tiene por objetivo identificar el patrón de consumo de alcohol entre los estudiantes de medicina. Se trata de un estudio transversal realizado en la Universidad Veracruzana, en México en 2006."/>
```

- Use palabras en las URL's
 - Evitar
 - Codigos y parametros (&id=XXX&...)
 - Nombres de paginas genericas “articulo1.html”
- Tener una estructura simple

<http://universidad/biblioteca/tesis/consumo-de-consumo-de-alcohol-entre-estudiantes-de-medicina>

- HTML Sitemap
 - Simple pagina html con los enlaces para todas las paginas del sitio
- XML Sitemap
 - Permite verificar si los buscadores están indizando el contenido del sitio (herramientas de webmaster)
 - Google, Yahoo, Ask y Microsoft Bing han acordado recientemente el uso del formato originalmente creado por Google.

Gracias!

**No se va. La Presentación de los
CASES ya va empezar**