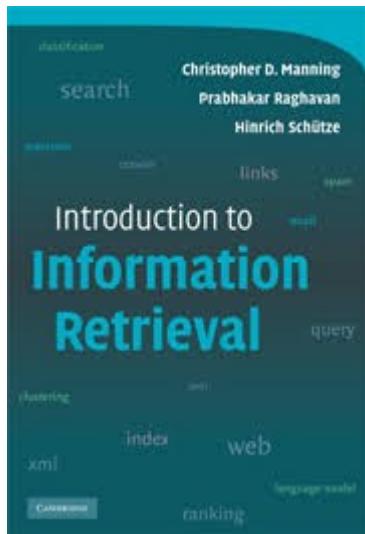


Sistemas de Recuperación de Información

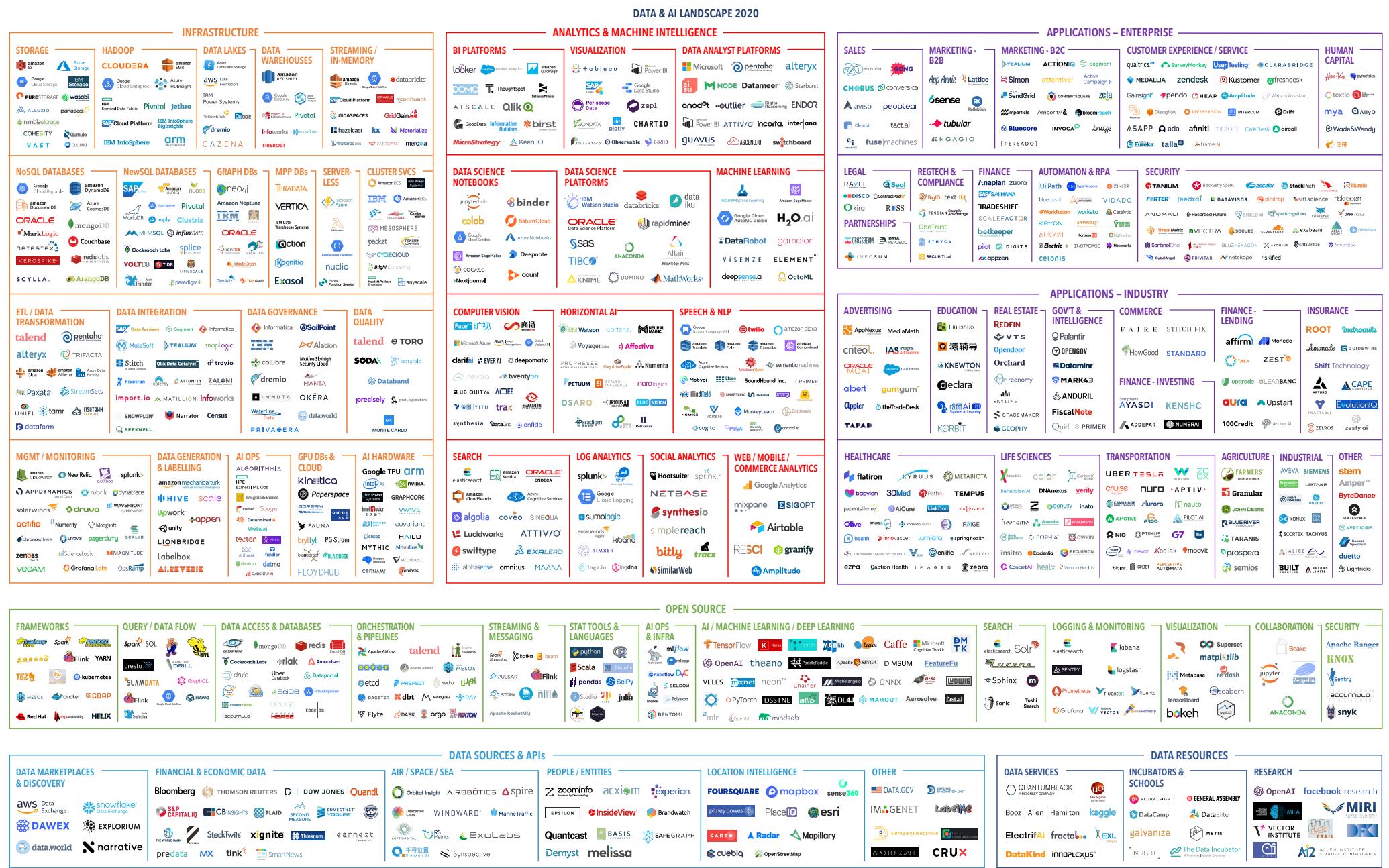


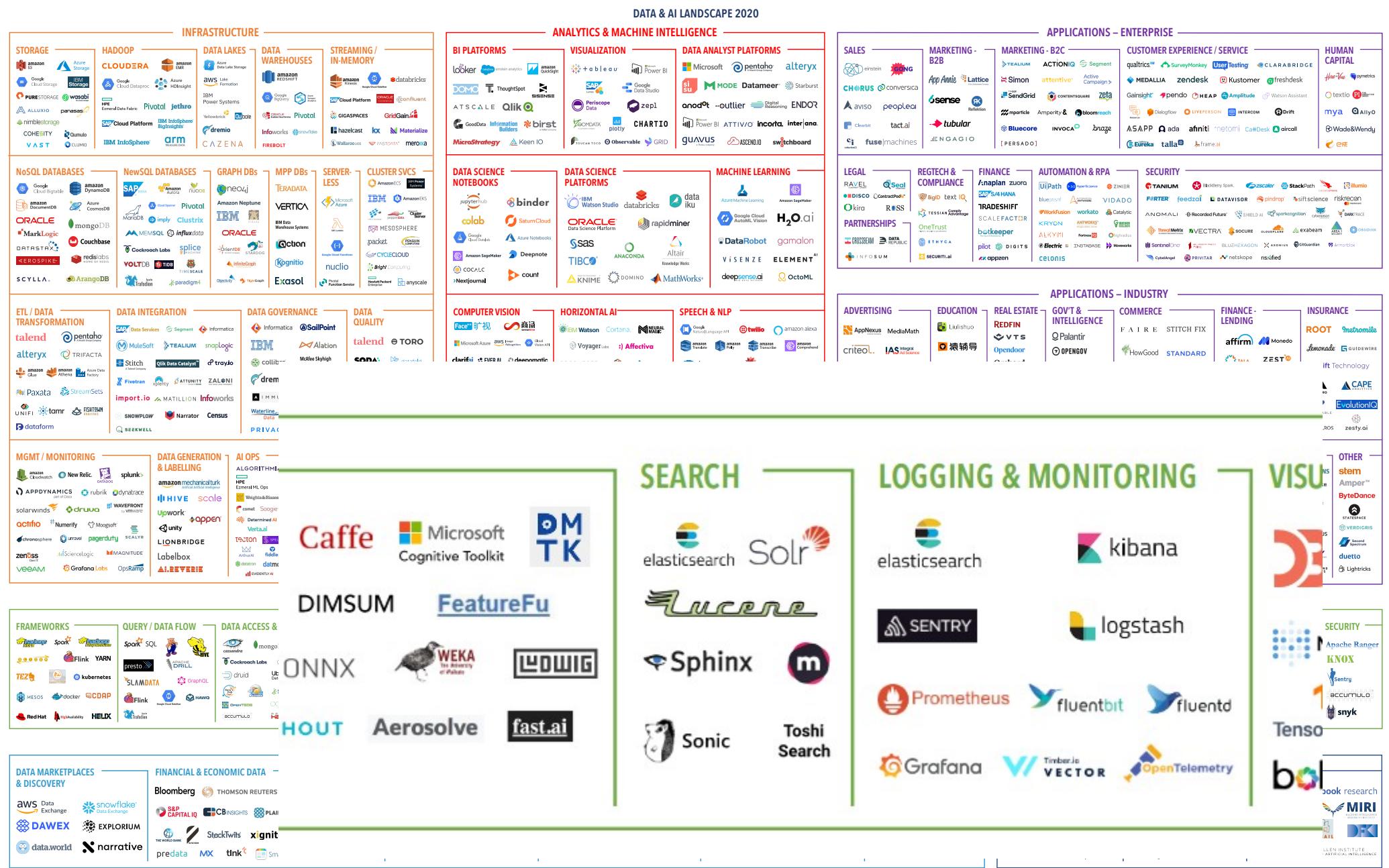
- “Big data is a term that describes large **volumes** of high **velocity**, complex and **variable** data that require advanced techniques and technologies to enable the capture, storage, distribution, management, and analysis of the information.” (TechAmerica Foundation’s Federal Big Data Commission, 2012)
- Variabilidad, datos heterogéneos.
 - Los datos estructurados (tablas o base de datos relacionales) representan el 5% de los datos existentes (Cukier, 2010)
 - Datos no estructurados: **texto**, imágenes, videos, etc. carecen **a priori** de la estructura que necesitan un computador para su análisis



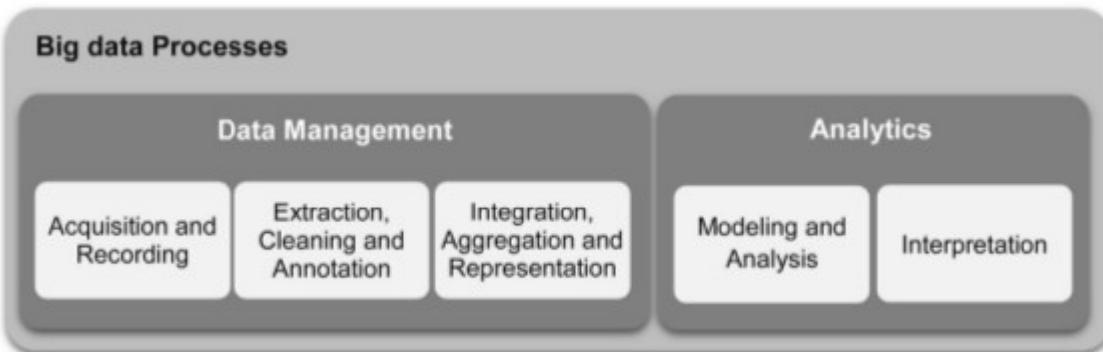
Recuperación de Información





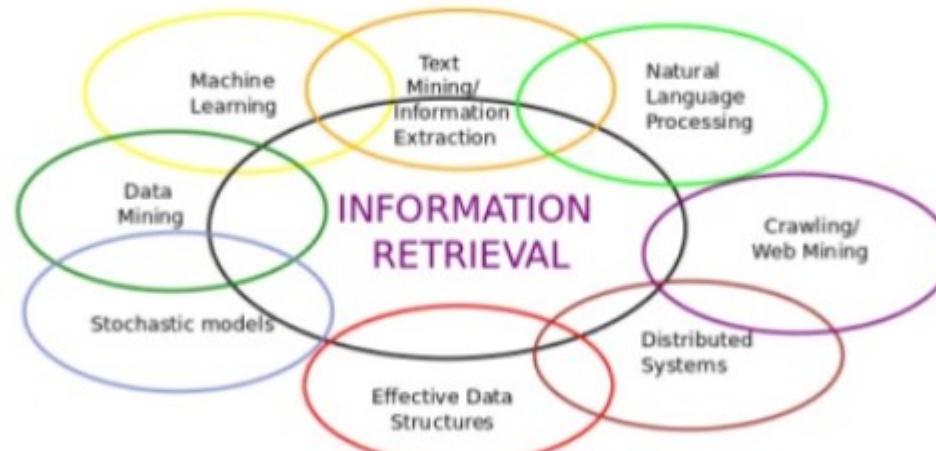


- El texto por si solo no posee información si no se procesa para permitir tomar decisiones.



Recuperación de Información

- Es la primera aplicación del Paradigma Big Data
- Como disciplina importa/exporta elementos con otras muchas



Objetivos

- Definir Recuperación de Información
 - Breve historia
 - Visión General



Recuperar Información => búsquedas

Web Search



UNIVERSIDAD DE GRANADA

Redes Sociales | Accesibilidad | ELVIRA | Asistente Virtual | CEIBIOTIC | Bóveda del crucero del Hospital Real

cei CÁMPUS DE EXCELENCIA INTERNACIONAL

Acceso ugr | Estudiantes | Estudiantes Internacionales | Profesorado

Presentación | Gobierno de la UGR | Facultades, Escuelas y Departamentos | Escuela Internacional de Posgrado | Institutos y centros de investigación | Otros centros | Gestión y servicios | Calidad | Estudios

Inicio > Buscador

Buscador

Búsqueda avanzada | Búsqueda con Google

Buscar: master

en: Universidad de Granada

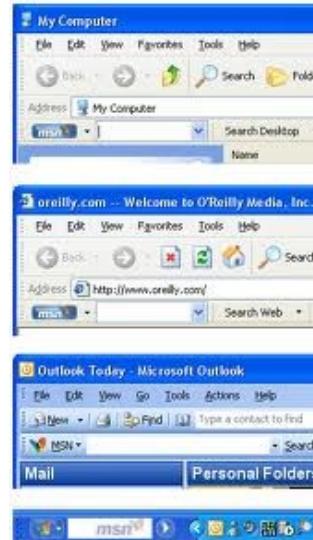
Toda la Universidad

Idioma: Español

Tipo: Todos

Buscar

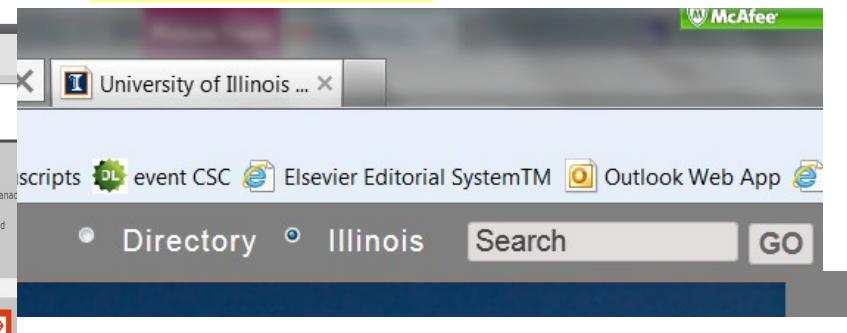
Desk Search



Enterprise Search



Site Search



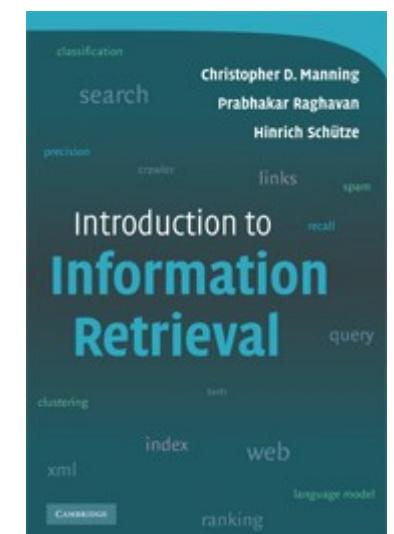
Social Media Search



Definición de RI



“RI es la disciplina que trata de encontrar material (típicamente documentos) de una naturaleza desestructurada (típicamente texto) que satisface una necesidad de información en una colección grande (típicamente almacenada en ordenadores)”
(Manning, Raghavan, 2008)



Detallemos un poco más ...

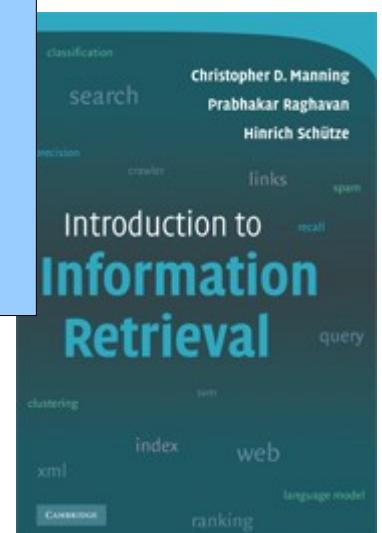


“RI es la disciplina que trata de encontrar material (típicamente **documentos**) de una naturaleza desestructurada (típicamente texto)



Información que se presenta al usuario:

- Texto: resúmenes, artículos, páginas web, capítulos de libro, correos electrónicos, frases, citas, ...
- y datos multimedia, como videos, música, o imágenes,





universidad de granada



SIGN IN

Web Images Maps Shopping More ▾ Search tools

About 3,810,000 results (0.27 seconds)

Cookies help us deliver our services. By using our services, you agree to our use of cookies.

OK

Learn more

[Universidad de Granada](#)

[www.ugr.es/en/](#) ▾ Translate this page

Apertura del curso académico 2013-2014 de la Universidad de Granada El ...

[Universidad de Granada](#)

Página oficial de la universidad.
Enlaces a estudios, centros ...

[Presentation](#)

Related links. UGR International The University of Granada ...

[International students](#)

Related links. UGR International The University of Granada ...

[More results from ugr.es »](#)

[University of Granada - Wikipedia, the free encyclopedia](#)

[en.wikipedia.org/wiki/University_of_Granada](#) ▾

The University of Granada (**UGR**) is a public university located in Granada, Spain that enrolls approximately 80,000 students. The university also has campuses ...

[Universidad de Granada - Wikipedia, la enciclopedia libre](#)

[es.wikipedia.org/wiki/Universidad_de_Granada](#) ▾ Translate this page

La Universidad de Granada (**UGR**) es una universidad pública española con sede en Granada y con campus en las ciudades de Granada, Ceuta y Melilla, ...

[Universidad de Granada. Centro de Enseñanzas Virtuales](#)

[www.ugr.es/](#) - Translate this page

3.7 ★★★★☆ 18 Google reviews · Write a review

Calle Real de Cartuja, 36 18012 Granada



Map data ©2013 Google, basado en BCN IGN España

University of Granada

[Directions](#)

The University of Granada is a public university located in Granada, Spain that enrolls approximately 80,000 students. The university also has campuses in Ceuta and Melilla. [Wikipedia](#)

Address: Calle Real de Cartuja, 36, 18012 Granada

Phone: 958 24 10 00

Enrollment: 78,656 (2010)

Founder: Charles V, Holy Roman Emperor

Founded: 1531

People also search for



University
of Seville
Seville



University
of Jaén
Jaén



Uned
Madrid



University
of Málaga
Málaga



Complut...
University
of Madrid
Madrid

Imagenes

Cookies help us deliver our services. By using our services, you agree to our use of cookies.

OK

Learn more



Logo



ugr

Universidad
de Granada



Universidad
de Granada



Universidad de Granada
Facultad de Medicina



Universidad de Granada



Detallemos un poco más ...



”“RI es la disciplina que trata de encontrar material (típicamente documentos) de una naturaleza **desestructurada** (típicamente texto) que satisface una necesidad de ...”



Datos estructurados:

Tienen campos a los que se asocia un valor
(ejemplo, hoja de cálculo, bases de datos)

Datos no estructurados:

Tienen estructura semántica arbitraria, como una
conversación humana, por ejemplo documentos empresa

Datos semi-estructurados

Mezcla de ambos, por ejemplo email

cei
CAMPUS DE EXCELENCIA INTERNACIONAL

Bóveda del crucero del Hospital Real

- | [Admission ugr](#) | [Students](#) | [International Students](#) | [Teaching Staff](#) | [Admin](#)

- | [Presentation](#)
- | [Governance and Administration](#)
- | [Faculties & Schools](#)
- | [Departments](#)
- | [Research Centres](#)
- | [Other Ce](#)
- | [Manag](#)
- | [Studies](#)
- | [Research](#)

Noticias ugr CANALUGR CcCu Agenda Cultural

 Apertura del curso académico 2013-2014 de la Universidad de Granada
El próximo viernes, 27 de septiembre, a las 11 h., en el Crucero del...

26/09/2013 | See more

Article Talk

Read Edit source View history

 Wiki Loves Monuments: Photograph a monument, help Wikipedia and win!

University of Granada

From Wikipedia, the free encyclopedia

The **University of Granada (UGR)** is a public university located in **Granada, Spain** that enrolls approximately 80,000 students. The university also has campuses in **Ceuta** and **Melilla**. Founded in 1531 by **Emperor Charles V**, has over 480 years of history and it is one of the oldest and most traditional universities in Spain. The University of Granada is the fourth largest Spanish university.

According to several rankings,^[1] the University of Granada is ranked among the top ten best Spanish universities and holds first place in **Translation and Interpreting Studies**.

The university has an important heritage, through its policy of using buildings of historical and cultural value. The **Madrasah of Granada** represents one such example. Similarly, the university has major new facilities committed to innovation, such as the **Parque Tecnológico de Ciencias de la Salud**.

Every year, over 2,000 European students enroll in the UGR through the **Erasmus Programme**, making it the most popular destination. The university's Center for Modern Languages (CLM) receives over 10,000 international students each year.^[2]

Contents [hide]

- 1 History
- 2 Campus
- 3 Academics
- 4 Center of Modern Languages
- 5 Famous alumni
- 6 Notes and references
- 7 See also
- 8 External links

History [edit source]

7 December 1526, a college was founded in Granada by **Holy Roman Emperor Charles V** for the teaching of **logic, philosophy, theology and canon law**.^[3] 14 July 1531, the establishment of a **studium generale** with the **faculties of theology, arts and canon law** was granted by a **papal bull** by **Clement VII**, marking the birth hour of the university.^{[3][4]}

Twitter

Buscar

Tweets

- | Siguendo
- | Seguidores
- | Favoritos
- | Listas

Sigue a Universidad Granada

Nombre completo

Correo electrónico

Contraseña

Regístrate

12.285 TWEETS 54.276 SIGUIENDO 55.016 SEGUIDORES

Universidad Granada @CanalUGR Twitter oficial de la Universidad de Granada. Una de las universidades de mayor prestigio europeo, fundada en 1531 y con 65.000 estudiantes. Granada, España · ugr.es

Universidad Granada @CanalUGR 12h Un saludo. Ver conversación

Universidad Granada @CanalUGR 12h @RMRubert Hola Ricardo, las asignaturas dentro del acuerdo de estudios se matrículan ahora (en el período de automatrícula online) Ver conversación

Universidad Granada @CanalUGR 12h @erreguez Es un estudio que realiza la UGR todos los años sobre los egresados, sus estudios, sus valoraciones, etc Un saludo. Ver conversación

Universidad Granada @CanalUGR 14h Si finalizaste tus estudios en la UGR en el año 2010, ayúdanos y participa en el Estudio de Egresados.

Fotos y vídeos

Tendencias Globales - Cambiar

#apple #London #fact

Detalles



“RI es la
encontrar
docume
desestruc

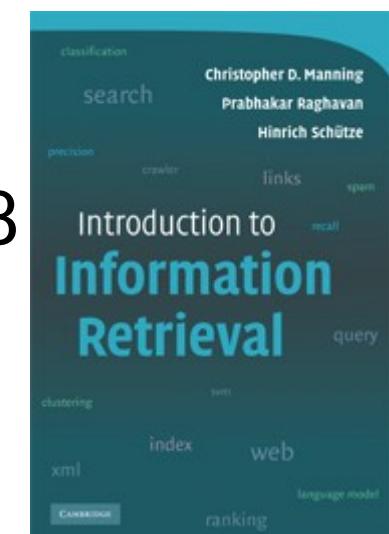


que satisface una necesidad de
información en una **colección grande**
(típicamente almacenada en
ordenadores)”

repositorio de documentos del cual se va
recuperar la información

- búsqueda en Internet, esta colección es la web
- Intranets, etc

Manning, Raghavan, Schütze, 2008





Sistema Inteligente para el acceso a las Colecciones Documentales

Buscar Diario de Sesiones o Boletines Oficiales

Legislatura: VIII

Tipo de D
3 de 4 se

Rango nº Documento: -

Fecha publicación: 2013 - 2013

Ninguno de los dos anteriores

Consulta:

crisis economica salud

[Sugerir Materias](#)

Materias

Seleccionar...

Buscar: ✓ Todas ✕ Ninguna

- politica sanitaria
- recesion economica
- regimen economico
- economista
- politica economica
- analisis economico
- ciencia economica
- region economica
- prevision economica



[La Biblioteca de la UGR](#) [Catálogos](#) [Biblioteca electrónica](#) [Servicios](#) [La bibli](#)

Catálogo

búsqueda rápida: 2.0

Use palabras completas o con * a la derecha

búsqueda avanzada: cualquier palabra

Busca las PALABRAS introducidas en cualquier orden en el campo seleccionado. Si desea que aparezcan en un orden determinado, introduzcas entre comillas. Téclee * (asterisco) para truncar a la derecha. Ejs.: Univers* Pulse Retorno o Enviar

Cualquier campo
Cualquier campo

● Limitar por

búsqueda experta: índices y subcatálogos

Seleccione índice

Donde empieza todo ...

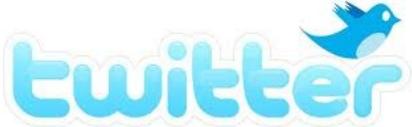
- Necesidad de Información del usuario:
 - información de la que el usuario precisa para alcanzar un objetivo
 - Puede ser muy precisa y clara o muy vaga, incluso (parcialmente) inconsciente
 - Y es dinámica: puede variar considerablemente a lo largo de una sesión de búsqueda
 - El usuario formula su necesidad en forma de consulta al sistema de IR
 - La formulación típica es una lista de palabras, pero hay otras...
 - ...formularios de búsqueda por campos
 - ...preguntas en lenguaje natural (question answering)
 - ...un documento



Necesidad de información

- Ejemplos de necesidad de información
 - Encontrar todos los documentos oficiales que abordan el papel de la Junta de Andalucía en la gestión de los expedientes de regulación de empleo
 - En qué año se estreno la primera película de Woody Allen
 - Piso de alquiler en Granada, con 3 dormitorios y dos cuartos de baño,
 - Talleres para llevar el coche a revisión
- Entre la necesidad de información y la consulta hay un “salto”: Considerar el texto, tal cual, como consulta no tiene porque ser una buena estrategia,
 - Capacidad y habilidad del usuario para traducir necesidades en consultas
 - Expresividad del lenguaje de consulta del sistema
(qué consulta usaríais)

Importante la eficiencia !

	# Queries /Day	X 1 sec	X 10 sec
	7,766,000,000	~2,150,000 hrs	~21,500,000 hrs
	2,100,000,000	~580,000 hrs	~5,800,000 hrs
	2,000,000	~550 hrs	~5,500 hrs
	• • • • • •		

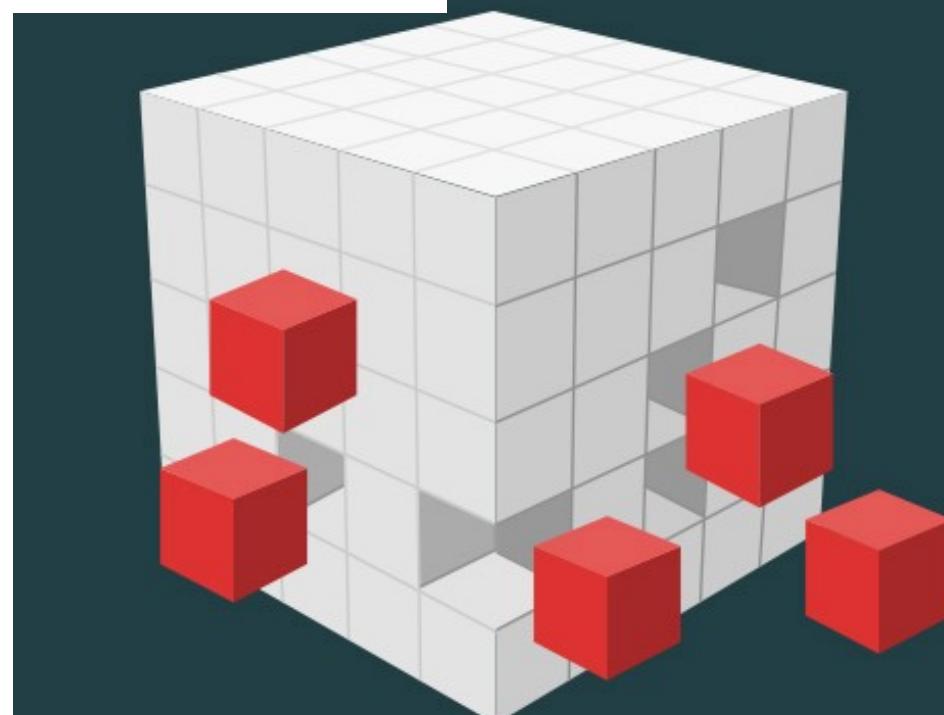
Fuentes:

Google, Twitter (actualizadas a 2015): <http://www.statisticbrain.com/>

Datos actualizados ... **Internet Live Stats**

Relevancia

- Ante la consulta de un usuario, el objetivo fundamental del SRI es encontrar la información que es útil o relevante para el usuario, pero recuperando el menor número posible de elementos no relevantes



El concepto de **relevancia** es clave en RI

¿ cómo se mide la relevancia de un documento a una consulta ?

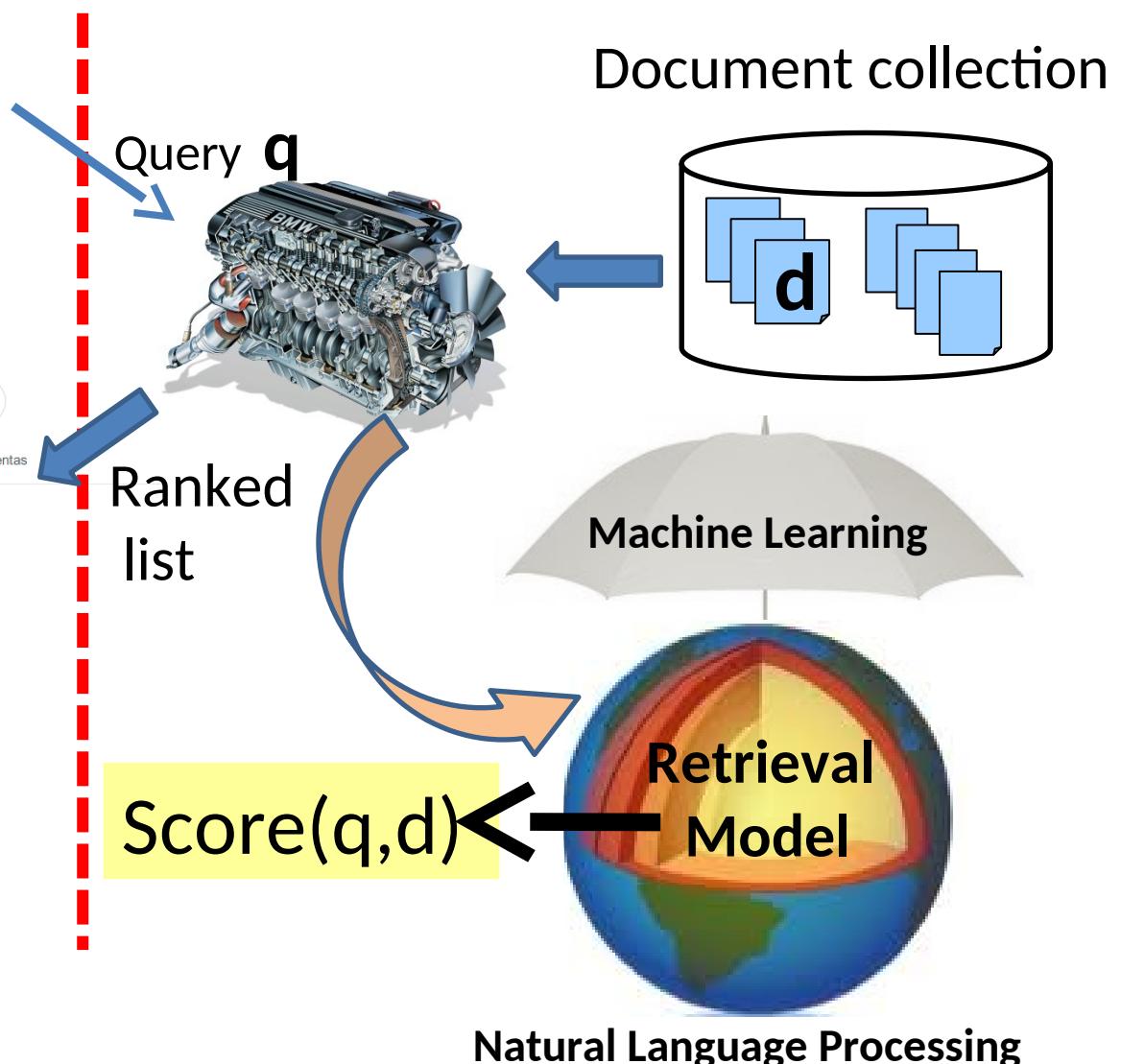
Recuperación de Información

Detrás de las cajas de búsqueda...

Google bing™

number of queries search engines

A screenshot of a search engine interface. At the top is a search bar with the placeholder "number of queries in search engines". Below it is a search results page with a header showing "Aproximadamente 16.500.000 resultados (0,82 segundos)". The main content area displays several search results, each with a title, snippet, and link. On the left side, there is a sidebar titled "Otras preguntas de los usuarios" containing a list of user questions like "What are the 3 main types of search queries?", "How many queries does a search engine answer every second?", and "How many queries does Google get a day?". At the bottom, there is a footer with links to "www.statista.com" and "U.S. search engines: number of core searches 2020 | Statista".



RI vs Bases de Datos



RI vs Bases de Datos

• DATOS:

BD:

- Conjunto de datos bien **estructurados** de elementos relacionados.
- Los elementos son atómicos con una clara interpretación

RI

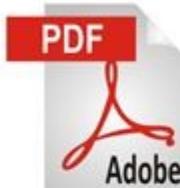
- Colecciones de Documentos (información)
- Información **sin estructura** o semiestructurado (título, fecha, asunto...)
- Puede tener muchas interpretaciones
- Puede ser muy imprecisa

Structured Data



0.103	0.176	0.387	0.300
0.333	0.384	0.564	0.587
0.421	0.309	0.654	0.729
0.266	0.750	1.056	0.936
0.225	0.326	0.643	0.337
0.187	0.586	0.529	0.340
0.153	0.485	0.560	0.428

Customer ID	Company	Contact	Contact Title	Address	City	Region	Postal Code
ALFKI	Alfreds Futterkiste	Maria Anders	Sales Repre.	Oberstraße 57	Berlin		12209
ANATR	Ana Trujillo	Ana Trujillo	Owner	Ave. de la Constitución 910	México D.F.		05021
ANTON	Antonio Moreno	Antonio Moreno	Owner	Moladeros 84	México D.F.		05023
AROUT	Around the Horn	Thomas Hendrickson	Sales Repre.	120 Hanover Street	London	WA1 1DP	
BERGS	Berglunds snabbköp	Christina Berglund	Order Admin.	Berguvsvägen 22	Luleå		S-958 22
BLAUS	Blauer See	Hanna Moos	Sales Repre.	Fasanenstr. 57	Mannheim		68306
BLONP	Blondel père et fils	Frédérique Cuvier	Marketing M.	24, place Kléber	Strasbourg		67000
BOLID	Bólido Comercio Exterior S.A.	Martín Sommer	Owner	C/ Araquil, 67	Madrid		28023
BONAP	Bon app'	Laurence Lebihan	Owner	12, rue des Rosiers	Marseille		13008
BOTTM	Bottom-Dollar Móveis e Decorações	Elizabeth Li	Accounting	23 Tswaane Street	Tsawassen	BC	T2F 8M4
BSBEV	B's Beverages	Victoria Ashworth	Sales Repre.	Fauntleroy	London		EC2 5NT
CACTU	Cactus Compañía de Desarrollo	Patricia Serrano	Sales Agent	Centro 333	Buenos Aires		1010
CENTC	Centro comercial Málaga	Francisco Cáceres	Marketing M.	Sierras de Málaga	Málaga		05022
CHOPS	Chop-suey	Yang Wang	Owner	Hauptstr. 29	Bern		3012
COMM1	Comércio Mineiro	Pedro Afonso	Sales Assoc.	Av. dos Lusíadas, 25	São Paulo	SP	05432-043
CONSH	Consolidated Holdings	Elizabeth Berger	Sales Repre.	Berkeley Glass	London		WX1 6LT
DRACD	Drachenblut	Sven Ottieb	Order Admin.	Weiseweg 21	Aachen		52066
DUMON	Du monde entier	Jeanne Labrune	Owner	67, rue des Carmes	Nantes		44000
EASTC	Eastern Connection	Ann Devon	Sales Agent	35 King George	London		WC3 6FW
ERNSH	Ernst Handel	Roland Mendel	Sales Manager	Kirchgasse 6	Graz		8010
FAMA	Famila Arquitectos	Aria Cruz	Marketing A.	Rua Orós, 92	São Paulo	SP	05442-030
FISSA	FISSA Fabricación Industrial S.A.	Diego Roel	Accounting	C/ Moralzarzal, 33	Madrid		28034



RI vs Bases de Datos

- **CONSULTAS:**

- En BD implica imponer unas restricciones que deben satisfacer los elementos de una tabla y por tanto la salida esta bien definida
- En RI la consulta esta formada por un conjunto de términos, que son una vaga especificación del contenido esperado de un doc.

Ejemplo

```
SELECT nombres, apellidos, edad  
FROM estudiantes  
WHERE edad > 21
```

	NOMBRES	APELLIDOS	EDAD
▶	1 Esteban	... Riquelme	22
	2 Ernesto	... Cardenas	25



RI vs Bases de Datos

- **SALIDA:**

- En BD podemos obtener como resultado una columna de la tabla
- En RI recuperamos un documento completo.
- En BD todos los elementos recuperados satisfacen las necesidades del usuario
- En RI es importante encontrar los mejores documentos que satisfacen dichas necesidades
 - Ranking de documentos

Ejemplo

```
SELECT nombres, apellidos, edad  
FROM estudiantes  
WHERE edad > 21
```

The diagram shows a SQL query example and its result in a table. The query is:

```
SELECT nombres, apellidos, edad  
FROM estudiantes  
WHERE edad > 21
```

The result is a table with three columns: NOMBRES, APELLIDOS, and EDAD. The data is:

	NOMBRES	APELLIDOS	EDAD
▶ 1	Esteban	Riquelme	22
2	Ernesto	Cardenas	25

Base de datos - Wikipedia, la encyclopédie libre
https://es.wikipedia.org/wiki/Base_de_datos ▾
Una base de datos es un conjunto de datos pertenecientes a un mismo contexto y debido al desarrollo tecnológico de campos como la Informática y la electrónica, las bases de datos están en formato digital, ...
Sistema de gestión de bases de - Base de datos relacional - Base de datos de red

¿Qué son las bases de datos? - Maestros del Web
www.maestrosdelweb.com/que-son-las-bases-de-datos/ ▾
26 oct. 2007 - Desde el punto de vista informático, la base de datos es un sistema: conjunto de datos almacenados en discos que permiten el ...

Otras preguntas de los usuarios

- ¿Qué es una base de datos en informática?
- ¿Qué es la BBDD?
- ¿Qué tipo de bases de datos existen?
- ¿Qué es una programación de base de datos?

¿Qué es Base de datos? - Definición en WhatIs.com - Search
<https://searchdatacenter.techtarget.com/es/definicion/Base-de-datos> ▾
En Informática, las bases de datos a veces se clasifican de acuerdo a su enfoque enfoque más frecuente es la base de datos relacional, una ...

¿Qué es una Base de datos? - Definición de Base de datos
www.masadelante.com/faqs ▾
Una base de datos es una colección de información organizada de forma que un ordenador pueda seleccionar rápidamente los fragmentos de ...

Aproximadamente 99.900.000 resultados (0,50 segundos)

Resultados de search engine optimization
Ver resultados de search engine optimization

Un recordatorio de privacidad de Google

RECORDARME MÁS TARDE LEER

search engine optimization

Definiciones en la Web

El posicionamiento en buscadores o Optimización de motores de búsqueda es el proceso de mejora de visibilidad de un sitio web en los resultados orgánicos de los diferentes buscadores. También es frecuente nombrarlo por su título inglés, SEO. ...
http://es.wikipedia.org/wiki/Search_engine_optimization

Enviar correo

Posicionamiento en buscadores - Wikipedia, la encyclopédie libre

https://es.wikipedia.org/wiki/Posicionamiento_en_buscadores ▾El posicionamiento en buscadores, optimización en motores de búsqueda o SEO (del inglés, Search Engine Optimization), es un conjunto de acciones ...
Historia · Técnicas para mejorar el ... · Link building · Posicionamiento y la ...

Search engine optimization - Wikipedia



search engine optimization (seo)

ina
a website or

Todos Imágenes Vídeos Mapas Noticias | Mis elementos guardados Land

3.510.000 Resultados Fecha Idioma Región

Engine Land.

Posicionamiento Seo Guru | posicionamientoseo.guru

<http://www.posicionamientoseo.guru> ▾Anuncio: Posicionamiento Seo Profesional Exito garantizado sin permanencia
Agencia Especialista en posicionamiento web seo sem, Adwords y ...

Busca en Ask.com - Toda la Información - ¿Buscas Información?

<es.ask.com/Resultados/Ahora> ▾Anuncio: ¿Buscas Información? Encuentra Resultados en Ask.com rápidamente
Más de 1 millón de usuarios visitaron ask.com el mes pasado

Aprende Online Sobre SEO - Aprende Nuevas Competencias

<www.nexuto.com/ClasesOnline/CursoSEO> ▾Anuncio: Optimiza El Proceso De Búsqueda y Posiciona Tu Página Web Con Palabras Claves.
Blog - Responsive Web Design - Marketing - Desarrollo Web

Seo search engine optimization - Encuentra en Smarter.com

<www.smarter.com/Seo search engine optimization/No busques más> ▾Anuncio: Busca Información para Seo search engine optimization aquí
Búsquedas populares · Últimas noticias · Noticias en red
Destinos: Madrid, Barcelona, Valencia, Sevilla

Posicionamiento en buscadores - Wikipedia, la encyclopédie ...

https://es.wikipedia.org/wiki/Posicionamiento_en_buscadores ▾El posicionamiento en buscadores, optimización en motores de búsqueda o SEO (del inglés, Search Engine Optimization), es un conjunto de acciones orientadas a ...
Discusión · Cargar Una Página al Azar · Bing · Cita Requerida · Acerca de Wikipedia · Contacto

Search Engine Optimization (SEO) Starter Guide - ... Traducir esta página

<https://support.google.com/webmasters/answer/7451184> ▾

Who is this guide for? If you own, manage, monetize, or promote online content via Google Search, this guide is meant for you. You might be the owner of a growing and ...

RI vs Bases de Datos

- **USUARIO:**

- En BD el usuario no afecta a la consulta ni a la salida
- En RI el usuario forma parte del modelo de recuperación.
 - la relevancia varía con el usuario, es importante entender sus necesidades (adecuar el sistema).

search engine optimization (seo)

Todo Imágenes Noticias Vídeos Maps Más Configuración Herramientas

Aproximadamente 61.000.000 resultados (0,45 segundos)

¿Qué es el SEO y por qué lo necesito? - 40defiebre

<https://www.40defiebre.com> ▾ Guía SEO "Cómo dominar a Google en 10 pasos" ▾
¿De dónde surge el término SEO (Search Engine Optimization)? ... El SEO se centra en los resultados de búsqueda orgánicos, es decir, lo que no son pagados:.

Posicionamiento en buscadores - Wikipedia, la encyclopédie libre

https://es.wikipedia.org/wiki/Posicionamiento_en_buscadores ▾
El posicionamiento en buscadores, optimización en motores de búsqueda o SEO (del inglés, Search Engine Optimization), es un conjunto de acciones ...
Historia · Posicionamiento natural u ... · Técnicas para mejorar el ... · Link building

What Is SEO / Search Engine Optimization? - Search Engine Land

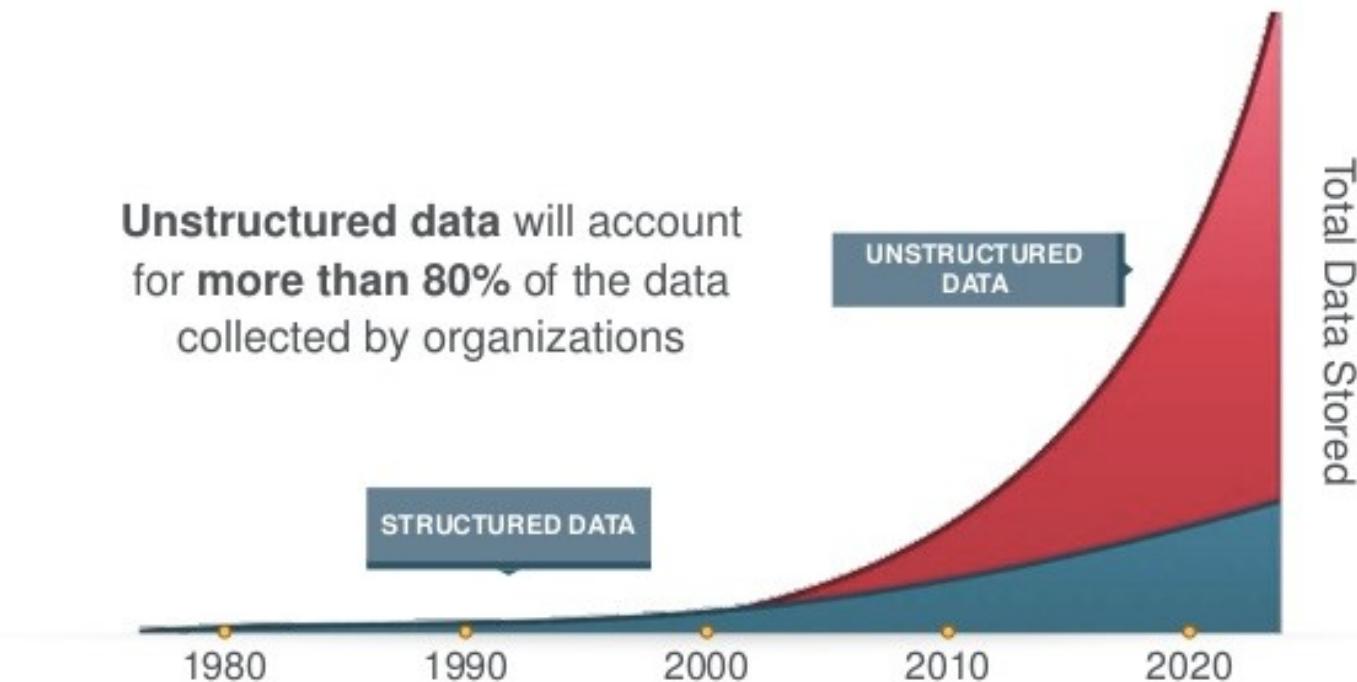
<https://searchengineland.com/guide/what-is-seo> ▾ Traducir esta página
Get started learning all about SEO from the industry's most trusted source, Search Engine Land.

Recuperación de Información vs Datos

De James Allan

	Bases De Datos	Recuperación de Información
Datos	Estructurados	No estructurados (semiestructurados)
Campos	Semántica clara (Teléfono, edad)	No hay campos (sólo texto)
Consultas	Definida (Álgebra relacional, SQL)	Texto libre (lenguaje natural, booleano)
Recuperación	Crítica (control de concurrencia, operaciones atómicas)	Minimizar
Comparación	Exacta (resultados son siempre “correctos”)	Imprecisa (se debe considerar la efectividad)

Un motivo para estudiar R.I.

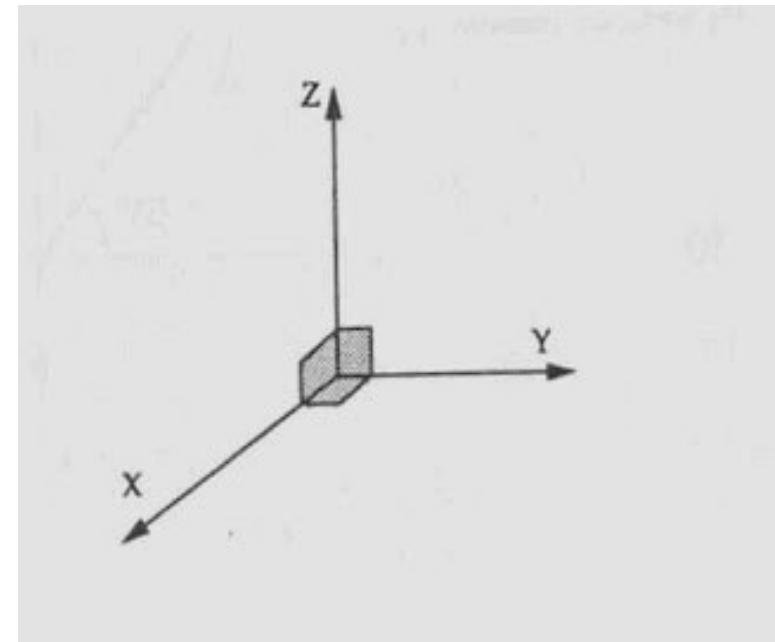


Source: Human-Computer Interaction & Knowledge Discovery in Complex Unstructured, Big Data

© 2014 MapR Technologies **MAPR** 4

Dimensiones de RI

1. Acceso a la Información
2. Tipos de Información
3. Colecciones



- Cuando consideramos valores específicos para cada una de ellas obtenemos distintos problemas de búsqueda
 - <http://www.thesearchenginelist.com/>

Tres dimensiones de RI

1. Acceso a la Información

- 1. Búsqueda Ejemplos ?
 - 2. Clasificación
 - 3. Agrupamiento
 - 4. Filtrado
 - 5. Recomendación
 - 6. Extracción de Información, resumen
 - 7. Búsqueda de expertos, búsqueda de respuestas
 - 8. Análisis de Sentimientos

Tres dimensiones de RI

2. Tipos de Información:

Implican estrategias de búsqueda diferentes

1. Documentos completos, pág. Webs, pdf, ...
2. Partes de un documento, XML
3. Anuncios, búsqueda patrocinada
4. Multimedia
5. Correos electrónicos
6. Geográfica
7. Noticias
8. Productos

Tres dimensiones de RI

3. Colecciones

1. Colecciones personales

disco duro

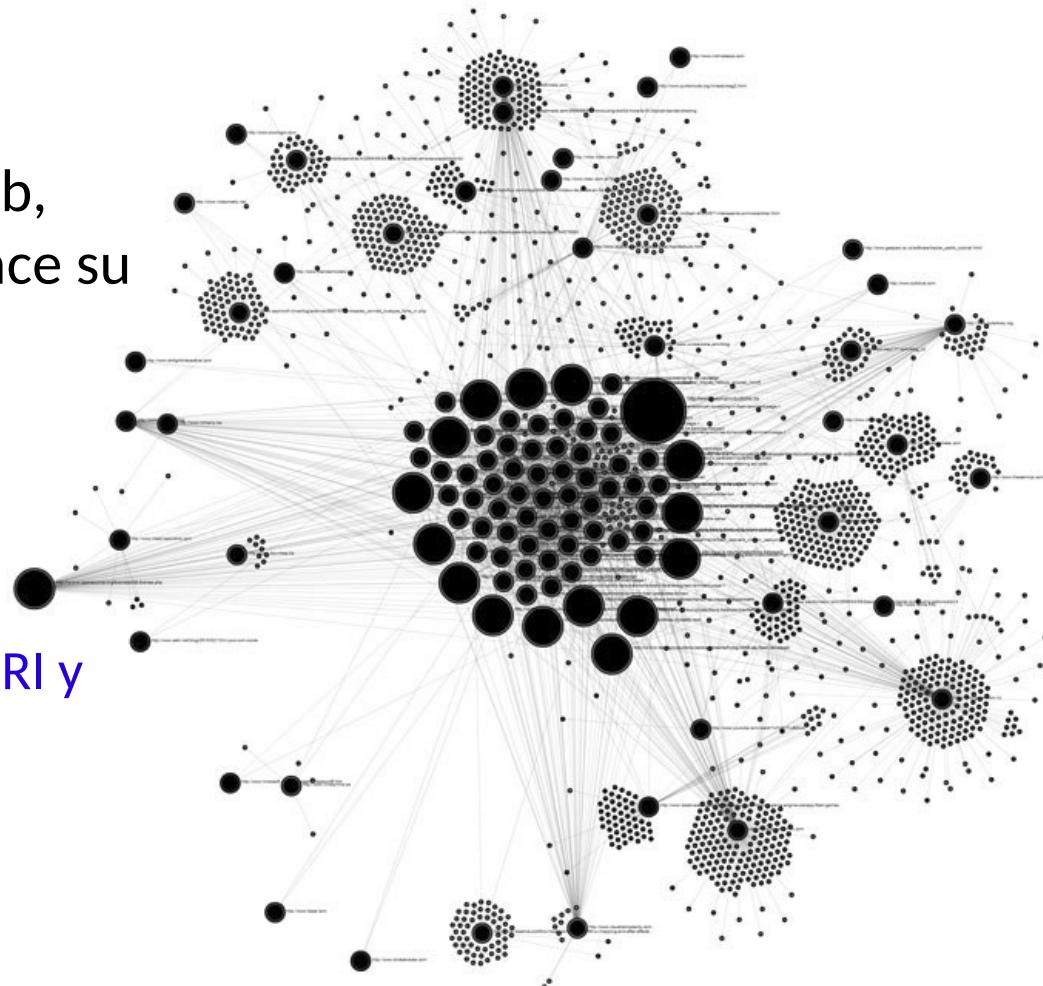
2. Colecciones corporativas,

búsqueda se adapta a los requerimientos de la empresa

3. Web

Actualmente

- Hasta hace poco, RI es un área focalizada en bibliotecónomos y expertos en información
- Sin embargo, la introducción de la Web, como repositorio de conocimiento, hace su uso extensivo.
 - Dado el tamaño de la web, encontrar información útil requiere realizar una búsqueda,
 - Buscar en la web está relacionado con RI y sus tecnologías.
- De la noche a la mañana, RI y sus tecnologías ha alcanzado un papel predominante en el mundo.



Historia de la RI

<http://www.searchenginehistory.com/>

Recuperación de Información

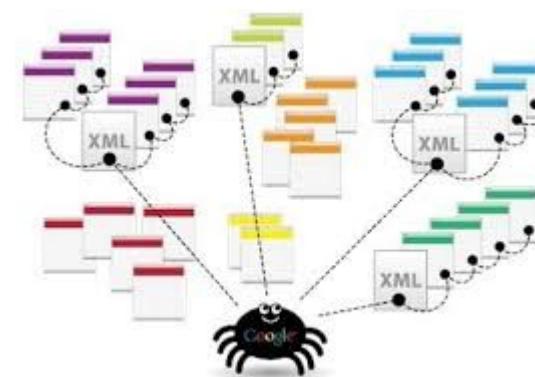
- ◆ Information Retrieval (IR), desde los 50
- ◆ La mayoría de las técnicas básicas se desarrollaron en los 60-70
 - Modelos, algoritmos, metodologías de evaluación
- ◆ Boom de la Web en los 90
 - Enorme impulso al campo de IR
 - Cuando pensamos en IR, pensamos en un buscador Web

Pero IR abarca más cosas: recomendación, QA, resúmenes automáticos, detección de topics, clasificación, búsqueda multimedia, vertical, geográfica...

- ◆ Muchos cambios desde mediados de los 90
 - PageRank, técnicas probabilísticas, machine learning, personalización, recomendación, diversidad, computación masiva, redes sociales...
- ◆ Sigue siendo un área muy dinámica
 - Investigación teórica, ciencia aplicada, ingeniería (academia + industria)

Cambios al considerar la Web

- Búsqueda Web es hoy la aplicación más destacada de RI
 - Elementos importantes son los algoritmos de ranking e indexación.
- En primer lugar, RI se ve afectada por las **características de la propia colección de documentos**
 - La web se compone de páginas distribuidas a través de millones de sitios, conectadas a través de hipervínculos
 - Esto requiere la recopilación de todos los documentos y el almacenamiento de copias de ellos en un repositorio central, antes de la indexación (crawling)



Cambios al considerar la Web

- En segundo lugar, clave la **eficiencia** dado el tamaño de la colección
 - Elevado volumen de las consultas que se realizan al día.
 - Como consecuencia, el rendimiento y la escalabilidad se han convertido en características críticas del sistema de RI
- En tercer lugar: obtener resultados de calidad (**eficacia**) se complica
 - la predicción de relevancia es mucho más difícil en colecciones muy grandes
 - Afortunadamente, la Web también incluye nuevas fuentes de evidencia
 - Ej: los hipervínculos y clicks de los usuarios

Cambios al considerar la Web

- En cuarto lugar, la Web también es un medio para hacer negocio
 - Búsqueda se ha extendido más allá de la búsqueda de texto sino la búsqueda de otro tipo de información
 - Ej: el precio de un libro, el número de teléfono de un hotel, el enlace para la descarga de un software
- Finalmente, hay que gestionar el spam en la Web
 - Web Spam: abuso de información comercial disfrazada en forma de contenido informativo

Sistema de Recuperación de Información (SRI)

- El software que implementa un modelo de Recuperación de información.
- Objetivo:
 - Predecir, con un conocimiento previo sobre el usuario y la propia fuente de recursos, qué objetos son los más adecuados para que el usuario pueda resolver su problema.

1

WEB

IMÁGENES

VÍDEOS

NOTICIAS

COMPRAS

RESPUESTAS

MÁS ▾

Buscar en toda la Web en español sólo en España

6

BÚSQUEDAS
RELACIONADAS

viajes iberia

vuelos iberia

iberia plus

iberia líneas aéreas

iberia express

FILTRAR POR HORA

Cualquier hora

Ayer

La semana pasada

El mes pasado

4

Anuncios relacionados con iberia

Vuelos Iberia | iberia.com Sitio oficial

Oferta de Vuelos en la Web Oficial ¡Reserve Ya su Vuelo en Iberia.com!

 www.iberia.com/Oferatas-Vuelos

5

Barcelona Nueva York desde 485€

Vuelos Barcelona Lima desde 691€

Vuelos Barcelona Miami desde 667€

Vuelos Madrid Nueva York desde 490€

Vuelos de Madrid a Lima desde 789€

Vuelos de Madrid a Miami desde 369€

Iberia: Vuelos desde 47€ | Iberia.edreams.es

Todas las Ofertas de Iberia: Reserva tu vuelo Iberia al Mejor Precio.

Iberia.edreams.es

Precio Mínimo Garantizado - BCN-Madrid en Oferta - Vuelos 2013: 75% DTO.

Vuelos desde 19€ | iberia.rumbo.es

Reserva Ahora al Mejor Precio ¡Encuentra los Vuelos Más Baratos!

iberia.rumbo.esVuelos Baratos - Descubre el nuevo Vuelosbaratos.

Rápido, sencillo e innovador.

www.vuelosbaratos.esMás resultados patrocinados iberiaIberia - Últimas noticias

Noticias

Fotos

Iberia presentará su nueva imagen corporativa en octubre

- abc.es - Hace 5 horas

El próximo día 15 de octubre Iberia anunciará la que es, probablemente, la medida más simbólica de su nueva estrategia para reforzar la imagen de la compañía: el... [Noticia completa](#)

2



3

Iberia elimina la corona de su imagen en plena crisis institucional de ...

- El Confidencial - 28 sep 06:09

Barajas busca rutas a Iberoamérica y Asia para frenar la caída de ...

- El País - Hace 8 horas

El AVE corta las alas al avión

- abc.es - Hace 5 horas

COMPONENTES

About 34,300,000 results (0.22 seconds)

Cookies help us deliver our services. By using our services, you agree to our use of cookies.

[OK](#) [Learn more](#)

Ads related to iberia [ⓘ](#)

[iberia.com - Iberia](#)

[www.iberia.com/Vuelos-Baratos](#) ▾

Miles de Vuelos en Oferta con Iberia ¡Reserva al Mejor Precio!

Iberia Líneas Aéreas has 1,157 followers on Google+

Madrid Nueva York 490€

Barcelona Miami 667€

Barcelona Lima 691€

Madrid Miami 626€

Madrid Lima 789€

[Vuelos Baratos - Precio Mínimo Garantizado - Iberia.edreams.es](#)

[iberia.edreams.es](#) ▾

Aprovechate, Reserva Tu Vuelo Ya!

eDreams España – Tus vuelos al mejor precio has 288,827 followers on Google+

Vuelos 2013: 75% DTO. - BCN-Madrid en Oferta - Nueva York desde 524€

[IBERIA.COM in Spain. The Best prices for Iberia flights](#)

[www.iberia.com/?language=en](#) ▾

IBERIA.COM in Spain. The Best prices for Iberia flights. Highlighted offers.

[Iberia](#)

Vuelos - Iberia - Gestión de reservas - Vuelos baratos - ...

[More results from iberia.com »](#)

[Check-in online](#)

Reserva tu asiento e imprime tu tarjeta de embarque, y podrás ir ...

[Iberia, online check-in - Print out your boarding card and save time ...](#)

[www.iberia.es/online-checkin](#) ▾

Book your seat and print out your boarding card and you can go directly to the gate for boarding.

[Iberian Peninsula - Wikipedia, the free encyclopedia](#)

[en.wikipedia.org/wiki/Iberian_Peninsula](#) ▾

The Iberian Peninsula (Asturian, Galician, Leonese, Mirandese, Portuguese, and Spanish: Península Ibérica, Catalan: Península Ibèrica, Aragonese and ...

Iberia



Airline company

Iberia, Líneas Aéreas de España, S.A. Operadora, Sociedad Unipersonal, trading as IBERIA, is the flag carrier and one of the largest airlines of Spain. [Wikipedia](#)

CEO: Luis Gallego

Founded: June 28, 1927

[Track flight status](#)

Enter flight number

[Track flight](#)

Recent posts



Celebrando el Día Mundial del Turismo con Fitur, ¡estrenamos concurso! Contesta correctamente a 4 preguntas sobre el planeta azul y podrás ganar 30.000 ... 27 Sep 2013

[Feedback/More info](#)

See results about



[Caucasian Iberia](#)

Kingdom

In Greco-Roman geography, Iberia was the name for a kingdom of the ...



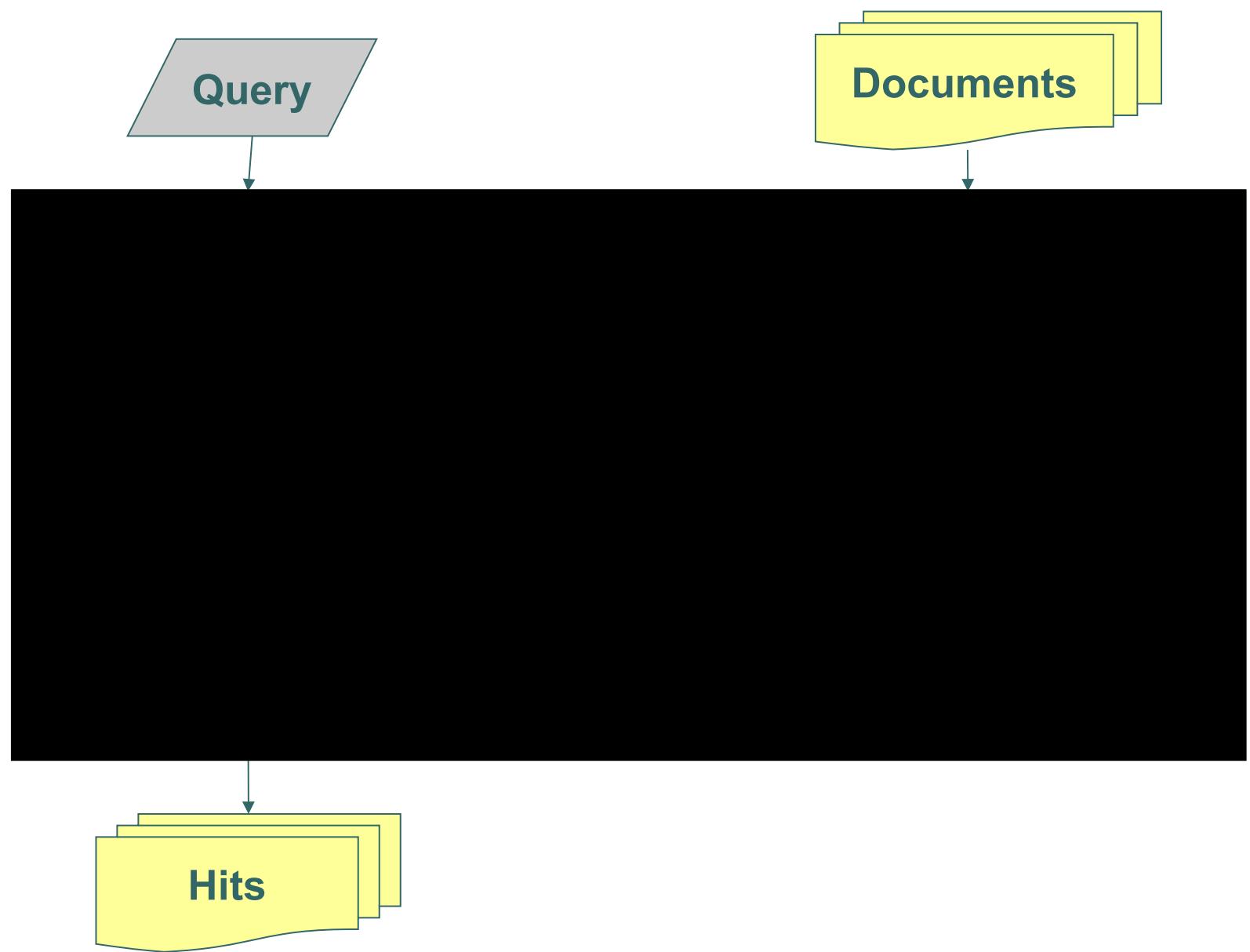
[Iberian Peninsula](#)

Peninsula in Europe

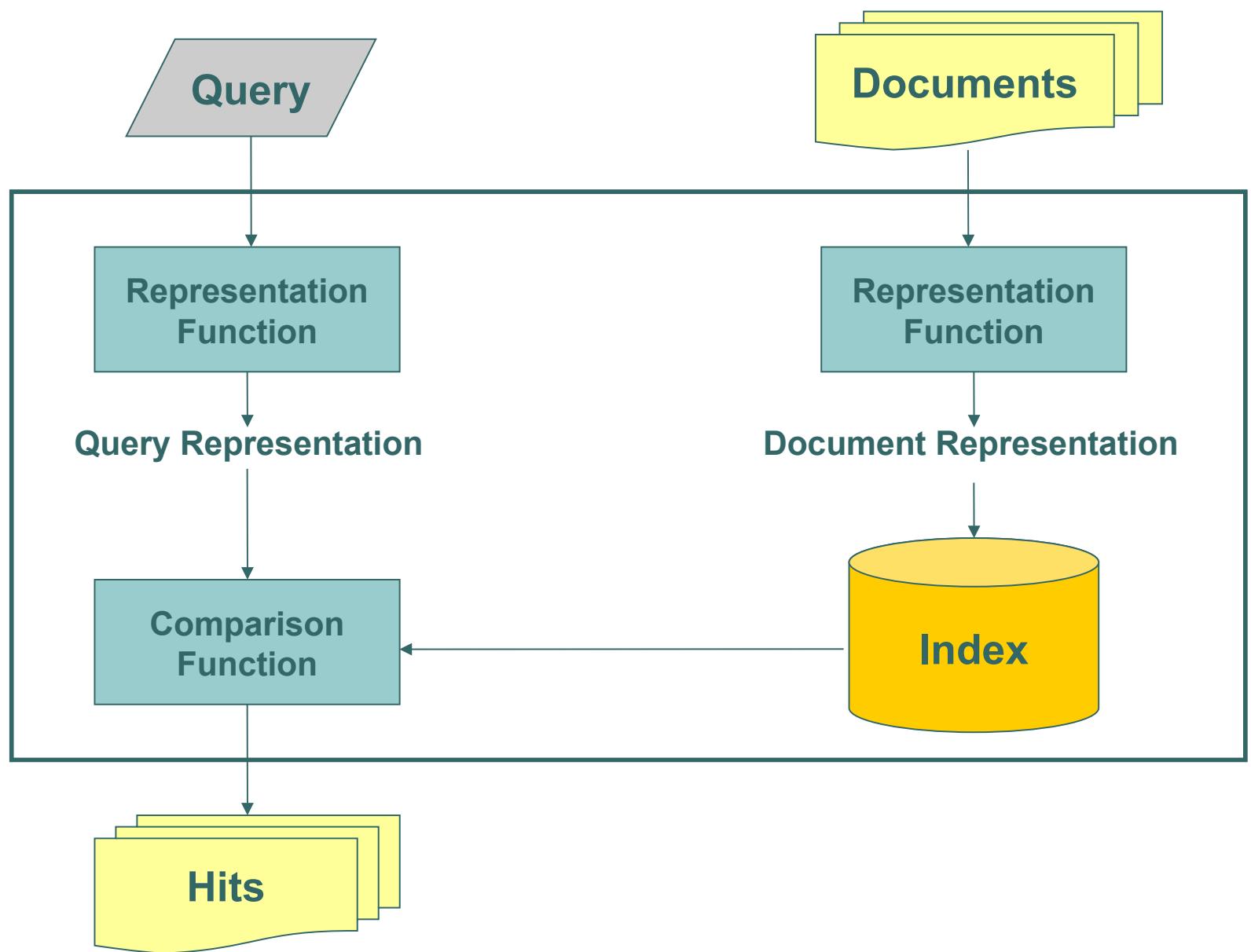
The Iberian Peninsula, commonly called Iberia, is a peninsula located in ...

Arquitectura de un Sistema de RI

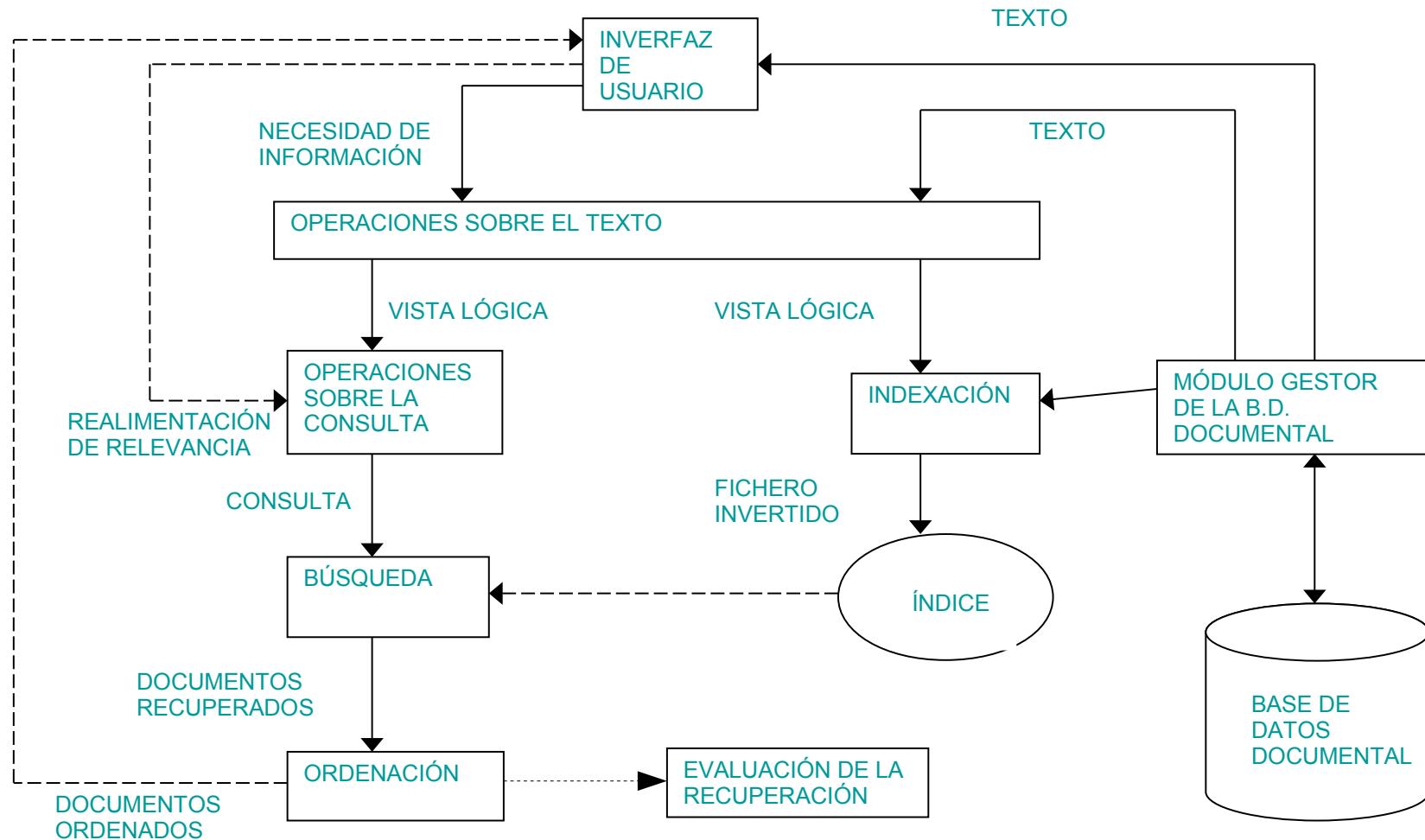
SRI como una caja negra



Dentro del SRI



Arquitectura de un SRI



Evaluación de los S.R.I.

- Determinar la calidad de un S.R.I.
- En función de:
 - Espacio.
 - Tiempo (eficiencia).
 - Facilidad de uso (usabilidad).
 - Calidad de la recuperación (eficacia).

Evaluación de los S.R.I.

- Métricas:
 - **Precisión**: proporción de documentos recuperados que son relevantes.
 - **Recall** (exahustividad): proporción de documentos relevantes recuperados.
 - **Average Precision**: proporción de documentos no relevantes recuperados.
 - **NDCG**: posiciones de los documentos relevantes en el ranking

Evaluación de los S.R.I.

	RELEVANT	NON-RELEVANT	
RETRIEVED	$A \cap B$	$\bar{A} \cap B$	B
NOT RETRIEVED	$A \cap \bar{B}$	$\bar{A} \cap \bar{B}$	\bar{B}
	A	\bar{A}	N

(N = number of documents in the system)

$$\text{PRECISION} = \frac{|A \cap B|}{|B|}$$

$$\text{RECALL} = \frac{|A \cap B|}{|A|}$$

$$\text{FALLOUT} = \frac{|\bar{A} \cap B|}{|\bar{A}|}$$

Recuperados

No recup.

Q1: R N R R N N N N N R **R N R R N N N N R R**

Q2: N N N R R R N N R R R N **N N N R N R N N**

Q3: R N R N R R N N R N N N R R R R N N R R

Q4: R R N R R R N N N N N N R R R R N N R R

Precisión Media

Es un indicador que usa un único número para resumir la eficacia de un sistema en relación a una consulta.

AP(q) precisión media (*average precision* en inglés) de la consulta q se define como

$$AP(q) = \frac{\sum_{i=1}^{|Rec|} (P(i) \times rel(i))}{|R^q|}$$

Donde

- $P(i)$ es la precisión en la posición i-esima
- $rel(i)$ es 1 si en la posición i-ésima se recupera un doc relevante y cero en caso contrario.
- $|R^q|$ hace referencia al número de doc relevantes recuperado para la consulta q

NDCG

- DCG es la ganancia acumulada en una posición p

$$DCG_p = \sum_{i=1}^p \frac{2^{rel_i} - 1}{log(1+i)}$$

utilizada por compañías web, hace énfasis en obtener documentos altamente relevantes.

El problema es que depende del número de documentos relevantes a la consulta, por ello se normaliza

$$NDCG_p = \frac{DCG_p}{DCG_{ideal}}$$

Evaluación de los S.R.I.

- Un buen S.R.I. debería
 - presentar un recall alto (recuperaría el mayor número posible de documentos relevantes),
 - y tener una precisión también alta (recuperar muy pocos no relevantes).