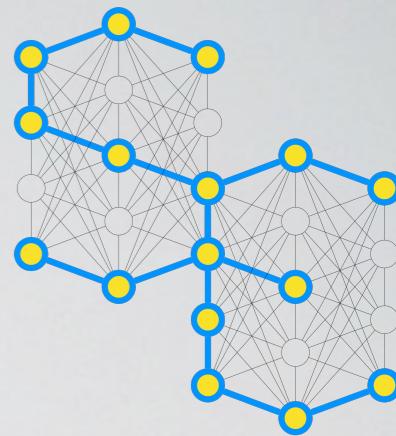


800
A N N I
1222 • 2022



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



Introduction to Information Retrieval

Search Engines

Master Degree in Computer Engineering

Master Degree in Data Science

Academic Year 2023/2024



DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

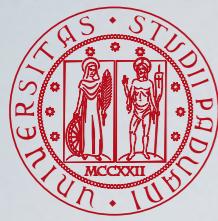
Nicola Ferro

Intelligent Interactive Information Access (IIIA) Hub

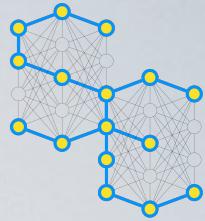
Department of Information Engineering

University of Padua





Outline



● What is Information Retrieval?

What is Information Retrieval?



THE UNIVERSAL DECLARATION OF Human Rights

WHEREAS recognition of the inherent dignity and of the equal and inalienable rights of all members of the human family is the foundation of freedom, justice and peace in the world,

WHEREAS disregard and contempt for human rights have resulted in barbarous acts which have outraged the conscience of mankind, and the advent of a world in which human beings shall enjoy freedom of speech and belief and freedom from fear and want has been proclaimed as the highest aspiration of the common people,

WHEREAS it is essential, if man is not to be compelled to have recourse, as a last resort, to rebellion against tyranny and oppression, that human rights should be protected by the rule of law,

WHEREAS it is essential to promote the development of friendly relations between nations,

WHEREAS the peoples of the United Nations have in the Charter reaffirmed their faith in fundamental human rights, in the dignity and worth of the

human person and in the equal rights of men and women and have determined to promote social progress and better standards of life in larger freedom,

WHEREAS Member States have pledged themselves to achieve, in cooperation with the United Nations, the promotion of universal respect for and observance of human rights and fundamental freedoms,

WHEREAS a common understanding of these rights and freedoms is of the greatest importance for the full realization of this pledge,

NOW THEREFORE THE GENERAL ASSEMBLY

Proclaims this Universal Declaration of Human Rights as a common standard of achievement for all peoples and all nations, to the end that every individual and every organ of society, keeping this Declaration constantly in mind, shall strive by teaching and education to promote respect for these rights and freedoms and by progressive measures, national and international, to secure their universal and effective recognition and observance, both among the peoples of Member States themselves and among the peoples of territories under their jurisdiction.

ARTICLE 1 All human beings are born free and equal in dignity and rights. They are endowed with reason and conscience and should act towards one another in a spirit of brotherhood.

ARTICLE 2 Everyone is entitled to all the rights and freedoms set forth in this Declaration, without distinction of any kind, such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status. Furthermore, no distinction shall be made on the basis of the political, jurisdictional or international status of the country or territory to which a person belongs, whether it be independent, trust, non-self-governing or under any other limitation of sovereignty.

ARTICLE 3 Everyone has the right to life, liberty and security of person.

ARTICLE 4 No one shall be held in slavery or servitude; slavery and the slave trade shall be prohibited in all their forms.

ARTICLE 14 (1) Everyone has the right to seek and to enjoy in other countries asylum from persecution. (2) This right may not be invoked in the case of prosecutions genuinely arising from non-political crimes or from acts contrary to the purposes and principles of the United Nations.

ARTICLE 15 (1) Everyone has the right to a nationality. (2) No one shall be arbitrarily deprived of his nationality nor denied the right to change his nationality.

ARTICLE 16 (1) Men and women of full age, without any limitation due to race, nationality or religion, have the right to marry and to found a family. They are entitled to equal rights as to marriage, during marriage and at its dissolution. (2) Marriage shall be entered into only with the free and full consent of the intending spouses. (3) The family is the natural and fundamental group unit of society and is entitled to protection by society and the State.

ARTICLE 17 (1) Everyone has the right to own property alone as well as in association with others. (2) No one shall be arbitrarily deprived of

an existence worthy of human dignity, and supplemented, if necessary, by other means of social protection. (4) Everyone has the right to form and to join trade unions for the protection of his interests.

ARTICLE 21 Everyone has the right to rest and leisure, including reasonable limitation of working hours and periodic holidays with pay.

ARTICLE 25 (1) Everyone has the right to a standard of living adequate for the health and wellbeing of himself and of his family, including food, clothing, housing and medical care and necessary social services, and the right to security in the event of unemployment, sickness, disability, widowhood, old age or other lack of livelihood in circumstances beyond his control. (2) Motherhood and childhood are entitled to special care and assistance. All children, whether born in or out of wedlock, shall enjoy the same social protection.

ARTICLE 26 (1) Everyone has the right to education. Education shall be free, at least in the elementary and secondary education. Higher



THE UNIVERSAL DECLARATION OF Human Rights

Article 19.

“Everyone has the right to **freedom of opinion and expression**; this right includes freedom to hold opinions without interference and to **seek, receive and impart information and ideas through any media and regardless of frontiers.**”

ARTICLE 1 All human beings are born free and equal in dignity and rights. They are endowed with reason and conscience and should act towards one another in a spirit of brotherhood.

ARTICLE 2 Everyone is entitled to all the rights and freedoms set forth in this Declaration, without distinction of any kind, such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status. Furthermore, no distinction shall be made on the basis of the political, jurisdictional or international status of the country or territory to which a person belongs, whether it be independent, trust, non-self-governing or under any other limitation of sovereignty.

ARTICLE 3 Everyone has the right to life, liberty and security of person.

ARTICLE 4 No one shall be held in slavery or servitude; slavery and the slave trade shall be prohibited in all their forms.

ARTICLE 14 (1) Everyone has the right to seek and to enjoy in other countries asylum from persecution. (2) This right may not be invoked in the case of prosecutions genuinely arising from non-political crimes or from acts contrary to the purposes and principles of the United Nations.

ARTICLE 15 (1) Everyone has the right to a nationality. (2) No one shall be arbitrarily deprived of his nationality nor denied the right to change his nationality.

ARTICLE 16 (1) Men and women of full age, without any limitation due to race, nationality or religion, have the right to marry and to found a family. They are entitled to equal rights as to marriage, during marriage and at its dissolution. (2) Marriage shall be entered into only with the free and full consent of the intending spouses. (3) The family is the natural and fundamental group unit of society and is entitled to protection by society and the State.

ARTICLE 17 (1) Everyone has the right to own property alone as well as in association with others. (2) No one shall be arbitrarily deprived of

an existence worthy of human dignity, and supplemented, if necessary, by other means of social protection. (4) Everyone has the right to form and to join trade unions for the protection of his interests.

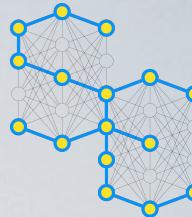
ARTICLE 21 Everyone has the right to rest and leisure, including reasonable limitation of working hours and periodic holidays with pay.

ARTICLE 25 (1) Everyone has the right to a standard of living adequate for the health and wellbeing of himself and of his family, including food, clothing, housing and medical care and necessary social services, and the right to security in the event of unemployment, sickness, disability, widowhood, old age or other lack of livelihood in circumstances beyond his control. (2) Motherhood and childhood are entitled to special care and assistance. All children, whether born in or out of wedlock, shall enjoy the same social protection.

ARTICLE 26 (1) Everyone has the right to education. Education shall be free, at least in the elementary and secondary education. Higher



Why Do We Need Information?



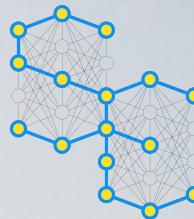
- To make decisions
- To identify or structure a problem or opportunity
- To put problem or opportunity in context
- To generate alternative solutions
- To choose the best alternative
- ...



[credits to Maarten de Rijke]

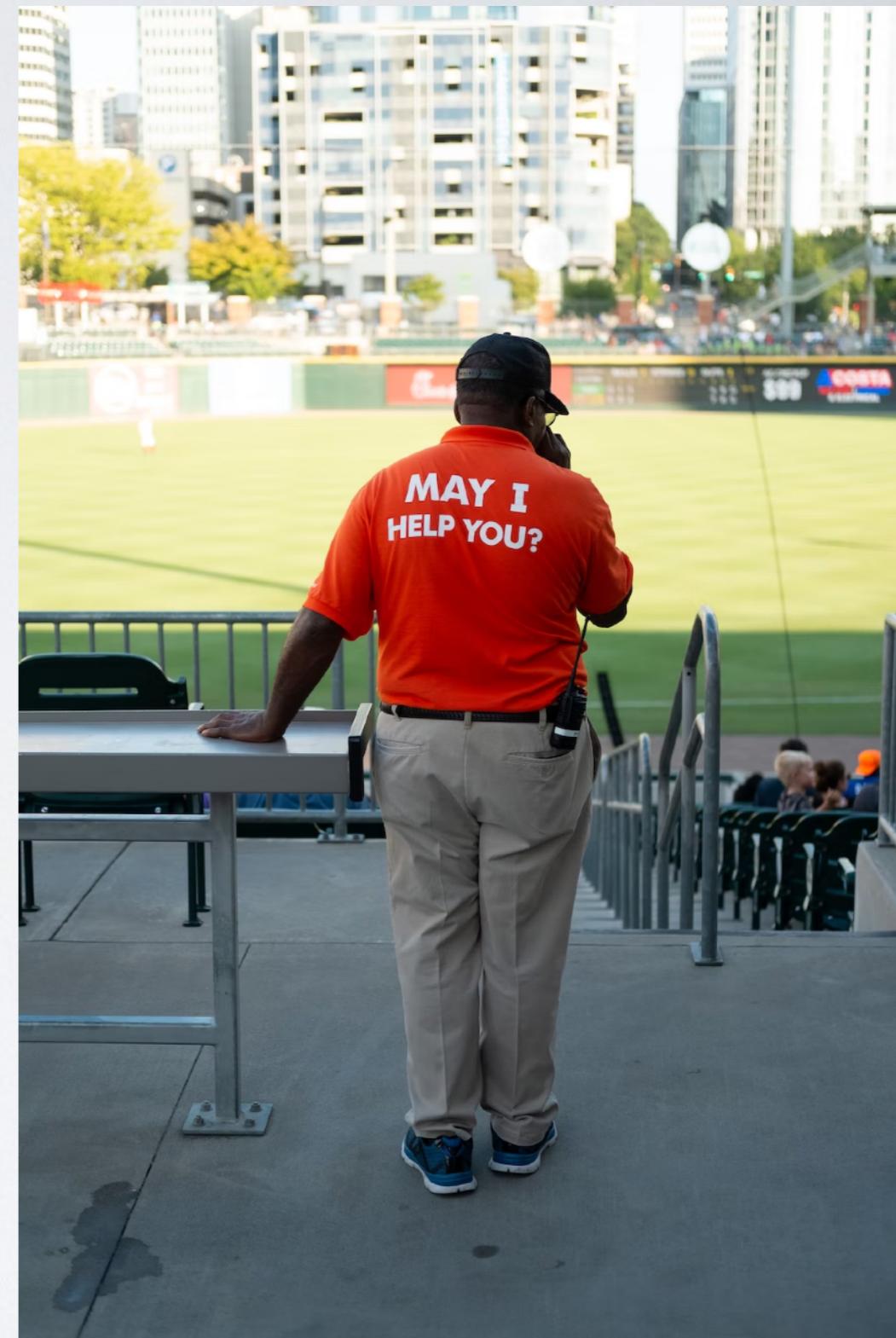


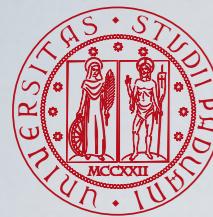
Why Do We Need Information? Examples



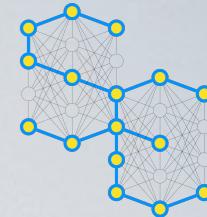
I need to **find** some **information**

- Who is the head of the School of Engineering? (precise need)
- How to get to the LUX city centre from LUX airport? (precise need)
- What is the upcoming topic in IR research? (vague need)
- What to do this weekend in Padua? (vague need)





IR: Web Search



Google

Яндекс

NAVER



Baidu 百度

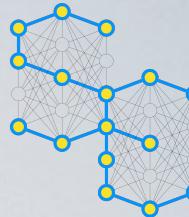
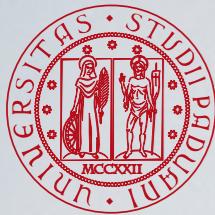


DuckDuckGo



Qwant

ECOSIA



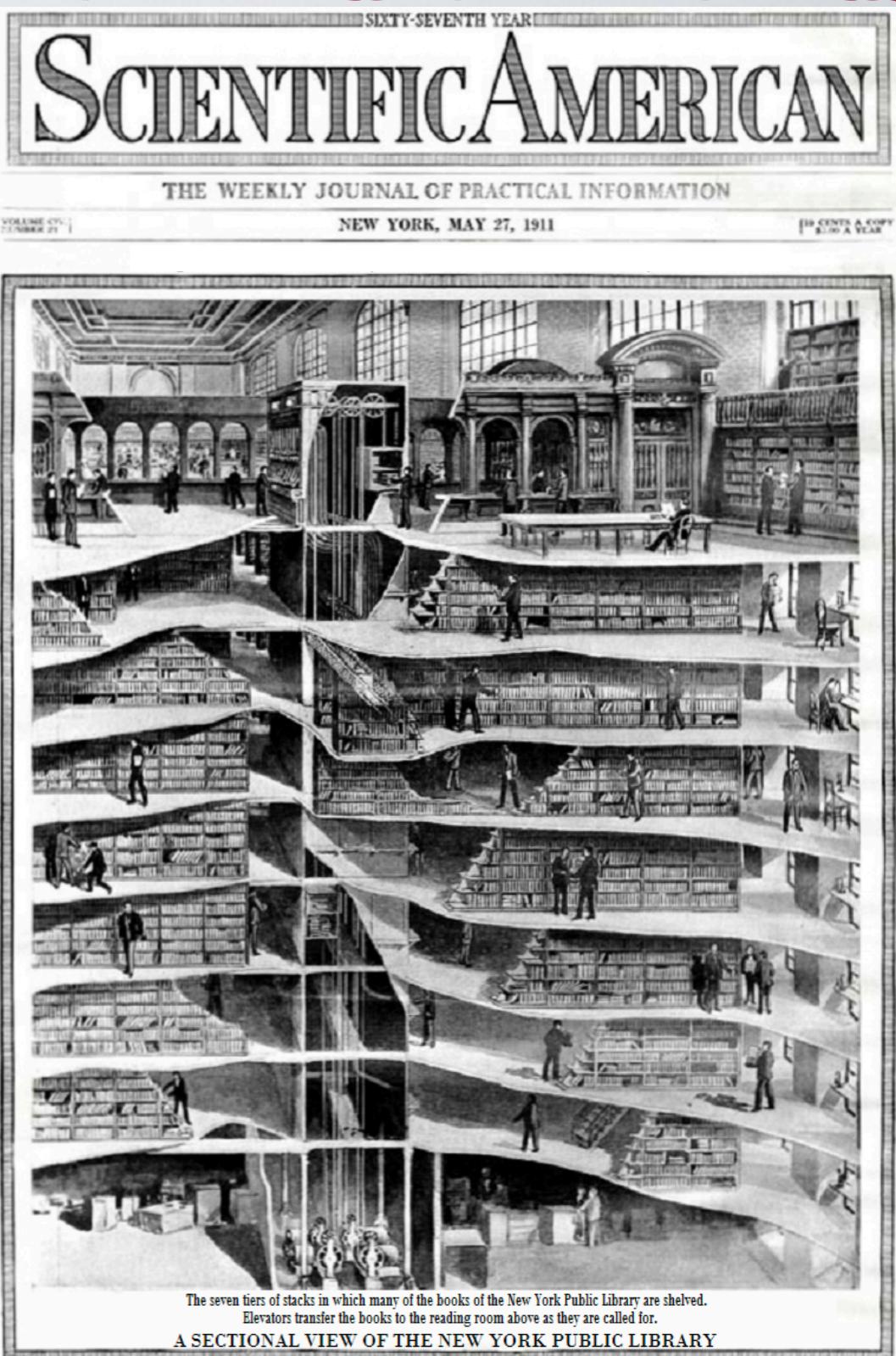
IR: Library Search

The seven tiers of stacks in which many of the books of the New York Public Library are shelved. Elevators transfer the books to the reading room above as they are called for.

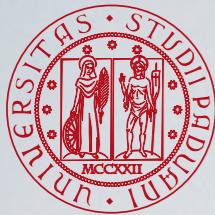
Scientific American, May 27, 1911, Volume 104, Issue 21

<https://www.scientificamerican.com/magazine/sa/1911/05-27/>

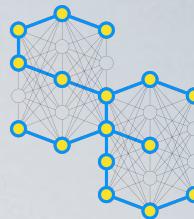
The screenshot shows the homepage of the University of Padova's library system, SBA (SISTEMA BIBLIOTECARIO DI ATENEO). The top navigation bar includes links for Staff, Chi sei, Accessibility, Help, Privacy, Italiano, English, and a search bar. The main menu categories are Digital Library, What are you looking for?, Tools, and About us. A banner at the bottom left promotes a virtual exhibition of Prof. Pecile's rare books. A green box highlights the search bar labeled "Search with GalileoDiscovery".



[credits to Diane Kelly]



IR: Social Media Search



LinkedIn search engines

Home My Network Jobs Messaging Notifications Me Try Premium Free for 1 Month

People Courses Companies Groups Jobs Schools Posts Events All filters

People

- Shiroy Choksey - 2nd Metrics, Measurement, Machine Learning, Offline measurement, Human Labelling, Search engines... Sunnyvale, CA
- Andreas Wagner - 2nd Helping Search Engines understand Humans Pforzheim Skills: Site: Search Analytics
- Junte Zhang - 2nd Making new things possible with search engines Amsterdam Current: Principal Search Engineer at Optimizely

See all people results

Courses

- SEO: E-commerce Strategies Course • 1h 2m By: Sam Dey • Released Sep 1, 2020 3.4K viewers
- Elasticsearch Essential Training Course • 1h 31m By: Ben Sullins • Released Nov 7, 2017 24.7K viewers
- Learning PPC with Google Ads Course • 1h 4m By: Duane Brown • Released Nov 9, 2020 1.6K viewers

See all course results

Companies

- Search Engines MD Automotive 22 followers

Search Engines MD is an internet marketing company that focuses on Facebook marketing, Search Engine Marketing and Search Engine Optimization. Their goal from...

Instagram search engines

We've Made Our

#searchengines 23,782 posts

search_engine_seo

searchenginesex

searchengineland

tagwalk

search_engine.optimization

deiunipd il mattino

«L'uso della mascherina ha evitato 30 mila contagi»

L'esperienza del ministero dell'Innovazione ha premiato la città di Padova come la migliore in Italia durante le prime undici settimane di aprile. Lo studio è stato condotto attraverso l'analisi dei dati sui contagi pubblicati quotidianamente sul sito dell'Istituto Superiore di Sanità. La corrispondenza tra i dati della Regione Veneto e quelli del ministero dell'Innovazione dimostra che la diffusione dei contagi in Veneto è stata contenuta grazie alla massiccia campagna di sensibilizzazione e alle misure di controllo impostate dal governo. Il ministero ha quindi deciso di premiare la città di Padova con il titolo di "Città più pulita d'Italia".

Il presidente della Regione Veneto, Luca Zaia, ha dichiarato: «A Padova abbiamo dimostrato che è possibile contenere i contagi senza ricorrere alla chiusura totale delle scuole. Abbiamo provveduto a limitare i contatti sociali e a garantire la sicurezza dei cittadini, dove ciò nonostante i

contagi sono stati contenuti. Questo è un esempio per tutti gli italiani di come si può combattere la pandemia in modo efficace e responsabile».

L'esperienza del ministero dell'Innovazione ha premiato la città di Padova come la migliore in Italia durante le prime undici settimane di aprile. Lo studio è stato condotto attraverso l'analisi dei dati sui contagi pubblicati quotidianamente sul sito dell'Istituto Superiore di Sanità. La corrispondenza tra i dati della Regione Veneto e quelli del ministero dell'Innovazione dimostra che la diffusione dei contagi in Veneto è stata contenuta grazie alla massiccia campagna di sensibilizzazione e alle misure di controllo impostate dal governo. Il ministero ha quindi deciso di premiare la città di Padova con il titolo di "Città più pulita d'Italia".

Il presidente della Regione Veneto, Luca Zaia, ha dichiarato: «A Padova abbiamo dimostrato che è possibile contenere i contagi senza ricorrere alla chiusura totale delle scuole. Abbiamo provveduto a limitare i contatti sociali e a garantire la sicurezza dei cittadini, dove ciò nonostante i

Liked by cuspadovacalcio5 and 82 others

deiunipd #deiresearch ... more

2 DAYS AGO

Add a comment... Post

Twitter search engines

Home

Esplora

Notifiche

Messaggi

Segnalibri

Liste

Profilo

Altro

Twitta

Popolari Recent Persone Foto Video

Danny #BlackLivesMatter @DannyAligh... - 9min A start would be for #schools to CHANGE the default search engines on their PCs, #tablets & #laptops to something other than the behavioural data scraping #Google - show the children that there are alternatives to big tech. The use of #Gmail & #Chrome should also be reevaluated

RiuNet @RiuNetUPV - 13min Crossing the academic ocean? Judit Bar-Ilan's oeuvre on search engines studies riunet.upv.es/handle/10251/1...

Growth Hacking Briefly @GrowthHack_b - 14min Using SEO to Obtain First Page Rankings (Adlibweb)

The measure of success for an online business site is where it ranks on search engines. Part of pa...

Add your highlights: briefly.co/anchor/Growth... #GrowthHacking #market...

Using SEO to Obtain First Page Rankings | Growth hacking | Briefly & briefly.co

Nicola Ferro @frncl

© Nicola Ferro

Filtri di ricerca

Personne Da chiunque Personne che seguì

Posizione Ovunque Vicino a te

Ricerca avanzata

Tendenze per te

Intrattenimento generale - Di tendenza Mark Ruffalo 10.700 Tweet

#LionsOfOurTime Sei un leone del nostro tempo? Sponsorizzato da Peugeot

Di tendenza nella seguente zona: Italia Laura Pausini 7.377 Tweet

Di tendenza nella seguente zona: Italia Stanley Tucci 2.823 Tweet

Di tendenza nella seguente zona: Italia Di Canio

Mostra altro

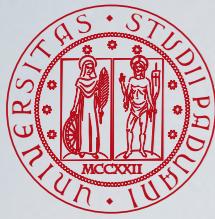
Chi segue

Mohammad Aliannejad @maliannejadi Ti segue Segui

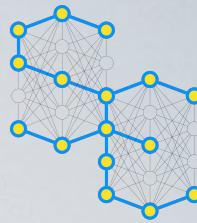
AIMultimediaLab @AIMultimed... Ti segue Segui

isi 2021 @isi2021 Ti segue Segui

Mostra altro



IR: Desktop Search

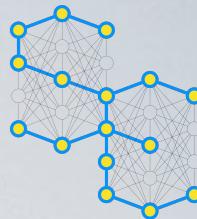


The image displays three separate windows illustrating desktop search capabilities:

- Mac OS X Finder Window:** Shows a search for "Macbeth" across "This Mac" and "Documents". Results include various PDF and EPUB files related to Shakespeare's play.
- Windows File Explorer Window:** Shows a search for "Windows" in the "Documents library". Results include several ISO files for different versions of Microsoft Windows.
- Terminal Window (Ubuntu):** Shows command-line search results for ".jpg" and ".png" files. It includes commands like "find", "grep", and "ls" to locate specific files.



IR: Mail Search



Tutte le caselle (La ricerca ha prodotto 699 risultati)

Ricevi Nuovo messaggio Stampa Archivia Elimina Indesiderata Rispondi Rispondi a tutti Inoltra Contrassegno Attiva notifiche Sposta

Caselle Caselle Cerca: Tutte Entrata - Ferro DEI Entrata (3) VIP Invia Contrassegnata Bozze

Entrata (1) Ferro DEI (2) Sofia Gmail (641) Ferro Gmail Contrassegnata (641) Bozze Invitata Sul mio Mac Ferro DEI Ferro PEC Sofia Gmail Ferro Gmail Indesiderata Cestino Archivio

Ordina per data

Zoom 09:55 VACCARO FABIO has joined your meeting - Web Application... Cestino - Ferro DEI Hi nicola.ferro@unipd.it, VACCARO FABIO has joined your meeting: Topic Web Applications - Lectures Meeting ID 881 9238 1370 Start Meeting Thank you for choosing Zoom. -The Zoom Team Copyright ©2020 Zoom Video Commun...

Zoom 09:52 FARINEA LORENZO has joined your meeting - Web Application... Cestino - Ferro DEI Hi nicola.ferro@unipd.it, FARINEA LORENZO has joined your meeting: Topic Web Applications - Lectures Meeting ID 881 9238 1370 Start Meeting Thank you for choosing Zoom. -The Zoom Team Copyright ©2020 Zoom Video Commun...

Zoom 09:50 stefano.romanello.1@studenti.unipd.it has joined your meeting... Cestino - Ferro DEI Hi nicola.ferro@unipd.it, stefano.romanello.1@studenti.unipd.it has joined your meeting: Topic Web Applications - Lectures Meeting ID 881 9238 1370 Start Meeting Thank you for choosing Zoom. -The Zoom Team Copyright ©2020 Zoom Video Commun...

Zoom 09:26 ESPOSITO VITTORIO has joined your meeting - Web Application... Cestino - Ferro DEI Hi nicola.ferro@unipd.it, ESPOSITO VITTORIO has joined your meeting: Topic Web Applications - Tutor Meeting ID 825 2740 9383 Start Meeting Thank you for choosing Zoom. -The Zoom Team Copyright ©2020 Zoom Video Commun...

Zoom 09:11 ESPOSITO VITTORIO has joined your meeting - Web Application... Cestino - Ferro DEI Hi nicola.ferro@unipd.it, ESPOSITO VITTORIO has joined your meeting: Topic Web Applications - Lectures Meeting ID 881 9238 1370 Start Meeting Thank you for choosing Zoom. -The Zoom Team Copyright ©2020 Zoom Video Commun...

Q web applications

Oggetto: L'oggetto contiene: web applications [WEB APPLICATIONS] chiarimento pr... WEB APPLICATIONS Richiesta Info R.I.A. - corso WEB APP... Modalità erogazione corso secondo s...

Gmail

Compose

Inbox Starred Snoozed Sent Drafts Book Club Europe Trip Planning finances misc communication

articles

articles articles thursday iphone articles articles friday

Search the web for "articles"

Shift + Enter

out 82 < > 31 10:22 AM Mar 6

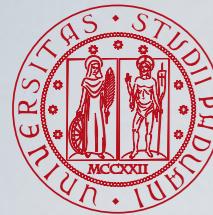
be great: "Wh... SAS | Business I...

Thursday 3/5 articles - " Best, Marissa -- Marissa L. Perino NYU GSAS | Business Insider... Mar 5

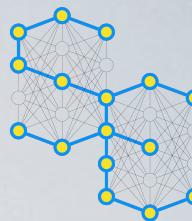
"How to add mu... "How to mute F...

Wednesday 3/4 articles - Facebook ." Best, Marissa -- Marissa L. Perino NYU GSAS | Bu... Mar 4

"How to add mu... "How to share a...



IR: Product Search

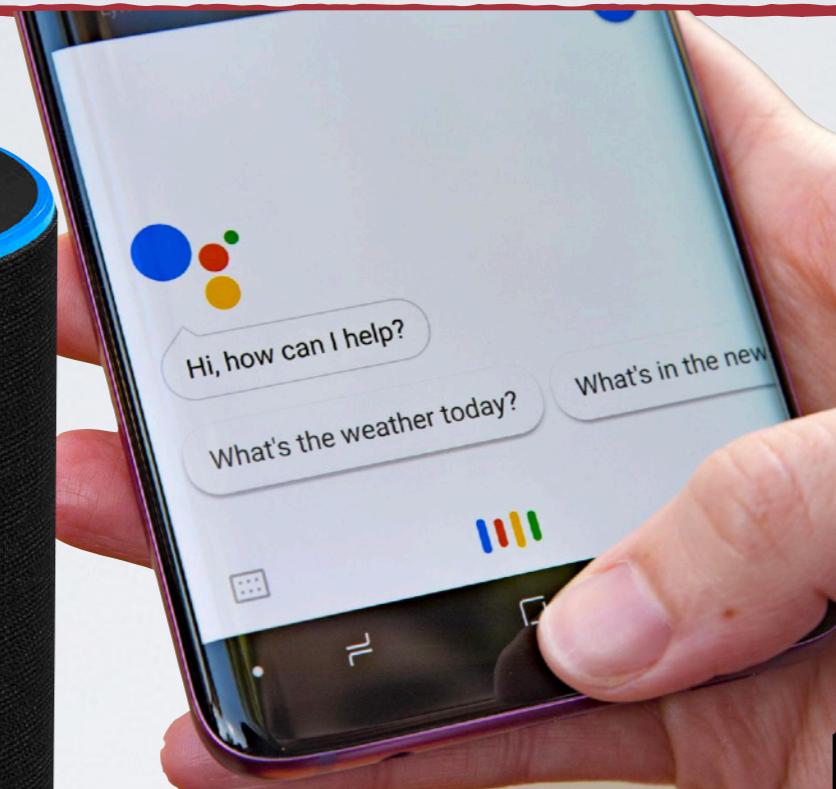
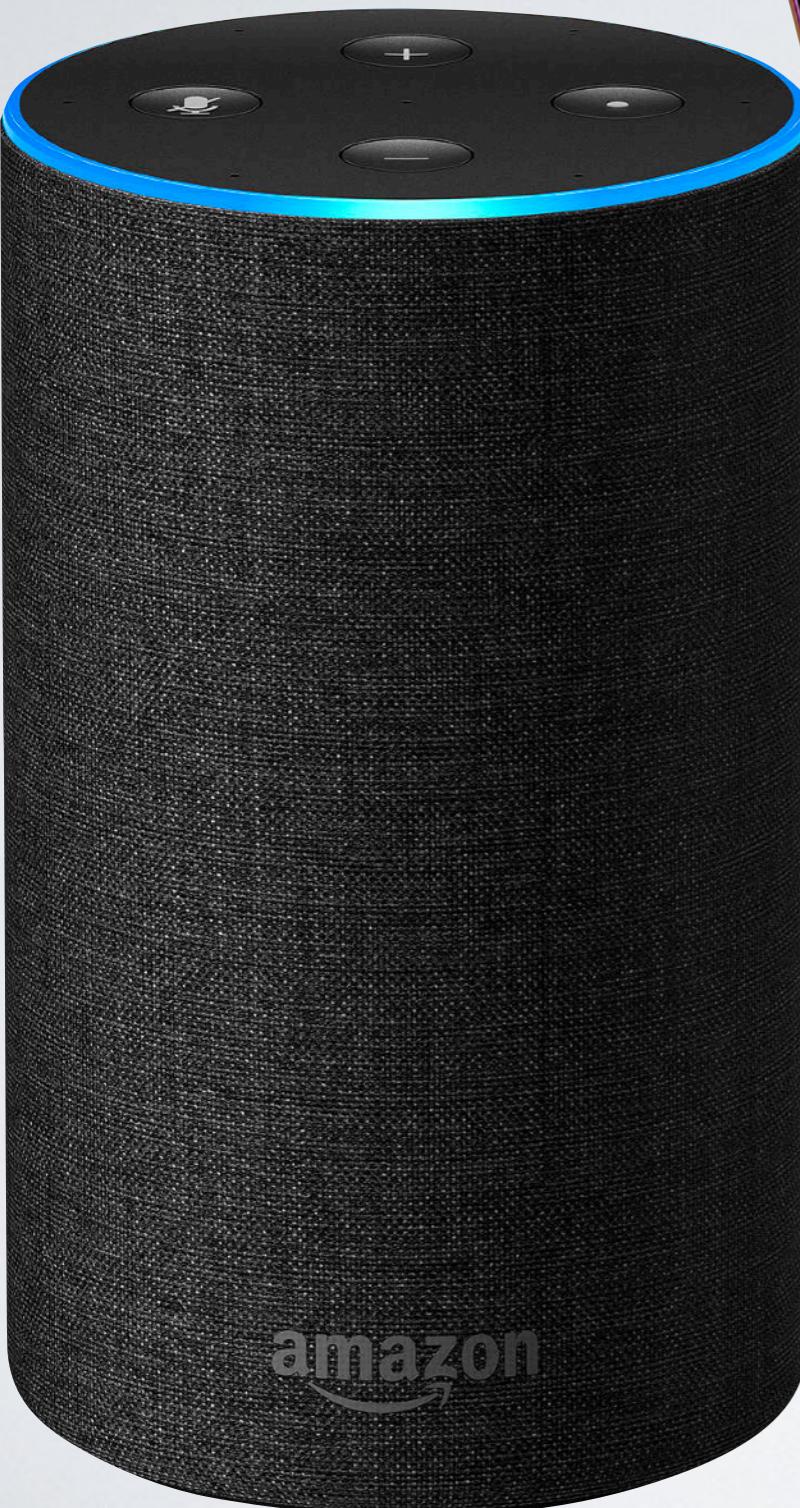
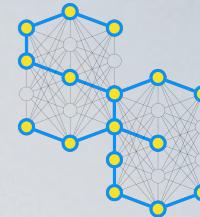


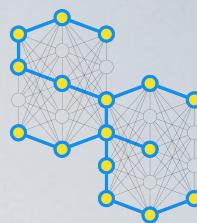
The collage consists of four separate web pages:

- Amazon Superstore:** Shows a search result for "search engines". It lists four products: "Search Engine Optimization All-in-One Dummies", "Librarian The First Search Engine", "SEO 2020", and "Divertente tazza SEO per SEO". Each product has a thumbnail, title, author, price, and a brief description.
- Spotify:** Shows a search result for "search engines" on the Spotify platform. It displays a "Top result" for the podcast "Search Engine Sex" and sections for "Songs", "Artists", and "Albums".
- Search engines (Product Hunt):** Shows a search result for "search engines" on the Product Hunt website. It lists several products: Ecosia, Quick Search - Find Any Item, Varians The Original Coffee Mug, Search Engine Switcher, DuckDuckGo Privacy Essentials, CopyPop - copy search eng..., and Webmii.
- YouTube:** Shows a search result for "search engines" on YouTube. It lists several video results: "Ecommerce Search Engine - Ricerca Veloce e Sicura", "I Migliori Search Engines For Website - Ricerca Protetta e Sicura", "20 Search Engines That Are Better Than Google!", "The Internet: How Search Works", "Most Popular Search Engines 1994 - 2019", and "Top 5 Best Search Engines That Do Not Track You!".



IR: Conversational Search





IR: Image Search

Google search engine

All Images Videos News Maps More Tools Collections SafeSearch

google logo icon yahoo! bing crawler web internet >

20 Great Search Engines You Can Use ...
searchenginejournal.com

Top 10 Search Engines In The World ...
reliablesoft.net

How Search Engines Work
aeroadmin.com

Top 8 Search Engines
wildstonesolution.com

Is Search Engine Submission Necessary?
seodesignchicago.com

Alternative Search Engines To Use in 2022
kinsta.com

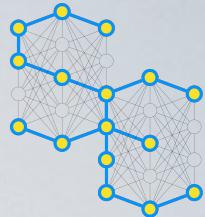
Browsers and Search Engines: What is ...
digitalnuisance.com

226 likes
maxvoltar Not deadmau5.
daveedgamboa Very cool
yurigoytacaz deadmnni3, then?
tantek deadfro9

© Nicola Ferro

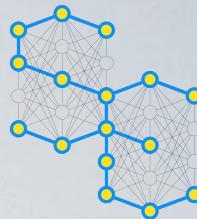


IR: Audio Search





IR: Legal Search



Adobe Flash Player non è più supportato

Home page

Normativa

Giurisprudenza

Collezioni

Newsletter

Utente

Aderenti al progetto:



Università degli Studi di
Modena e Reggio Emilia

- [Chi siamo](#)
- [Abbonamenti](#)
- [Redazione
scientifica](#)
- [Registrati adesso](#)
- [Contatti](#)
- [News](#)
- [F.A.Q.](#)

Cerca con infoLEGES.it

Ricerca semplice Cerca frase esatta

Normativa

[Normativa Comunitaria](#)

333.268 documenti

[Normativa Statale](#)

72.839 documenti

[Normativa Regionale](#)

40.921 documenti

[DoGi - Dottrina Giuridica](#)

576.138 documenti

[Giurisprudenza Comunitaria](#)

11.691 documenti

[Giurisprudenza Italiana](#)

183.600 documenti

Collezioni

[Gazzetta Ufficiale](#)

281.679 documenti

[Contratti collettivi](#)

949 documenti

[Concorsi pubblici](#)

63.609 documenti

Aderente a:

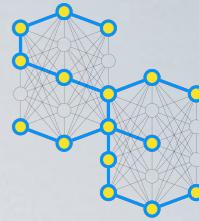
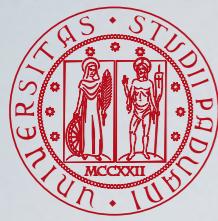


Partner scientifico:

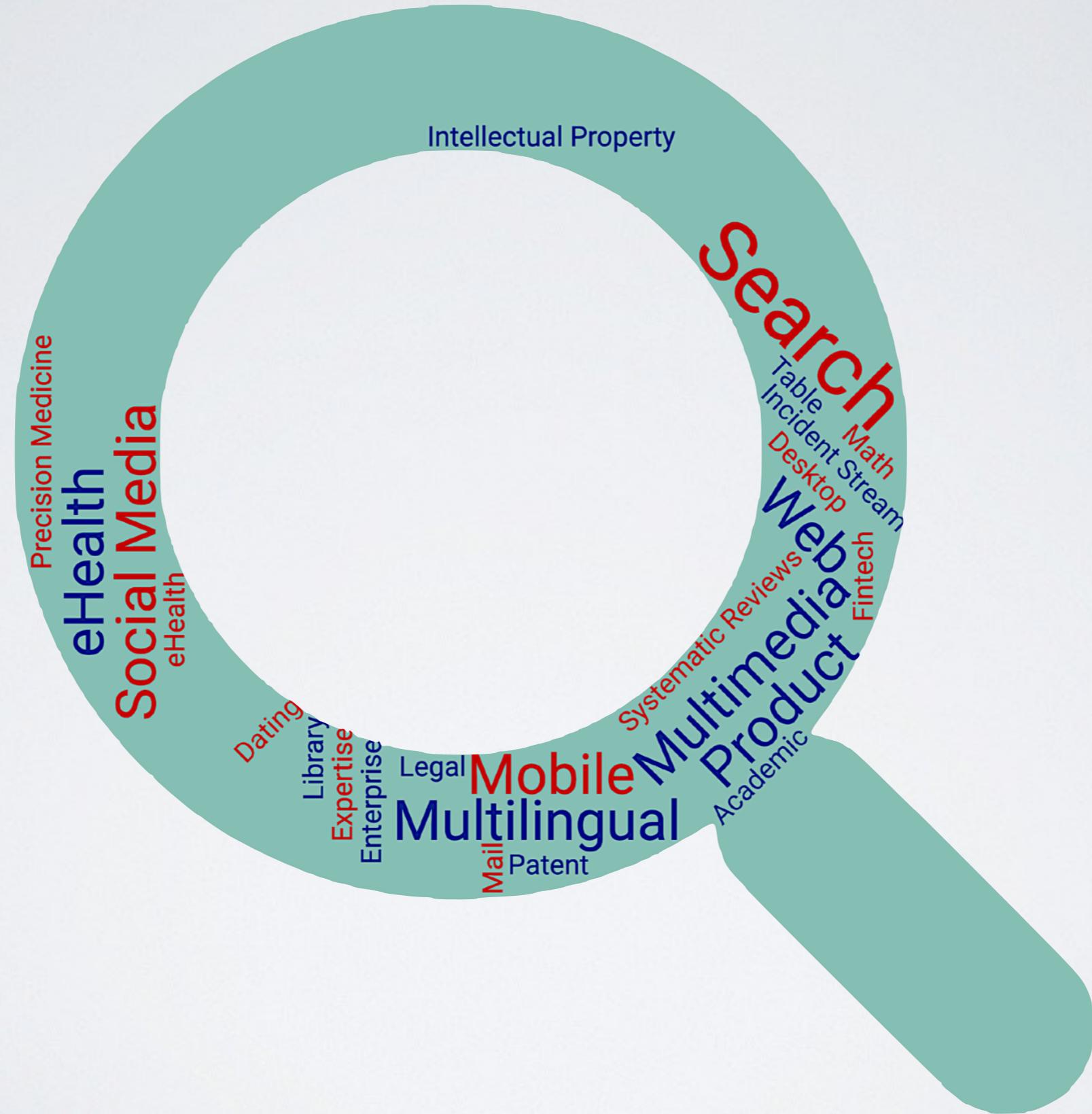


Partner tecnico:



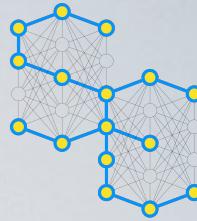


IR... IR... IR... Everywhere





R... IR... IR... Everywhere

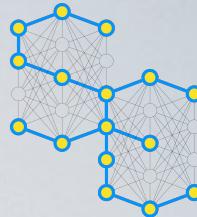


I am looking for someone
to share in an adventure.
Will you cover this gap
and be the next in IR?





What is Information Retrieval?

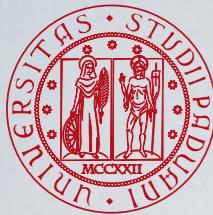


Information retrieval is a field concerned with the **structure, analysis, organization, storage, searching, and retrieval** of **information**.

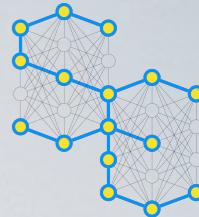
Salton, G. (1968). *Automatic Information Organization and Retrieval*. McGraw-Hill, New York, USA.



Gerald Salton



Where is Information? Documents!

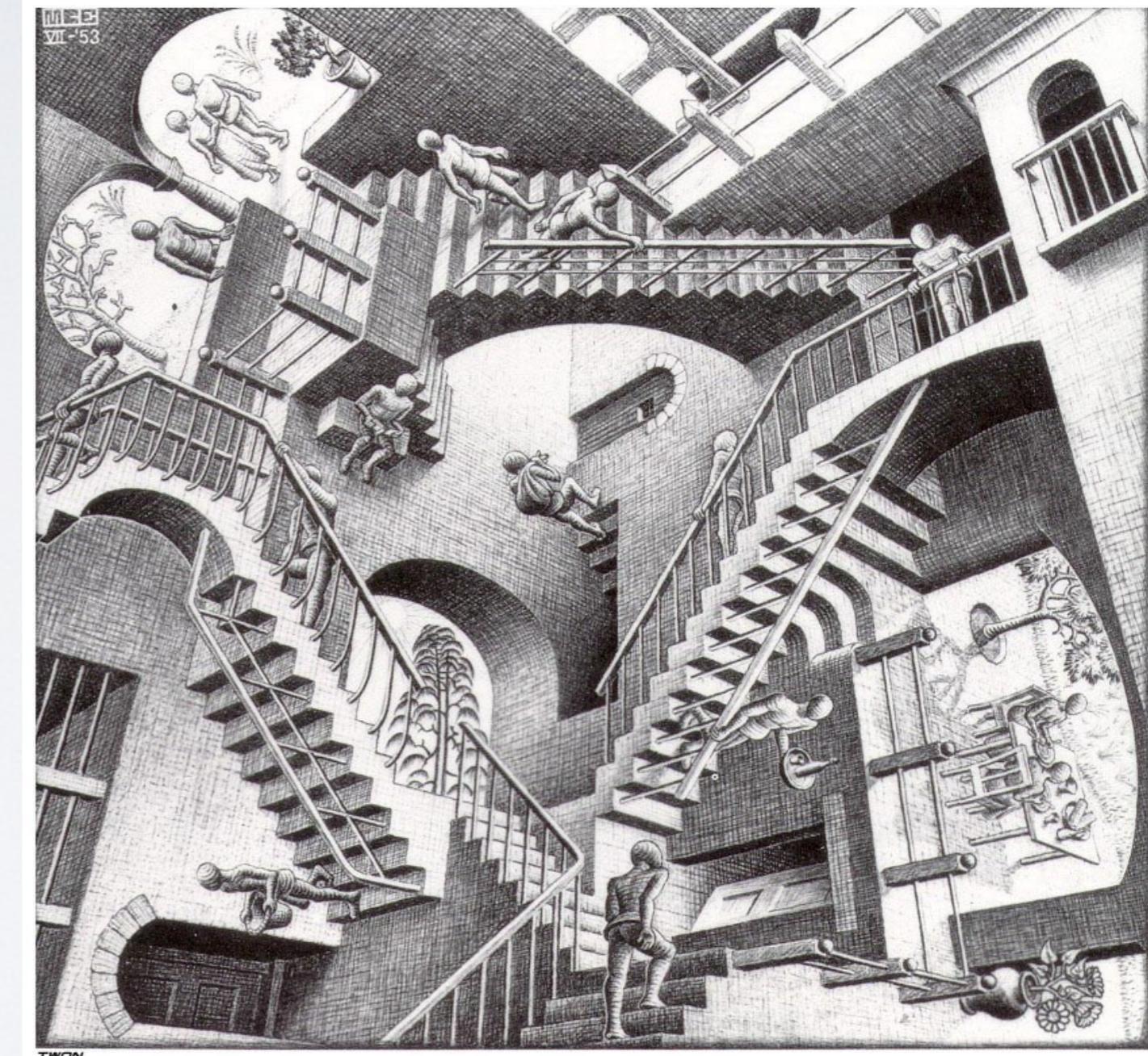


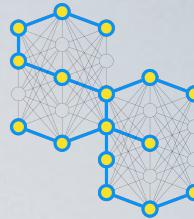
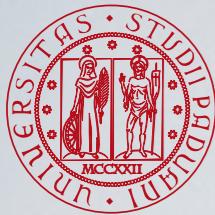
- Primary focus of IR since the 50s has been on *text* and **documents**
 - nowadays, routinely, also images, audio, video
- Some examples of documents
 - web pages, email, books, news stories, scholarly papers, text messages, Word™, Powerpoint™, PDF, forum postings, patents, IM sessions, etc.
- Documents may have some **organization**
 - title, author, date for papers
 - subject, sender, destination for emails
 - introduction, verse, pre-chorus, refrain, post-chorus, outro, ... for a song

What does retrieval mean?

Find something that you are looking for:

- **Ad hoc search**
 - Find documents “about this” topic-x
- **Known item search**
 - Find the University of Padua home page
- **Answer seeking**
 - What is the capital of Belgium?
- **Directed exploration**
 - Who builds video conferencing systems?
- **Decision making**
 - Best places to stay in Paris
- **Expert search**
 - Who knows about stochastic processes in my organisation?
- **Etc.**

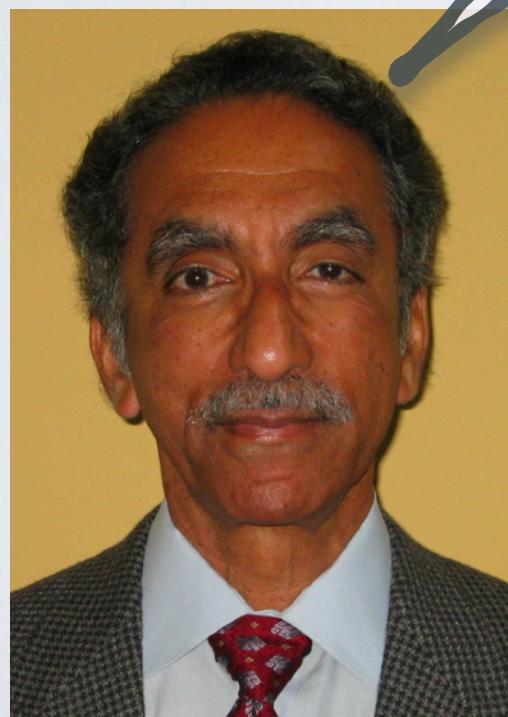




What is Information Retrieval?

A **database** is a collection of **related data**. By **data**, we mean **known facts** that can be **recorded** and that have implicit meaning. A **database management system** (DBMS) is a computerized system that enables users to **create** and **maintain** a database.

Elmasri, R. and Navathe, S. B. (2015). *Fundamentals of Database Systems*. Pearson Education Limited, USA, 7th edition.



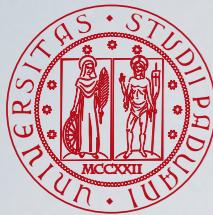
Information retrieval is a field concerned with the **structure, analysis, organization, storage, searching, and retrieval** of **information**.

Salton, G. (1968). *Automatic Information Organization and Retrieval*. McGraw-Hill, New York, USA.

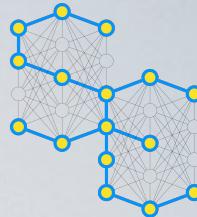
Shamkant B Navathe



Gerald Salton



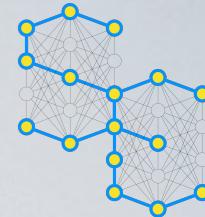
Documents vs Database Records



- Database records (or tuples in relational databases) are typically made up of well-defined *attributes*
 - bank records with account numbers, balances, names, addresses, social security numbers, dates of birth, etc.
- Easy to compare fields with **well-defined semantics** to queries in order to find matches
- Text is more difficult
 - even more, multimedia



Documents vs Database Records

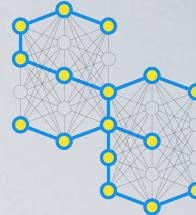
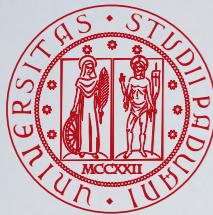


● Example bank database query

- Find records with balance > €5,000 in branches located in Padua, Veneto.
- Matches easily found by comparison with field values of records

● Example search engine query

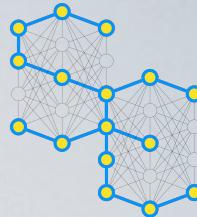
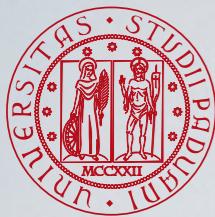
- bank scandals in Veneto region
- This text must be compared to the text of entire news stories



Comparing Text

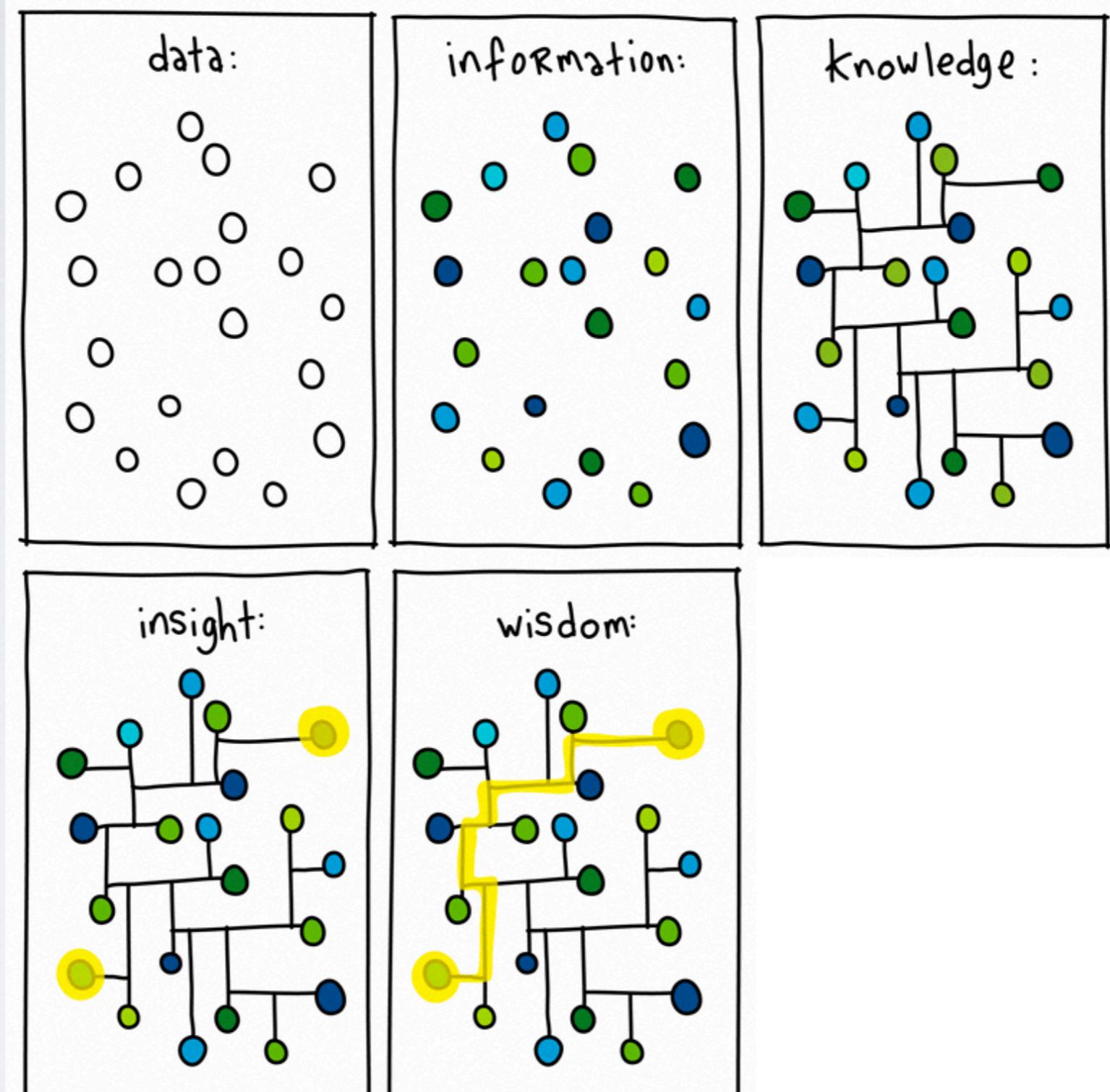
- Comparing the query text to the document text and determining what is a good match is the **core issue** of information retrieval
- Exact matching of words is not enough
 - Many different ways to write the same thing in a “natural language” like English
 - e.g., does a news story containing the text “ABN Amro wins Antonveneta control” match the query?
- Some stories will be better matches than others

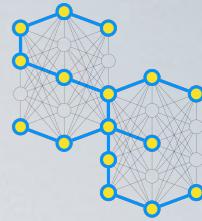
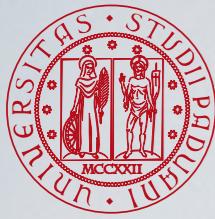
<https://en.wikipedia.org/wiki/Bancopoli>



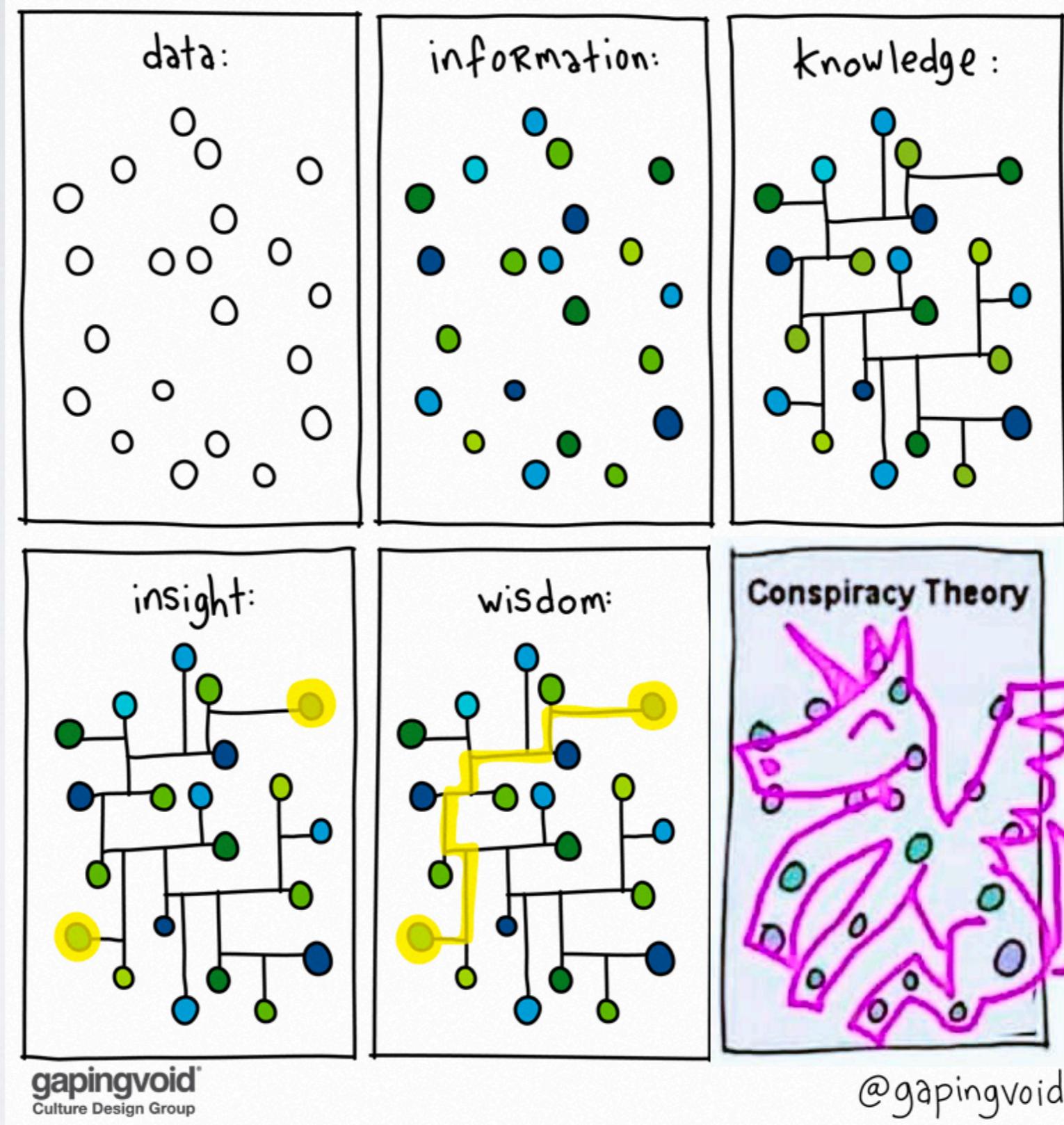
Data, Information, ...

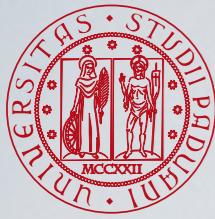
DATA INFORMATION
KNOWLEDGE AND
WISDOM pyramid



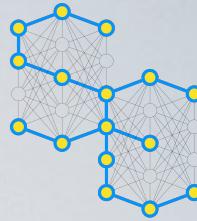


Data, Information, ...

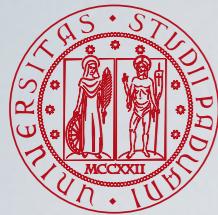




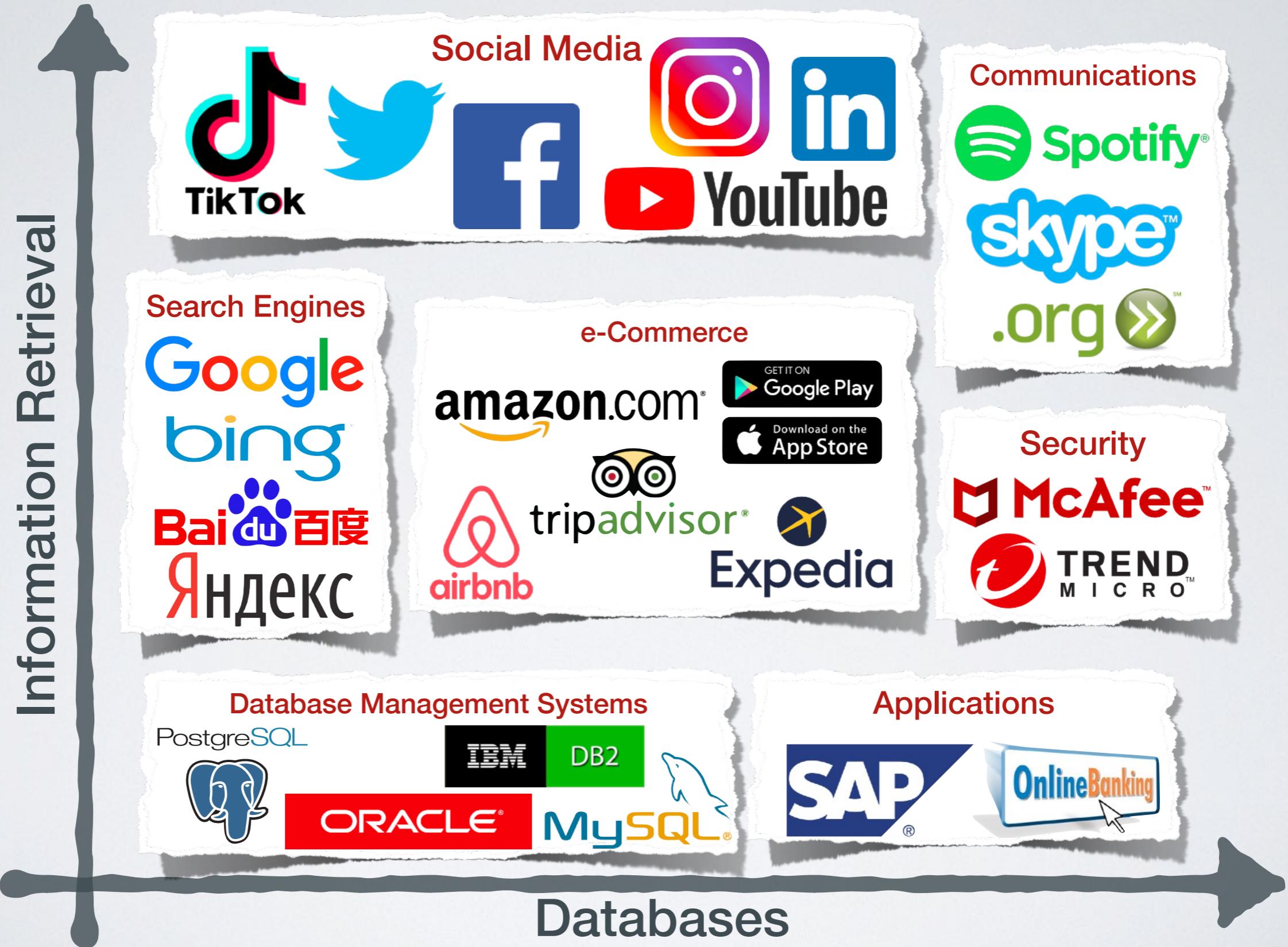
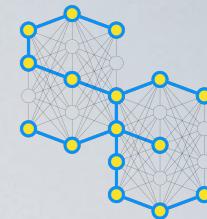
Databases vs Information Retrieval

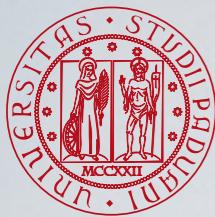


	Databases	IR
Unit	<u>Structured</u> data. <u>Clear semantics</u> based on a formal model <i>data types, attributes</i>	^{or semi-structured} Mostly <u>unstructured</u> data. Eg. free text with some organization
Queries	<u>Formally defined</u> queries. <u>Unambiguous</u> <i>not user-friendly</i>	Vague and imprecise information needs
Results	Correct in a formal sense. <u>Closed world assumption</u> <i>everything is in database, otherwise it does not exist</i>	^{depends on user} Sometimes relevant , often not. <u>Open world assumption</u>
Match	<i>sets: unique unordered elements</i> <u>Exact match</u> <i>evaluate proposition in a logic sense</i> No Ranking	<u>Best match</u> <u>Ranked</u>
Interaction	One-shot queries <u>No interaction</u>	<u>Interaction is fundamental</u>

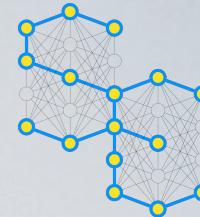


Databases vs Information Retrieval

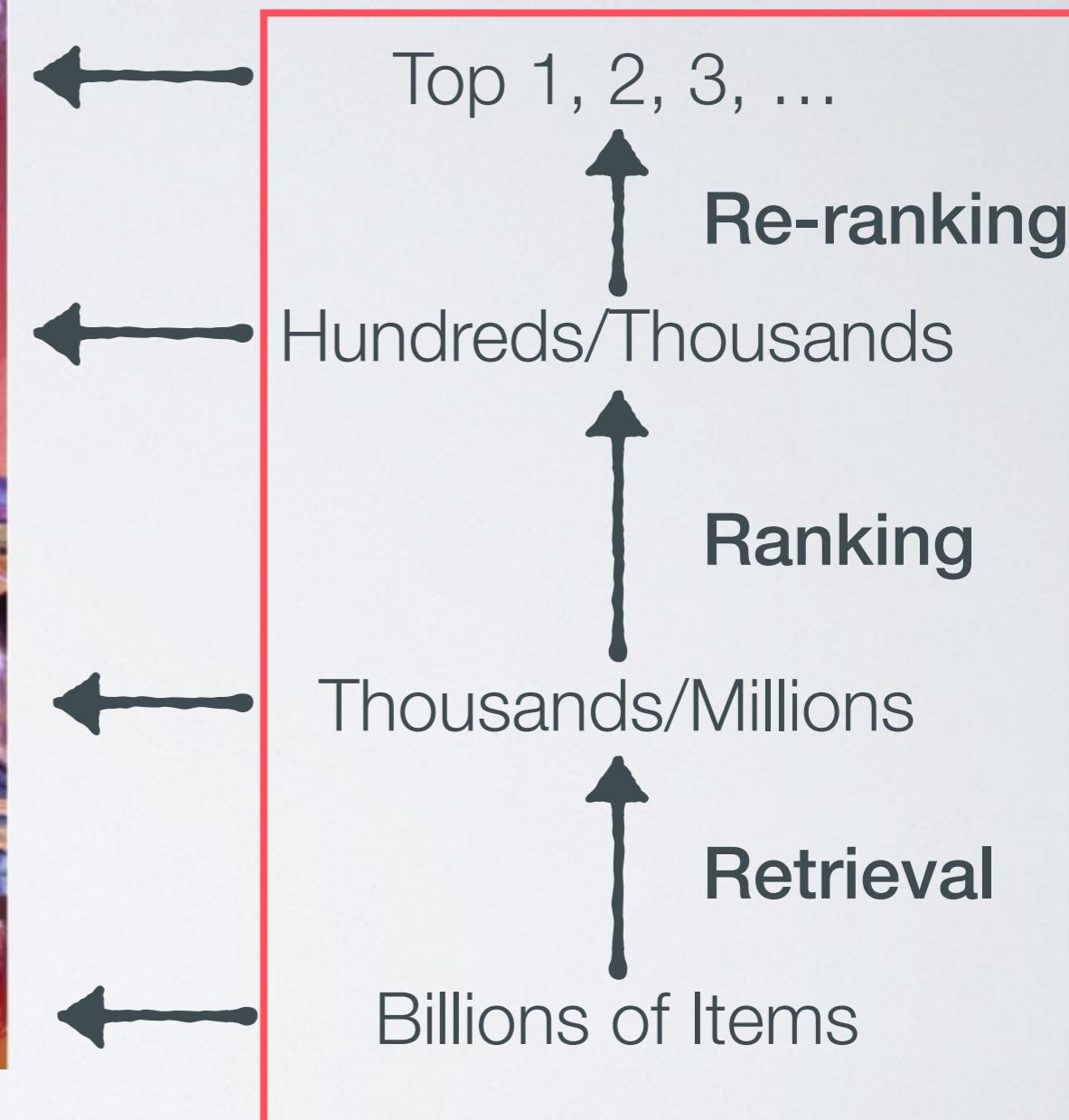
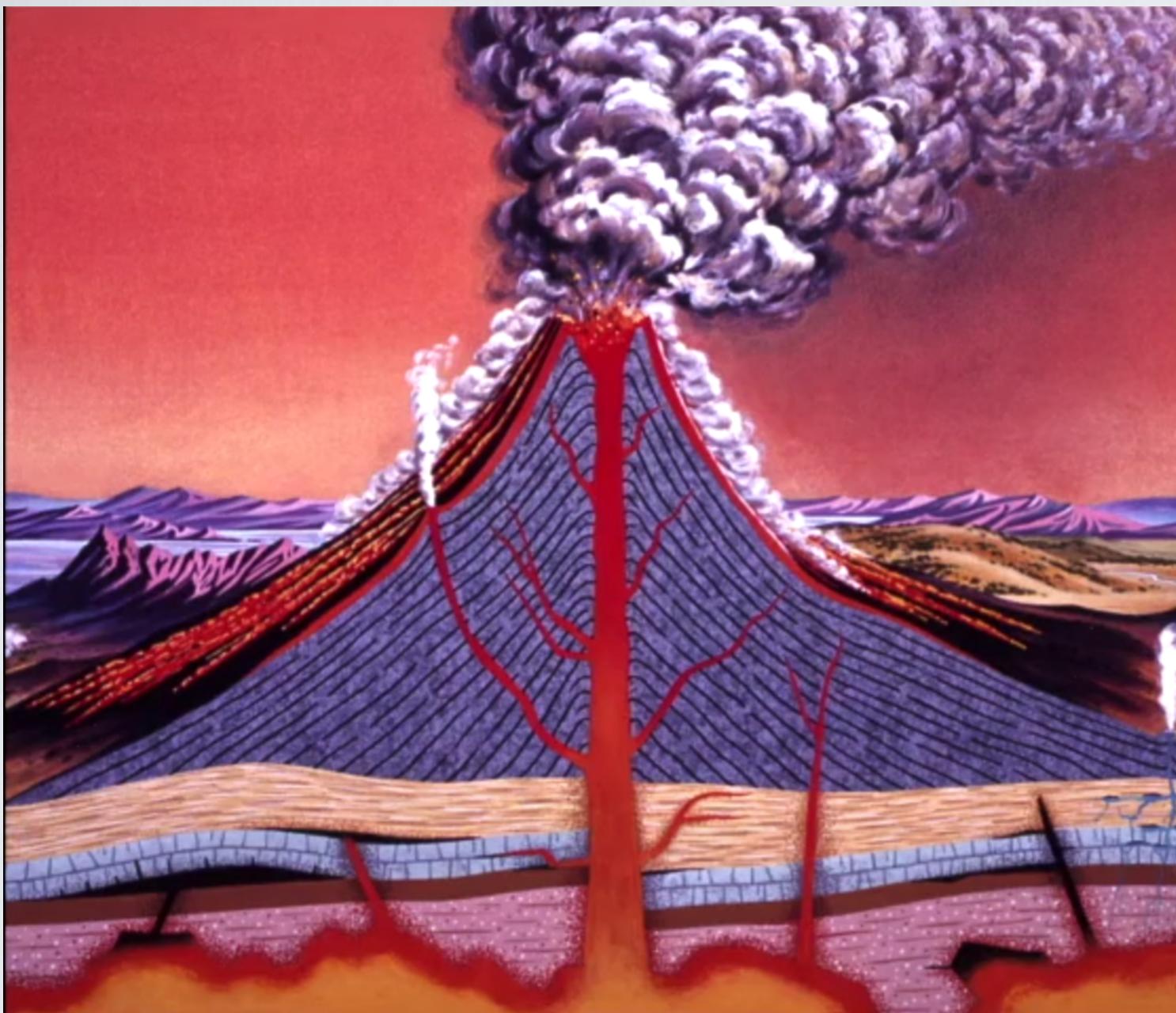




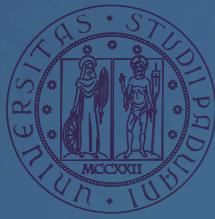
Large-scale Multi-stage Ranking



ranking methods: kinda old, but still widely used
beat DL in some applications



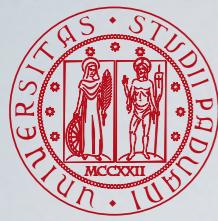
[credits to Charlie Clarke]



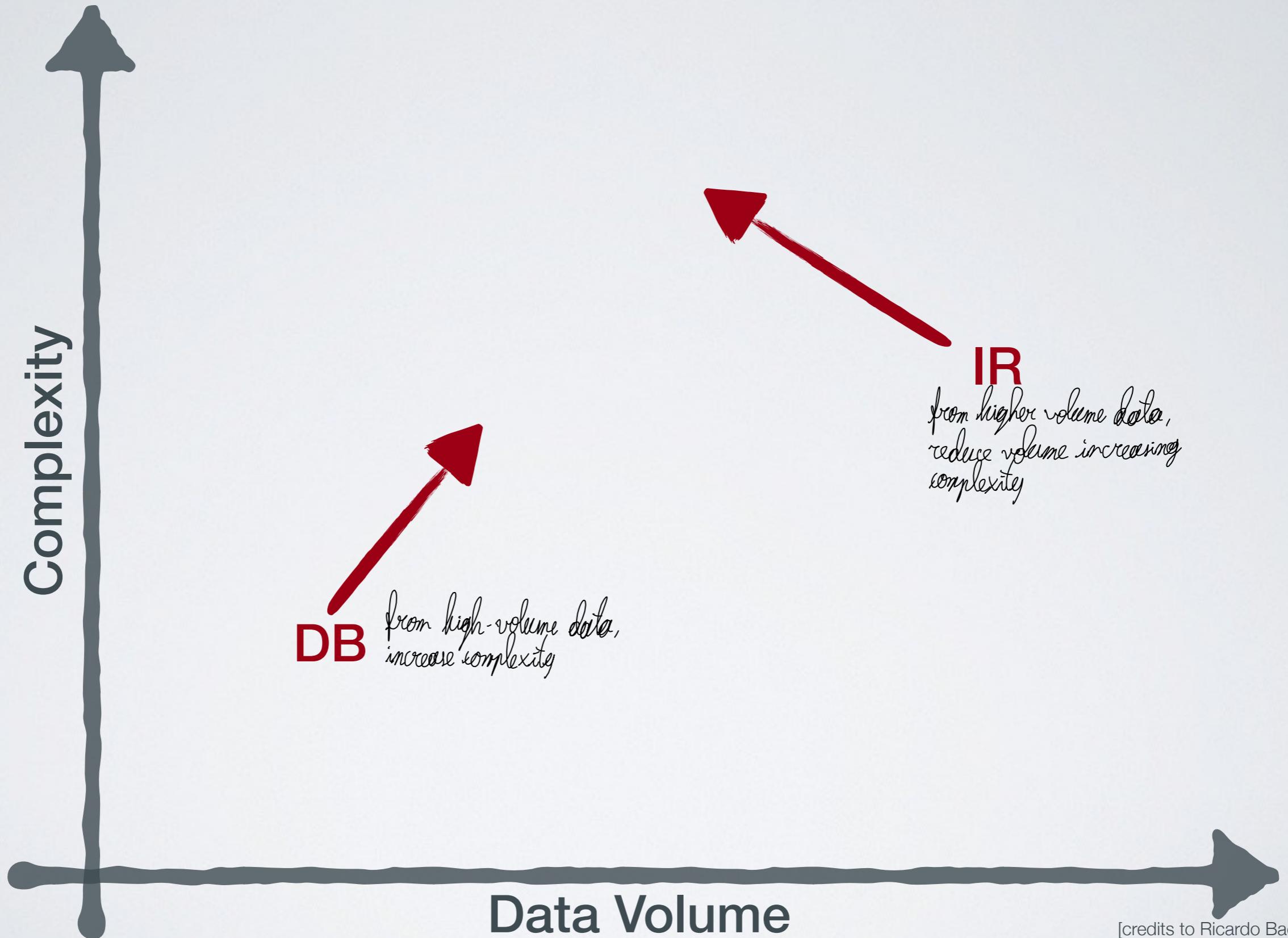
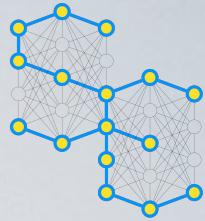
BIG DATA

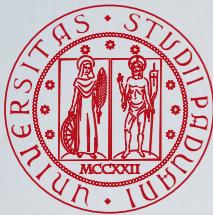


[credits to Charlie Clarke]



DB vs IR: Two Viewpoints on Data





Who Did Come First?



Probably we should teach information retrieval first and databases afterwards

DB: similar to engineering stuff, but far from reality
IR: process info in human-friendly way

Ricardo Baeza-Yates

Information Retrieval

P. BAXENDALE, Editor

A Relational Model of Data for Large Shared Data Banks

E. F. CODD

IBM Research Laboratory, San Jose, California

Future users of large data banks must be protected from having to know how the data is organized in the machine (the internal representation). A prompting service which supplies such information is not a satisfactory solution. Activities of users at terminals and most application programs should remain unaffected when the internal representation of data is changed and even when some aspects of the external representation are changed. Changes in data representation will often be needed as a result of changes in query, update, and report traffic and natural growth in the types of stored information.

Existing noninferential, formatted data systems provide users with tree-structured files or slightly more general network models of the data. In Section 1, inadequacies of these models are discussed. A model based on n -ary relations, a normal form for data base relations, and the concept of a universal data sublanguage are introduced. In Section 2, certain operations on relations (other than logical inference) are discussed and applied to the problems of redundancy and consistency in the user's model.

KEY WORDS AND PHRASES: data bank, data base, data structure, data organization, hierarchies of data, networks of data, relations, derivability, redundancy, consistency, composition, join, retrieval language, predicate calculus, security, data integrity

CR CATEGORIES: 3.70, 3.73, 3.75, 4.20, 4.22, 4.29

1. Relational Model and Normal Form

1.1. INTRODUCTION

This paper is concerned with the application of elementary relation theory to systems which provide shared access to large banks of formatted data. Except for a paper by Childs [1], the principal application of relations to data systems has been to deductive question-answering systems. Levein and Maron [2] provide numerous references to work in this area.

In contrast, the problems treated here are those of *data independence*—the independence of application programs and terminal activities from growth in data types and changes in data representation—and certain kinds of *data inconsistency* which are expected to become troublesome even in nondeductive systems.

Volume 13 / Number 6 / June, 1970

The relational view (or model) of data described in Section 1 appears to be superior in several respects to the graph or network model [3, 4] presently in vogue for noninferential systems. It provides a means of describing data with its natural structure only—that is, without superimposing any additional structure for machine representation purposes. Accordingly, it provides a basis for a high level data language which will yield maximal independence between programs on the one hand and machine representation and organization of data on the other.

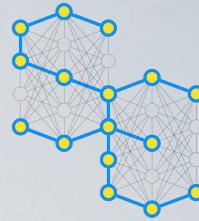
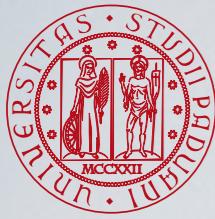
A further advantage of the relational view is that it forms a sound basis for treating derivability, redundancy, and consistency of relations—these are discussed in Section 2. The network model, on the other hand, has spawned a number of confusions, not the least of which is mistaking the derivation of connections for the derivation of relations (see remarks in Section 2 on the “connection trap”).

Finally, the relational view permits a clearer evaluation of the scope and logical limitations of present formatted data systems, and also the relative merits (from a logical standpoint) of competing representations of data within a single system. Examples of this clearer perspective are cited in various parts of this paper. Implementations of systems to support the relational model are not discussed.

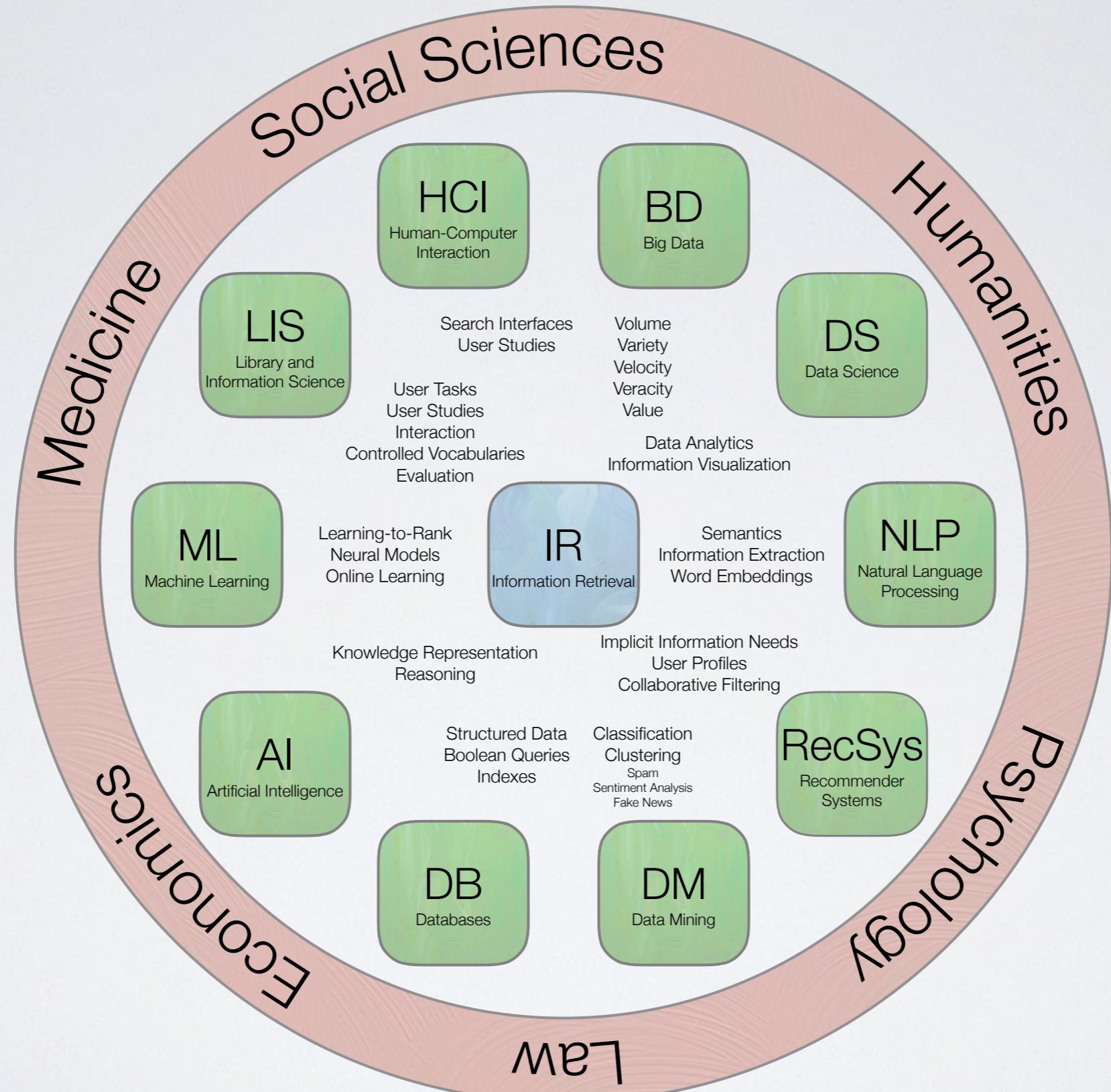
1.2. DATA DEPENDENCIES IN PRESENT SYSTEMS

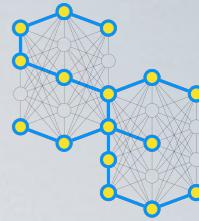
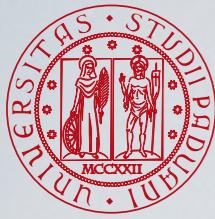
The provision of data description tables in recently developed information systems represents a major advance toward the goal of data independence [5, 6, 7]. Such tables facilitate changing certain characteristics of the data representation stored in a data bank. However, the variety of data representation characteristics which can be changed *without logically impairing some application programs* is still quite limited. Further, the model of data with which users interact is still cluttered with representational properties, particularly in regard to the representation of collections of data (as opposed to individual items). Three of the principal kinds of data dependencies which still need to be removed are: ordering dependence, indexing dependence, and access path dependence. In some systems these dependencies are not clearly separable from one another.

1.2.1. *Ordering Dependence.* Elements of data in a data bank may be stored in a variety of ways, some involving no concern for ordering, some permitting each element to participate in one ordering only, others permitting each element to participate in several orderings. Let us consider those existing systems which either require or permit data elements to be stored in at least one total ordering which is closely associated with the hardware-determined ordering of addresses. For example, the records of a file concerning parts might be stored in ascending order by part serial number. Such systems normally permit application programs to assume that the order of presentation of records from such a file is identical to (or is a subordering of) the



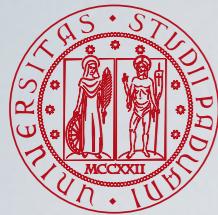
IR Constellation



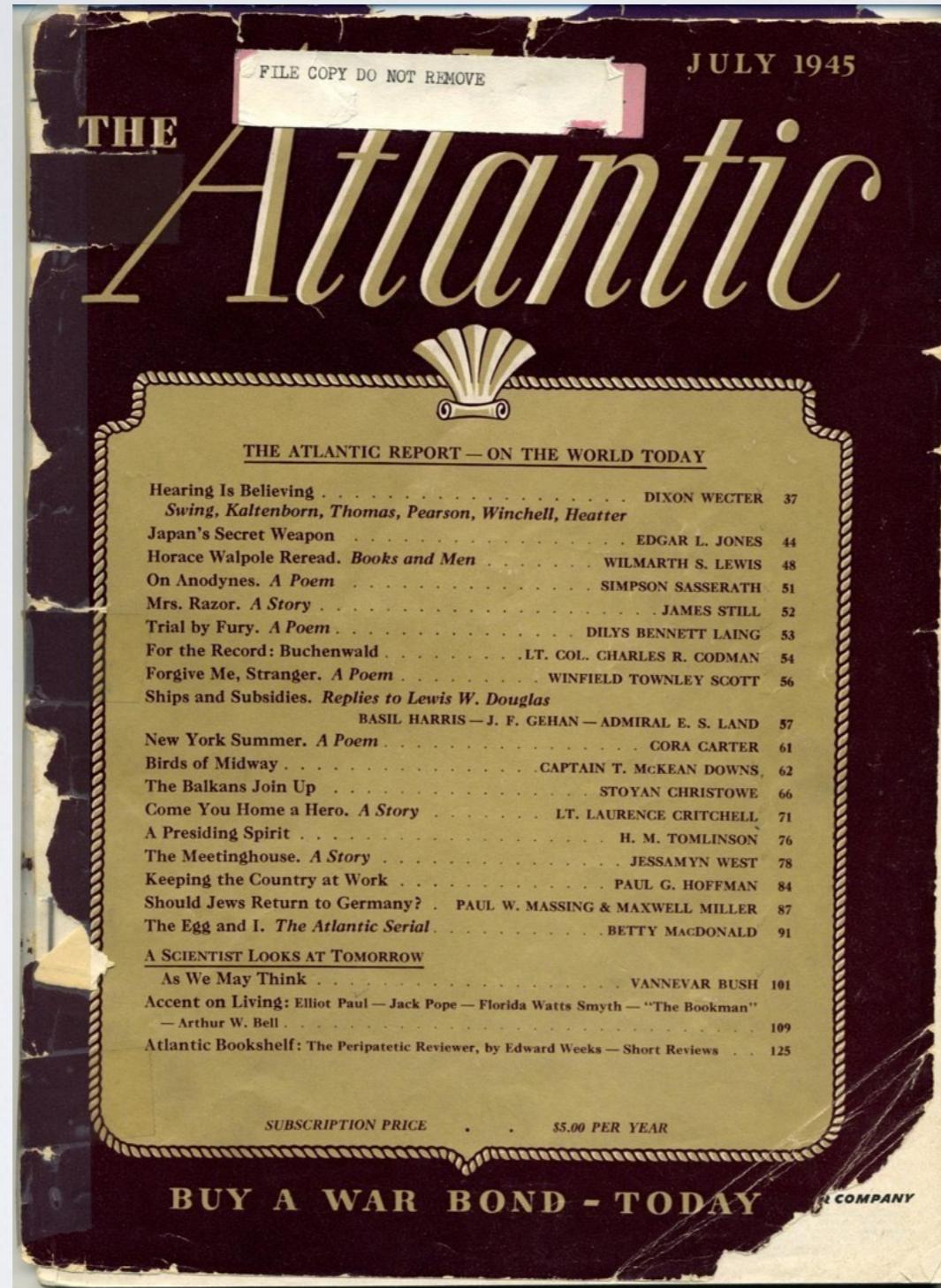
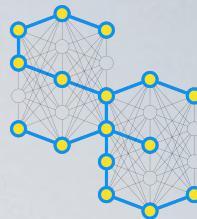


IR Constellation





Where Did All of This Come From?



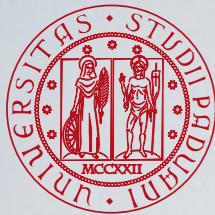
The MEMEX

Vannevar Bush (1890–1974)
Director of Office of Scientific Research and Development (OSRD)

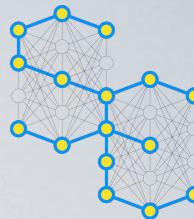


Bush, V. (1945). As We May Think. *The Atlantic Monthly*, 176:101–108.

<https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/>



Where Did All of This Come From?



The MEMEX

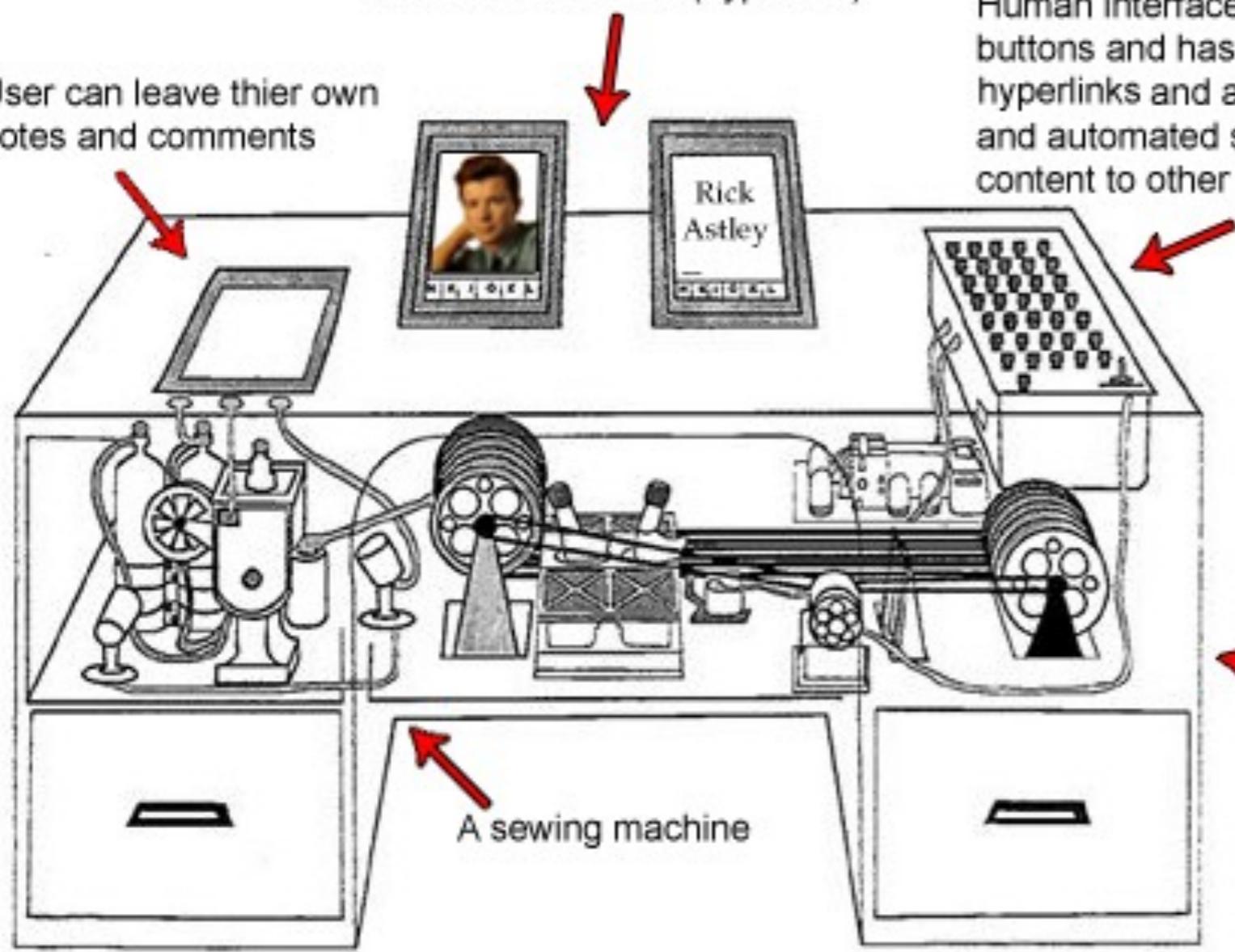


Dual projectors with the ability to cross reference text (hypertext).

User can leave their own notes and comments



Human interfaces with machine through buttons and has the ability to make new hyperlinks and associations (like bookmarks) and automated search and export user content to other memex machines



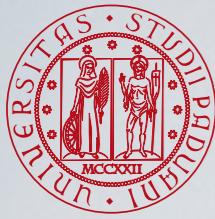
Var Bush (1890–1974)
Chair of Office of Scientific
Development (OSRD)



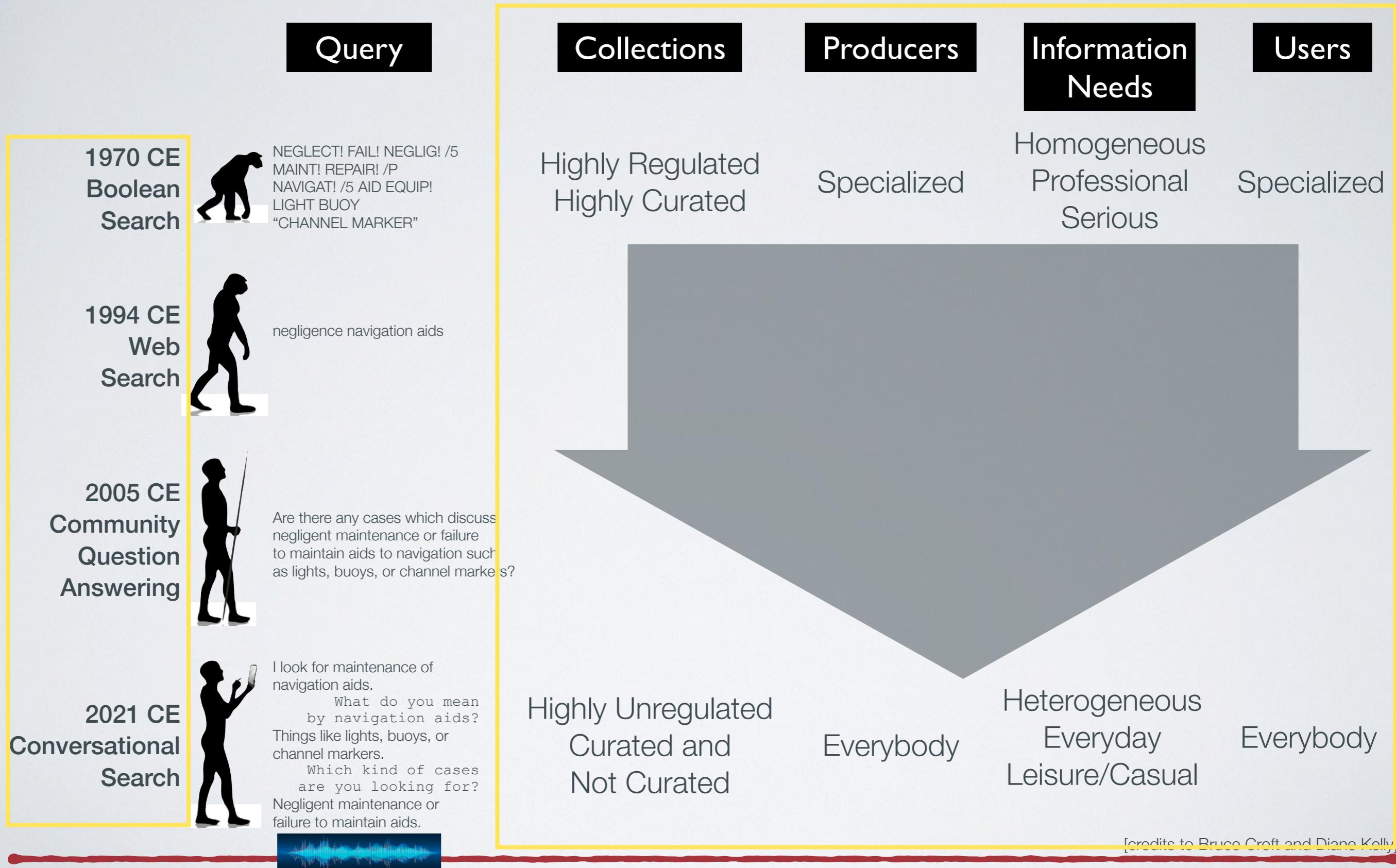
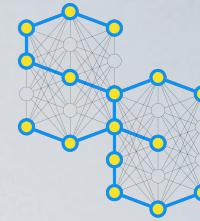
Sum of all human knowledge stored on micro-film in the form of text / image compositions

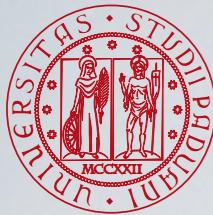
Bush, V. (1945). As We May Think. *The Atlantic Monthly*, 176:101–108.

<https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/>

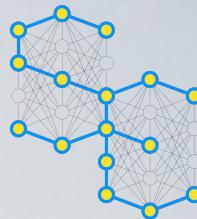


Where Are We Going?





Are We Already There?



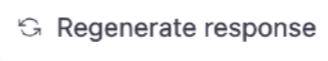
<https://openai.com/blog/chatgpt/>

 What is information retrieval?

 Information retrieval is the process of searching and retrieving information from a collection of data or information sources. The goal of information retrieval is to find relevant information that satisfies the user's information needs. This can be accomplished through a variety of methods, including keyword-based searching, semantic searching, content-based retrieval, and more.

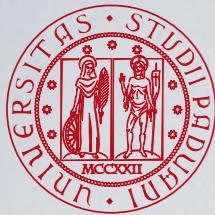
Information retrieval can be used to search for a wide range of information, including text, images, videos, and audio. The data sources for information retrieval can be digital or analog and can be stored in databases, websites, libraries, or other types of information systems.

Information retrieval is an important field of study that has a wide range of applications, including web search engines, digital libraries, and enterprise search systems. The goal of information retrieval is to make it easier for users to find the information they need, and to present the information in a way that is useful and accessible to the user.

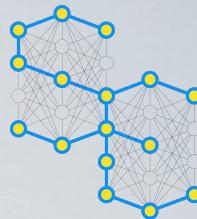
 Regenerate response

| 

ChatGPT Jan 30 Version. Free Research Preview. Our goal is to make AI systems more natural and safe to interact with. Your feedback will help us improve.



Are We Already There?



<https://blogs.microsoft.com/blog/2023/02/07/reinventing-search-with-a-new-ai-powered-microsoft-bing-and-edge-your-copilot-for-the-web/>

Microsoft Bing

what is information retrieval?

TUTTO NOTIZIE IMMAGINI VIDEO MAPPE CHAT ESPANDI

Informazioni sui risultati di 25.700.000 Data ▾

What is Information Retrieval?

A process of recovering specific pieces of information from the stored data. It is a system for organizing knowledge for subject retrieval. The phrase Information Storage and Retrieval (ISR) is of recent origin, coined by Calvin Moores. According to Moore, ISR is "Searching and Retrieval of information from a storage according to specification by subject." According to Lancaster, Activities involved in searching a body of literature in order to find items (documents) that deal with a particular subject area.

Information Retrieval Systems

- Information
- What is "information"?
- Retrieval
 - What do we mean by "retrieval"?
 - What are different types information needs?
- Systems
 - How do computer systems fit into the *human* information seeking process?

Information Retrieval

- Information retrieval (IR) is finding documents of an unstructured nature that satisfy an information need from a collection of documents (usually stored on computers).
- Information Retrieval
 - Deals with the representation, storage and access to information items
 - Modern Information Retrieval
 - General Objective: Minimize the overall cost of retrieving needed information

Recovery of information

information retrieval, **recovery of information**, especially in a database stored in a computer. Two main approaches are matching words in the query against the database index (keyword searching) and traversing the database using hypertext or hypermedia links.

Information retrieval | Definition, Methods, & Facts | Britannica

www.britannica.com/technology/information-retrieval

È stato utile? + -

Le persone hanno chiesto anche

What is a retrieval system?

Thus, retrieval systems provide users with online access to information that they may not be aware of, and they are not required to know or care about where the information is housed. Users can query all...

What is Information Retrieval (IR) i...

www.simplifying.com/tutorials/mach...

What is information retrieval in computing and Information Science?

Information retrieval (IR) in computing and information science is the process of obtaining information system resources that are relevant ...

Information retrieval - Wikipedia

en.wikipedia.org/w/index.php?title=Information_retr...

Information retrieval

Wikipedia

Information retrieval

L'information retrieval è l'insieme delle tecniche utilizzate per gestire la rappresentazione, la memorizzazione, l'organizzazione e l'accesso ad oggetti contenenti informazioni quali documenti, pagin...

Personne correlate

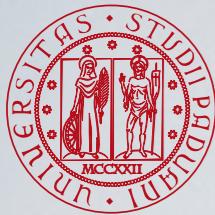
Susan Dumais C. J. van Rijsbergen Ricardo Baeza-Yates Karen Spärck... Rajeev Motwani

L'Information Retrieval

consiste nel **rendere** **accessibile** la conoscenza attualmente disponibile. Differentemente da quel che si possa pensare, questo processo non è una prerogativa dell'era digitale.

L'Information Retrieval (IR) è la **disciplina informatica** che analizza le tecniche di interrogazione e organizzazione dei dati in un archivio (database o base dati).

In altre parole, l'Information Retrieval è un **processo** con cui un utente può convertire il



Are We Already There?



<https://blogs.microsoft.com/blog/2023/02/07/reinventing-search-with-a-new-ai-powered-microsoft-bing-and-edge-your-copilot-for-the-web/>

Microsoft Bing

what is information retrieval?

TUTTO NOTIZIE IMMAGINI VIDEI

Informazioni sui risultati di 25.700.000

What is Information Retrieval?

A process of recovering specific pieces of information from the stored data. It is a system for organizing knowledge for subject retrieval. The phrase Information Storage and Retrieval (ISR) is of recent origin, coined by Celia Moore. According to Moore, ISR is "Searching and Retrieval of information from a storage according to specification by subject." According to Lancaster, Activities involved in searching a body of literature in order to find items (documents) that deal with a particular subject area.

Benvenuto sul nuovo Bing

Il motore di risposta basato sull'intelligenza artificiale

Chiedi qualsiasi cosa

Recovery of information

information retrieval, recovery of information from a computer. Two main approaches are matching words in the query against the database index (keyword searching) and traversing the database using hypertext or hypermedia links.

Information retrieval | Definition

www.britannica.com/technology/info

Le persone hanno chiesto anche

What is a retrieval system?

Thus, retrieval systems provide users with online access to information that they may not be aware of, and they are not required to know or care about where the information is housed. Users can query all...

What is Information Retrieval (IR) i...

www.superuser.com/tutorials/mack...

Microsoft Bing

what is information retrieval?

CERCA CHAT SCUOLA NOTIZIE IMMAGINI VIDEO MAPPE ESPANDI

Informazioni sui risultati di 27.200.000 Data Apri collegamenti in una nuova scheda

What is Information Retrieval?

A process of recovering specific pieces of information from the stored data. It is a system for organizing knowledge for subject retrieval. The phrase Information Storage and Retrieval (ISR) is of recent origin, coined by Celia Moore. According to Moore, ISR is "Searching and Retrieval of information from a storage according to specification by subject." According to Lancaster, Activities involved in searching a body of literature in order to find items (documents) that deal with a particular subject area.

Information Retrieval Systems

- Information
 - What is "Information"?
 - Retrieval
 - What do we mean by "retrieval"?
 - What are different types information needs?
 - Systems
 - How do computer systems fit into the **human** information seeking process?

Information Retrieval

- Information retrieval is the process of finding documents (or other information sources) of an unstructured nature that satisfy an information need (usually stored on a computer).
- Information Retrieval
 - Deals with the representation, access to information and organization of information.
 - Modern Information Retrieval
- General Objective: Minimizing the cost of locating needed information.

Recovery of information

Information retrieval, **Recovery of information**, especially in a database stored in a computer. Two main approaches are matching words in the query against the database index (keyword searching) and traversing the database using hypertext or hypermedia links.

Information retrieval | Definition, Methods, & Facts | Britannica

www.britannica.com/technology/information-retrieval

What are some examples of databases? What are hypertext and hypermedia? Ho bisogno di aiuto?

Digit a message... ➤

Commenti e suggerimenti ➡

Le persone hanno chiesto anche

What is a retrieval system?

Thus, retrieval systems provide users

What is information retrieval in computing and Information Science?

Information retrieval (IR) è la **disciplina informatica** che analizza le tecniche di interrogazione e organizzazione dei dati in un archivio (database o base dati).

In altre parole, l'Information Retrieval

W Wikipedia

Persons correlate ➤

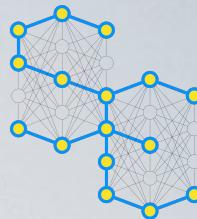
Susan Dumais C. J. van Rijsbergen Ricardo Baeza-Yates Karen Spärck Jones

L'Information Retrieval consiste nel rendere accessibile la conoscenza attualmente disponibile. Differentemente da quel che si possa pensare, questo processo non è una prerogativa dell'era digitale.

idf = log $\frac{\text{Numero di documenti}}{\text{Numero di documenti che contengono il termine}}$



Are We Already There?



<https://blogs.microsoft.com/blog/2023/02/07/reinventing-search-with-a-new-ai-powered-microsoft-bing-and-edge-your-copilot-for-the-web/>

Microsoft Bing CERCA CHAT Nicola 18 🔍

Condizioni per l'utilizzo | Informativa sulla privacy

what is information retrieval?

- ✓ Ricerca in corso per: **information retrieval**
- ✓ Ricerca in corso per: **information retrieval examples**
- ✓ Generazione delle risposte per l'utente in corso...

Information retrieval (IR) is the process of obtaining information system resources that are relevant to an information need from a collection of those resources ¹. Searches can be based on full-text or other content-based indexing ¹. Information retrieval is different from data retrieval, which is the process of accessing and retrieving structured data from a database ².

Some examples of information retrieval systems are:

- **Web search engines**, such as Google or Bing, which allow users to search for web pages, images, videos, etc. based on keywords or natural language queries ³ ⁴.
- **Library catalogues**, which allow users to search for books, journals, articles, etc. based on metadata such as title, author, subject, etc. ⁵.
- **Recommender systems**, which suggest items or products to users based on their preferences, ratings, or previous behavior ⁶.
- **Question answering systems**, which provide direct answers to natural language questions from a large corpus of documents ⁴.

Altre informazioni: 1. [bing.com](#) 2. [geeksforgeeks.org](#) 3. [digital-coach.com](#) +7 altro

Tell me more about web search engines. How do information retrieval systems work? What are the advantages of information retrieval?

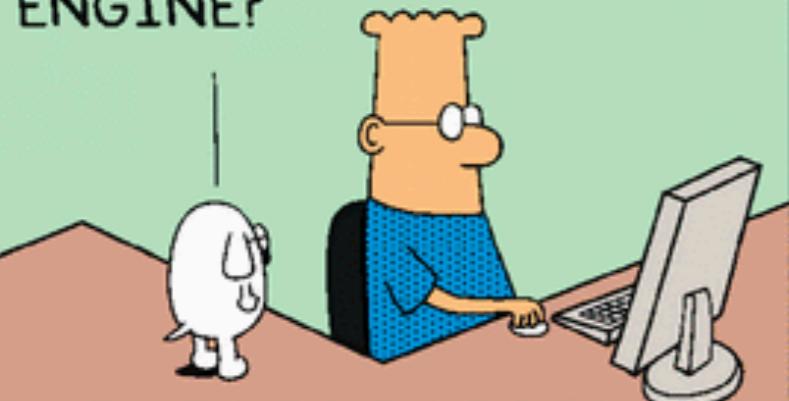
what is information retrieval?

30/2000

Feedback

questions?

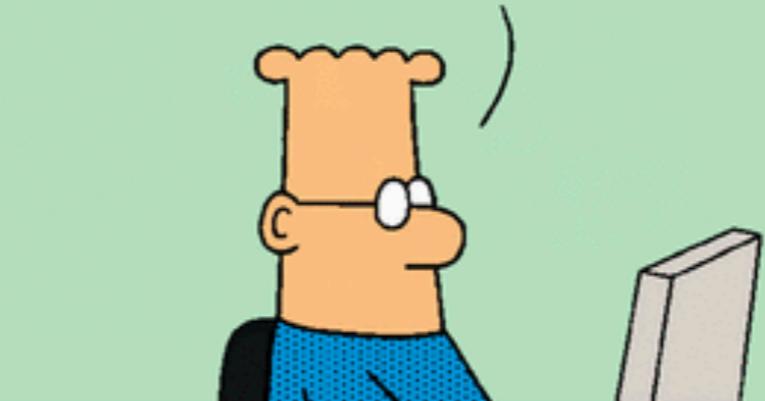
AREN'T YOU AFRAID
THAT GOOGLE WILL
TRY TO SQUASH YOU
FOR INVENTING A
BETTER SEARCH
ENGINE?



scottadams@aol.com

www.dilbert.com

THEIR COMPANY MOTTO
IS "DON'T BE EVIL." IT'S
NOT AS IF THEY HAVE A
DEATH RAY OR SOME-
THING.



5-16-06 © 2006 Scott Adams, Inc./Dist. by UFS, Inc.

GOOGLE HEADQUARTERS

NOW LOOK IN THE
BIG HOLE, ERIC.

