IL PENSIERO COMPUTAZIONALE

PERCORSO PER I DOCENTI DELLA SCUOLA SECONDARIA

Paolo Ferragina. *I motori di ricerca*

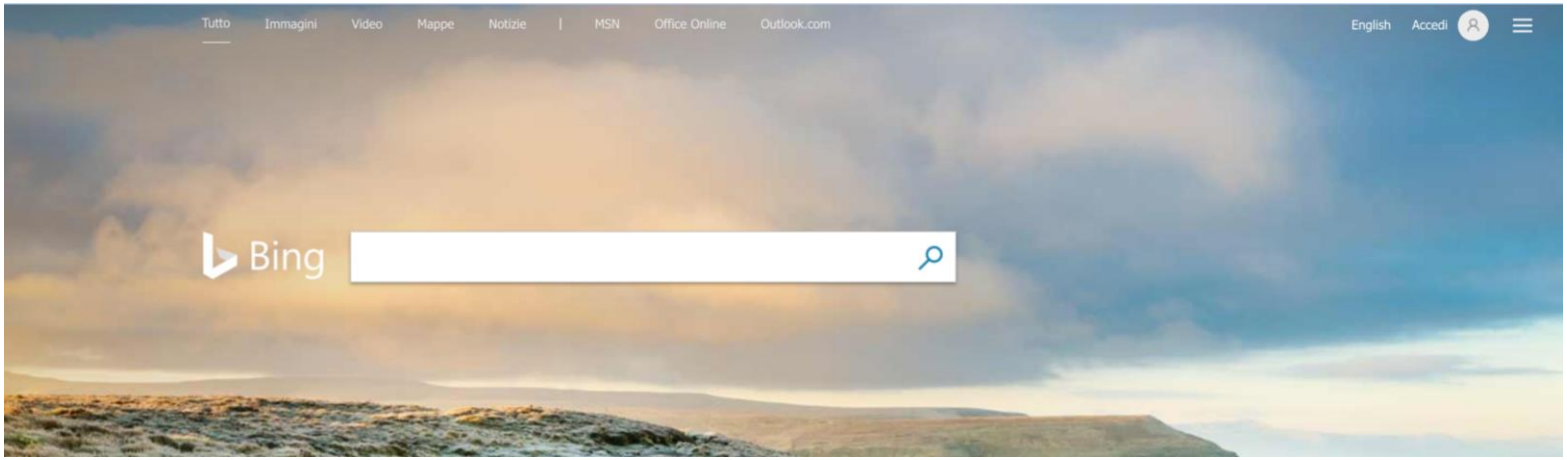
20 novembre 2019

UNIVERSITÀ DI PISA

Google

Italia

Cerca con Google | Mi sento fortunato



Alcune domande per iniziare

- Quando voi specificate una o più parole in una ricerca, Google/Bing «esplorano» il Web alla ricerca delle pagine che le contengono ?
- Dobbiamo specificare tutte le parole, anche articoli e preposizioni?
- La trigonometria è importante nella ricerca sul Web?
- Secondo quale «criterio di rilevanza» Google/Bing «ordinano» i risultati di una ricerca ?
- Come Google/Bing «trovano» la relazione tra Galileo e Copernico ?

Google galileo

Web Immagini Maps Video Notizie Altro Strumenti di ricerca

Circa 56.200.000 risultati (0,49 secondi)

Home | Galileo - Giornale di Scienza
www.**galileo**net.it/
News, magazine, recensioni e dossier monografici sui temi della scienza. Il primo magazine Online italiano su scienza e problemi globali. On line dal 1996.

Galileo Galilei - Wikipedia
it.wikipedia.org/wiki/**Galileo_Galilei**
Galileo Galilei (Pisa, 15 febbraio 1564 – Arcetri, 8 gennaio 1642) è stato un fisico, filosofo, astronomo e matematico italiano, considerato il padre della scienza ...
Processo a Galileo Galilei - Categoria:Galileo Galilei - Casa di Galileo Galilei

Galileo Galilei - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/**Galileo_Galilei** Traduci questa pagina
Galileo Galilei often known mononymously as **Galileo**, was an Italian physicist, mathematician, engineer, astronomer, and philosopher who played a major role ...

Aeroporto Galileo Galilei - Sito ufficiale - Aeroporto di Pisa ...
www.pisa-airport.com/
Aeroporto Internazionale **Galileo Galilei**. Include le informazioni, gli orari dei voli, le infrastrutture.
Hai visitato questa pagina molte volte. Ultima visita: 05/09/14

Liceo Classico Galilei Pisa
www.lc**galilei**.pisa.it/
Liceo Classico **Galileo Galilei** - Pisa. Il futuro ha un cuore antico. Home; Registro

Galileo Galilei
Fisico

Libri

Sidereus Nuncius
1610

Dialogo sopra i due massimi s...
1632

Discorsi e dimostraz... matemati...
1638

Il Saggiatore
1623

Lettera a Madama Cristina d...
1636

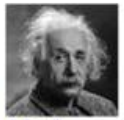
Ricerche correlate
Visualizza altri 15 elementi

Niccolò Copernico

Isaac Newton

Giovanni Keplero

Aristotele


Albert Einstein

Feedback

ALTAVISTA
Technology

View Multimedia From Our Vantage Point

Click here for advertising information - reach millions every month!

Search the Web and Display the Results in Standard Form

Submit

Search with Digital's Alta Vista [Advanced Search] [Add URL]

Contests
Make Me Laugh...

Creative Web
Create a Site...

Download free demo versions of AltaVista Technology software

ALTAVISTA

[Creative][Search][Humor][Email]

Solo contenuto testuale delle pagine

Search the web using Google!

10 results | Google Search | I'm feeling lucky

Index contains ~25 million pages (soon to be much bigger)

About Google!

Stanford Search Linux Search

Get Google! updates monthly!

your e-mail | Subscribe

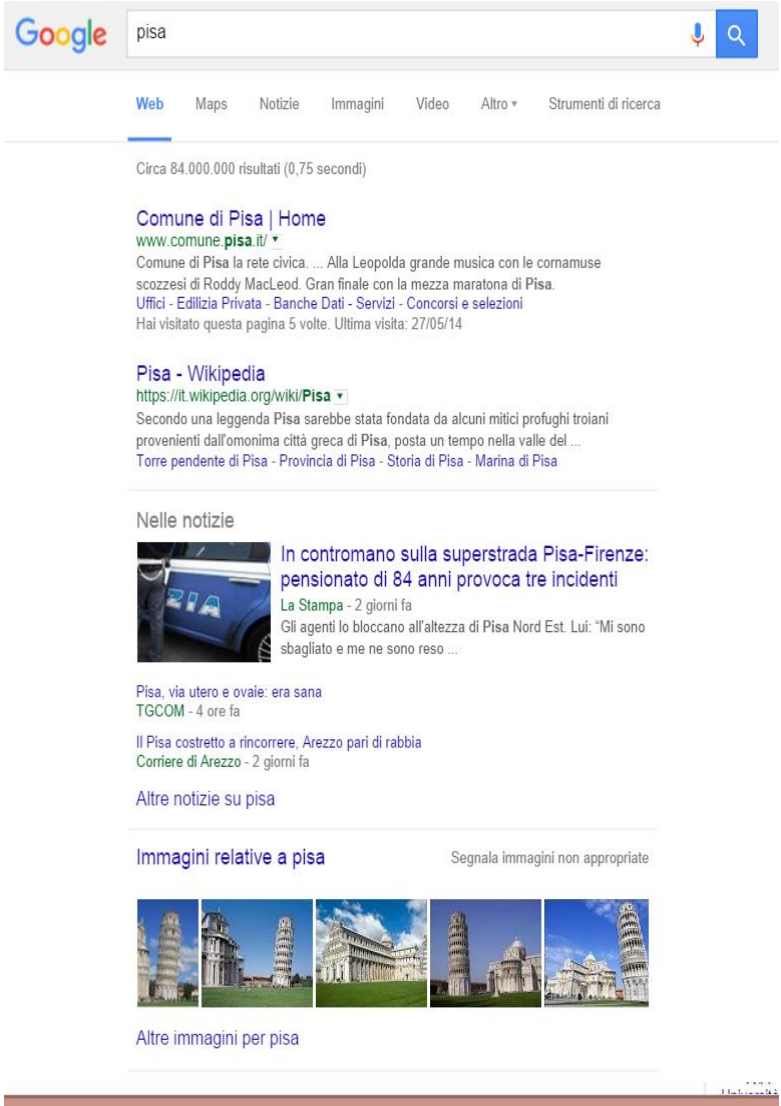
Copyright ©1997-8 Stanford Universit

- Testi àncora
- Rete degli hyperlink

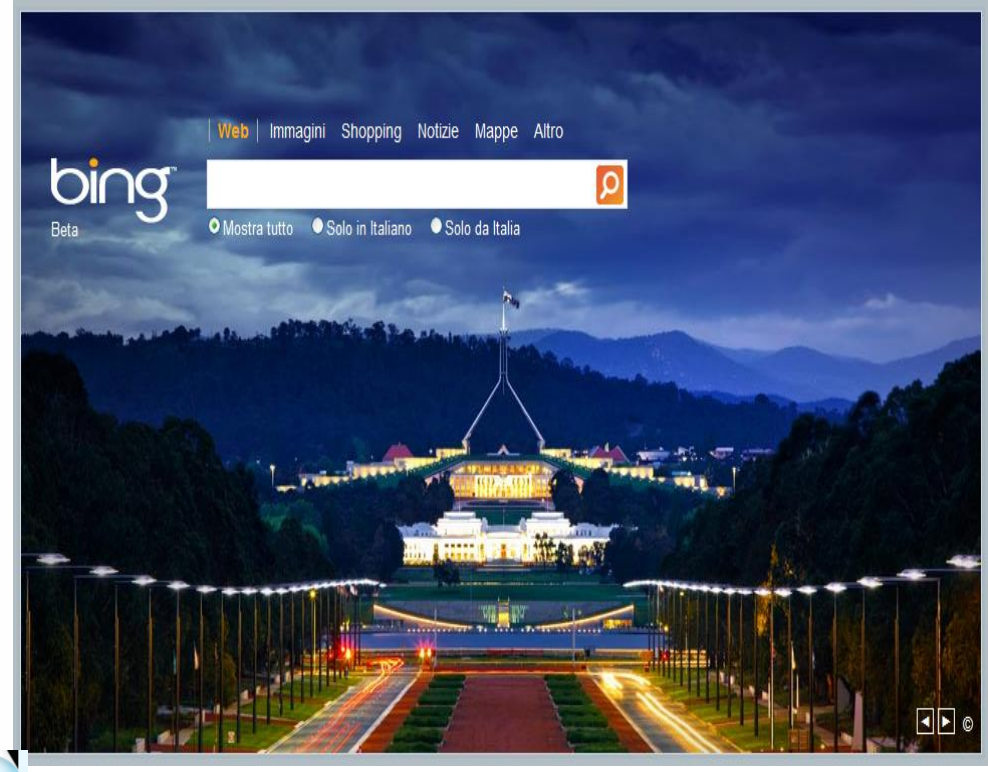
Terza generazione (≈ 2005)

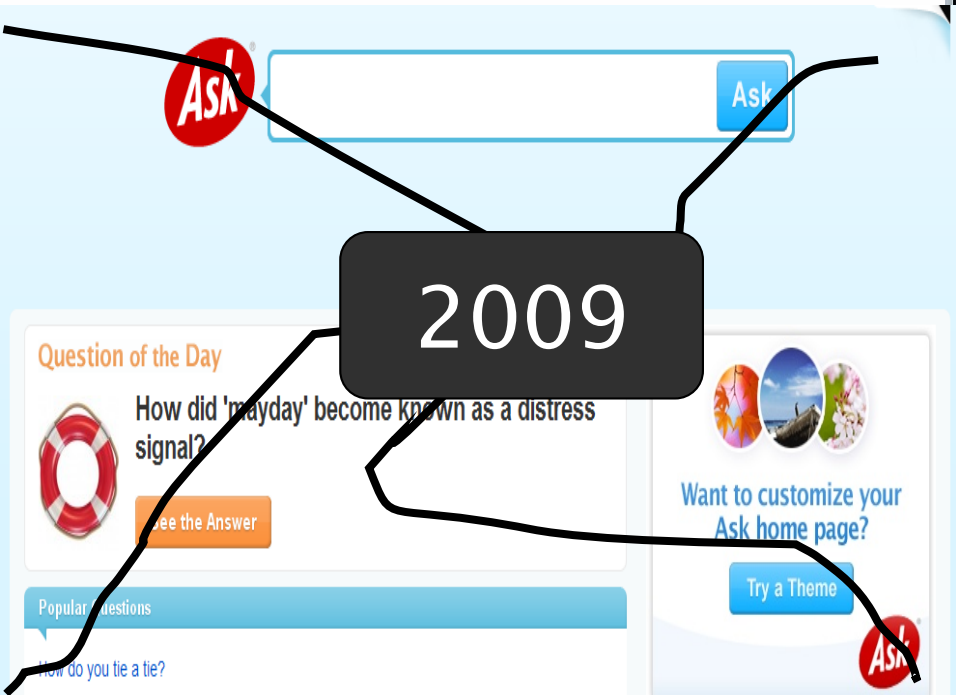
- Più sorgenti:
  - Notizie
  - Immagini
  - Mappe
  - Wikipedia
  - Dizionari
  - …

Rispondere ai bisogni informativi degli utenti



Google
Italia
Web Immagini News Maps Novità! Gruppi Desktop altro »
Ricerca avanzata
Preferenze
Strumenti per le lingue
Cerca con Google
Mi sento fortunato
Cerca: il Web pagine in Italiano pagine provenienti da: Italia
Novità! Usa Gmail e le altre applicazioni di Google Apps per la tua organizzazione.
Pubblicità - Soluzioni Aziendali - Tutto su Google - Google.com in English
Scegli Google come pagina iniziale!
©2007 Google

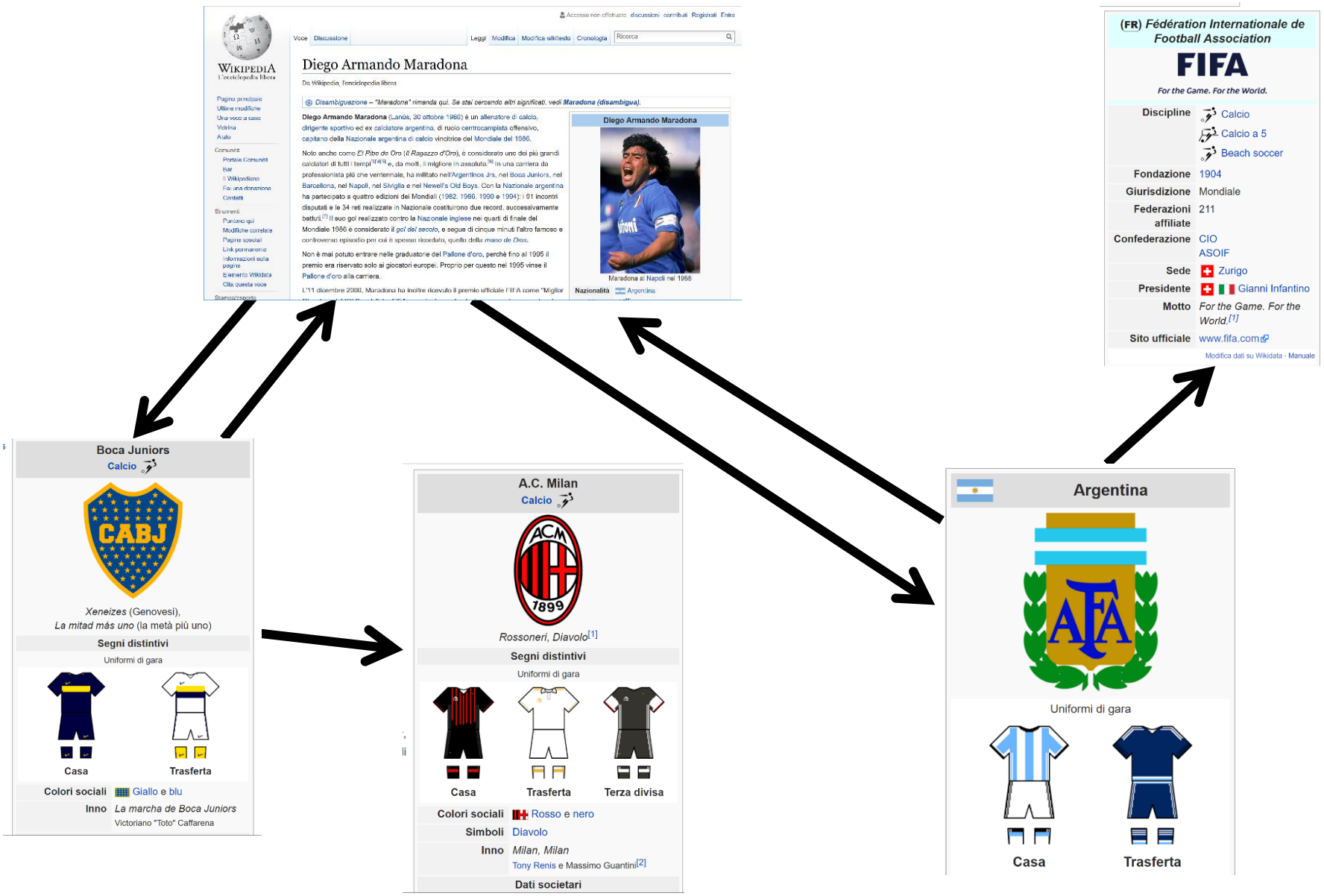

Web Immagini Shopping Notizie Mappe Altro
bing
Beta
Mostra tutto
Solo in Italiano
Solo da Italia


Ask
2009
Question of the Day
How did 'mayday' become known as a distress signal?
See the Answer
Popular Questions
How do you tie a tie?
Want to customize your Ask home page?
Try a Theme

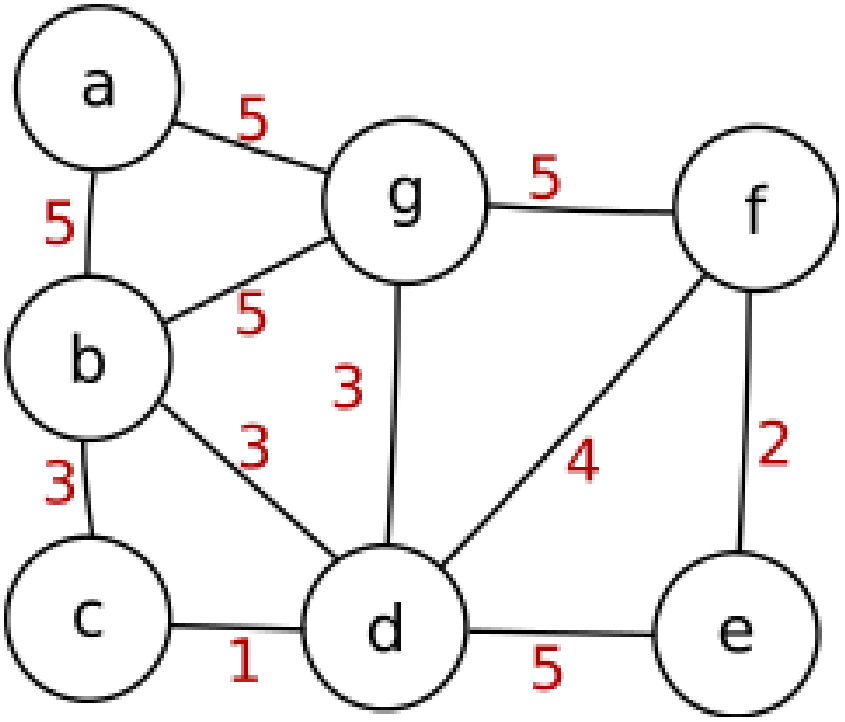
YAHOO!
Search:
Web Search
My Yahoo!
My Mail
Answers
Autos
Finance
Games
Groups
HotJobs
Maps
Mobile Web
Movies
Music
Personals
Real Estate
Shopping
Sports
Tech
Travel
TV
Yellow Pages
Featured
Sports
Video
Hi, Paolo
Sign Out
Mail
Messenger
Radio
Local
Horoscopes
2009-12
Flights
Cars
Deals
Vacations
Hotels
In the News
World
Small Business
Marketplace
Be a Better Movie Date
Watch movie trailers and clips
All Stars: Most Viewed Cars

Google | Inside Search
Home
Tips & Tricks
Features
Search Stories
Playground
Blog
Help
Search plus your world
Flight Search
Related Search Previews
Search by Image
Voice Search
Google Instant
Maggio 2012
The Knowledge Graph
Learn more about one of the key breakthroughs behind the future of search.
See it in action
Discover answers to questions you never thought to ask, powered by the Knowledge Graph.

Il Grafo di Wikipedia



Grafi: una struttura dati potente



Google

Quanto è alta la torre di pisa

Tutti | Notizie | Shopping | Immagini | Maps | Altro | Impostazioni | Strumenti

Circa 268.000 risultati (1,21 secondi)

Torre pendente di Pisa / Altezza

56 m

Torre Eiffel
300 m

Big Ben
96 m

Empire State Buil...
381 m

Feedback

amazon.it prime

Tutte le categorie

Invia a Paolo
Pisa 56123

Scegli per **categoria**

Amazon.it di Paolo | Offerte | Buoni Regalo | Vendere | Aiuto

Consigliati in base ai tuoi interessi

Simili a prodotti che hai già visto Visualizza altro

Altri prodotti da tenere presente Visualizza altro

facebook

E-mail o telefono: ilpensierocomputazionale

Password: ••••••••••

Accedi

Non ricordi più come accedere all'account?

Facebook ti aiuta a connetterti e rimanere in contatto con le persone della tua vita.

Iscriviti

È gratis e lo sarà sempre.

Nome | Cognome

Numero di cellulare o indirizzo e-mail

Nuova password

Data di nascita

17 | nov | 1993

Perché devo fornire la mia data di nascita?

Donna | Uomo

Cliccando su Iscriviti, accetti le nostre Condizioni. Scopri in che modo raccogliamo, usiamo e condividiamo i tuoi dati nella nostra Normativa sui dati e in che modo usiamo cookie e tecnologie simili nella nostra Normativa sui cookie. Potresti ricevere notifiche tramite SMS da noi, ma puoi disattivarle in qualsiasi momento.

Iscriviti

tripadvisor UNITED KINGDOM know better book better go better

Review JOIN LOG IN €

Hotels Flights Holiday Rentals Restaurants Things to do Best of 2016 More

Find: Hotels, Restaurants, Things to Do

Near: Enter a destination

Search

Find and book your ideal trip at the lowest prices

HOTELS HOLIDAY RENTALS THINGS TO DO RESTAURANTS FLIGHTS

Enter destination or hotel name

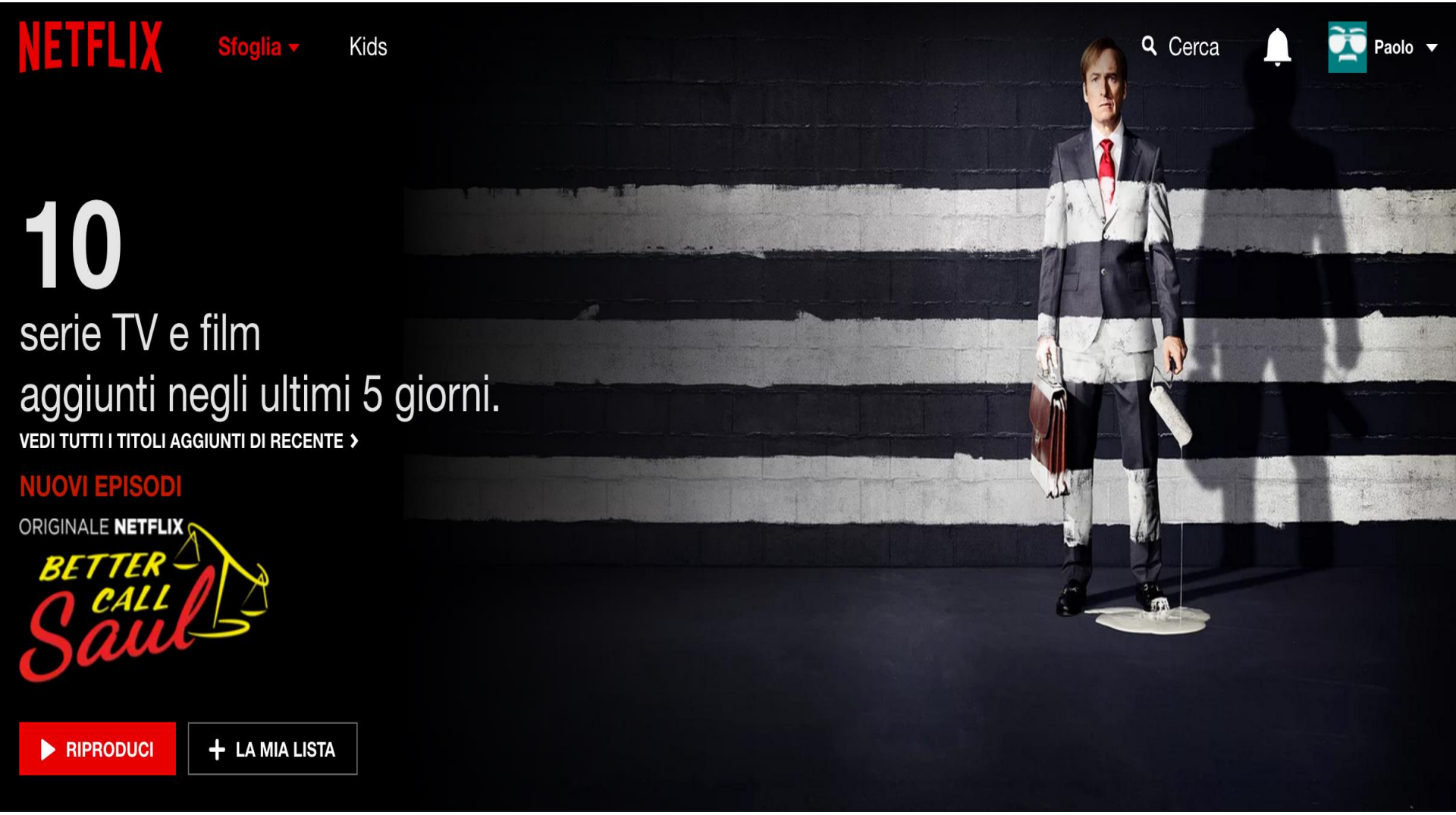
dd/mm/yyyy

dd/mm/yyyy

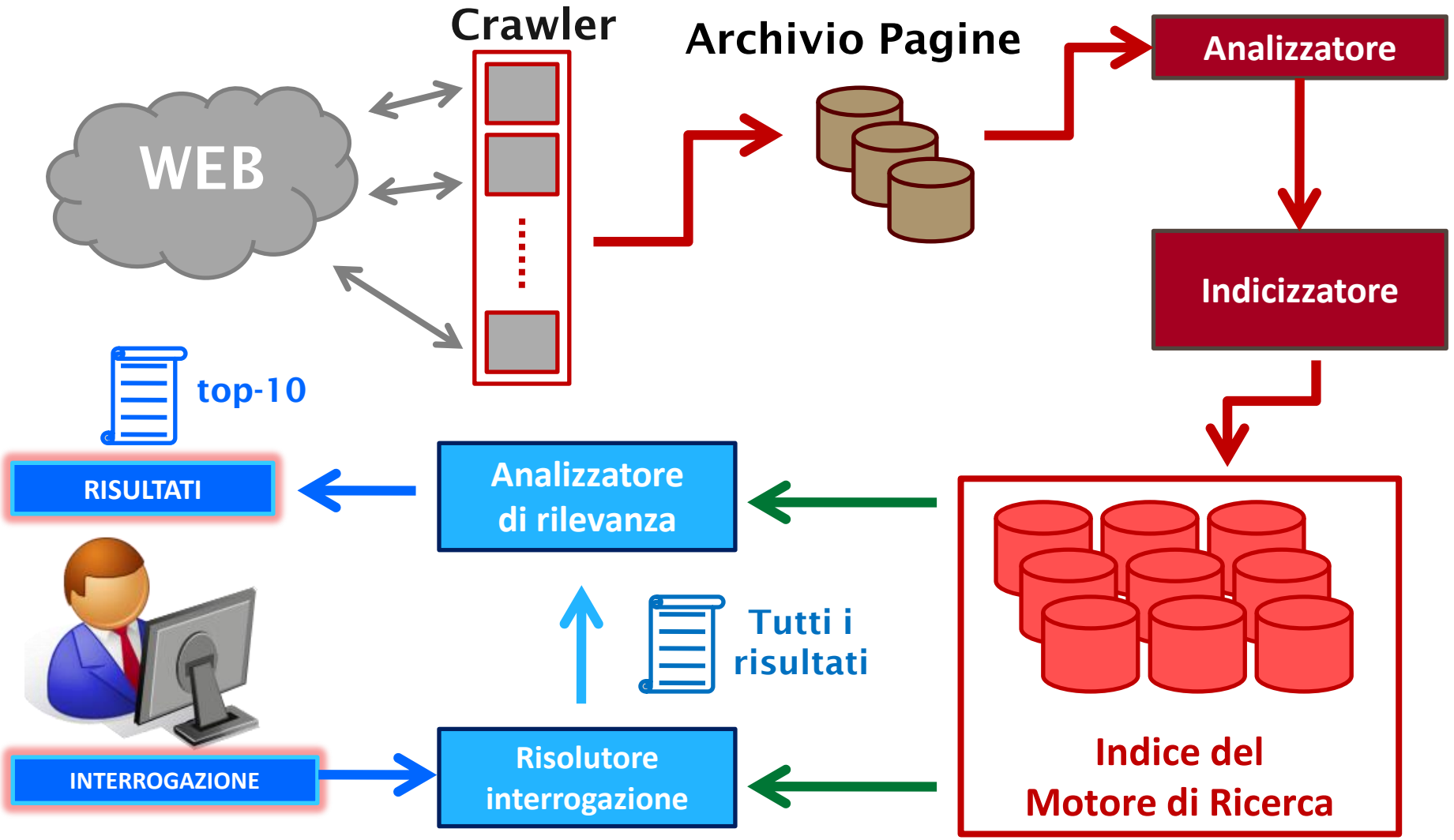
"Heaven on earth"

Review by 3Guapas

See all 41,234 reviews of Providenciales


NETFLIX
Sfoglia
Kids
Cerca
Paolo
10
serie TV e film
aggiunti negli ultimi 5 giorni.
VEDI TUTTI I TITOLI AGGIUNTI DI RECENTE ›
NUOVI EPISODI
ORIGINALE NETFLIX
BETTER CALL Saul
RIPRODUCI
+ LA MIA LISTA

La struttura di un motore di ricerca

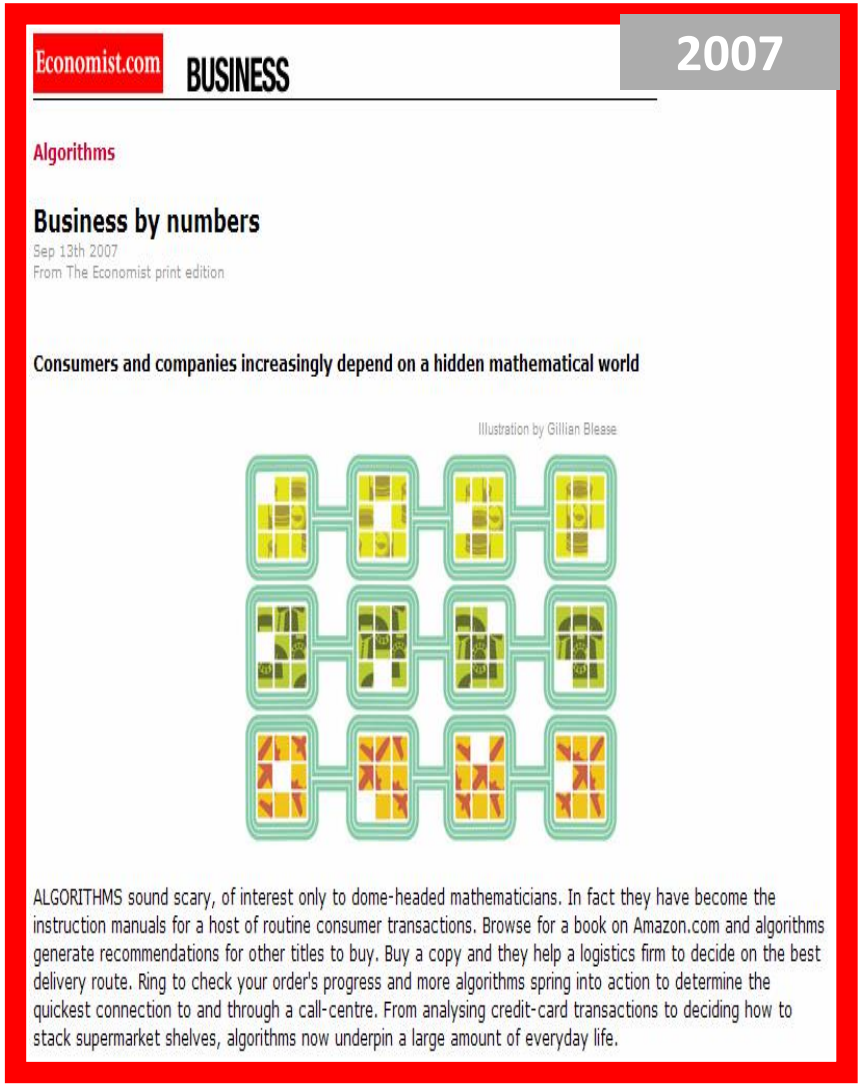


Council Bluffs, Iowa

Fiori di campo che fioriscono intorno alle torri di raffreddamento del nostro data center nell'Iowa.

Condividi +1 0

Algoritmi efficienti e «intelligenti»



Economist.com BUSINESS

2007

Algorithms

Business by numbers

Sep 13th 2007
From The Economist print edition

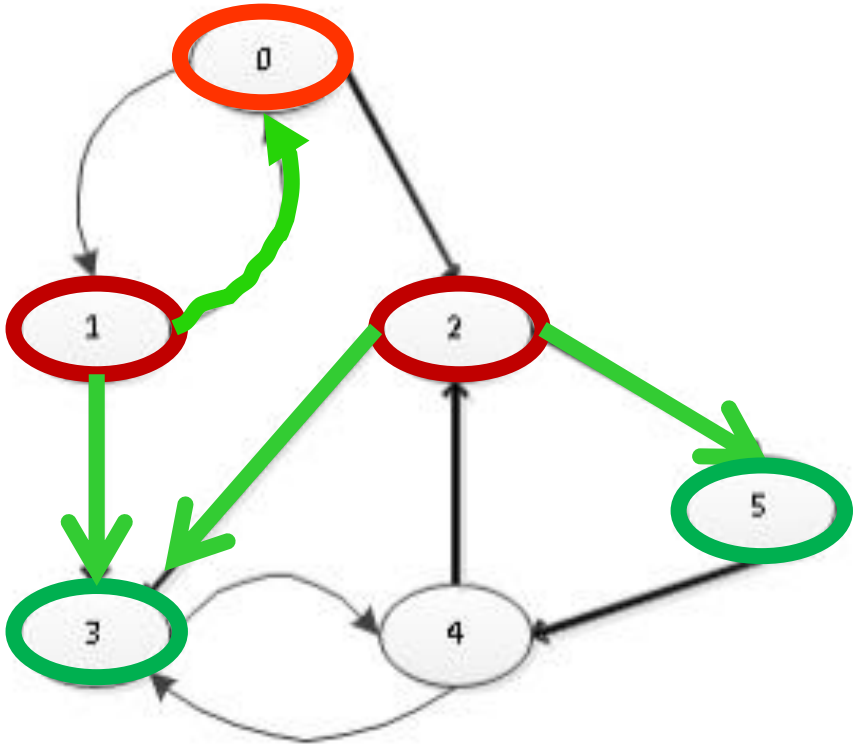
Consumers and companies increasingly depend on a hidden mathematical world

Illustration by Gillian Blease

ALGORITHMS sound scary, of interest only to dome-headed mathematicians. In fact they have become the instruction manuals for a host of routine consumer transactions. Browse for a book on Amazon.com and algorithms generate recommendations for other titles to buy. Buy a copy and they help a logistics firm to decide on the best delivery route. Ring to check your order's progress and more algorithms spring into action to determine the quickest connection to and through a call-centre. From analysing credit-card transactions to deciding how to stack supermarket shelves, algorithms now underpin a large amount of everyday life.

"This is rocket science but you don't have to be a rocket scientist to use it"

Un esempio di crawler



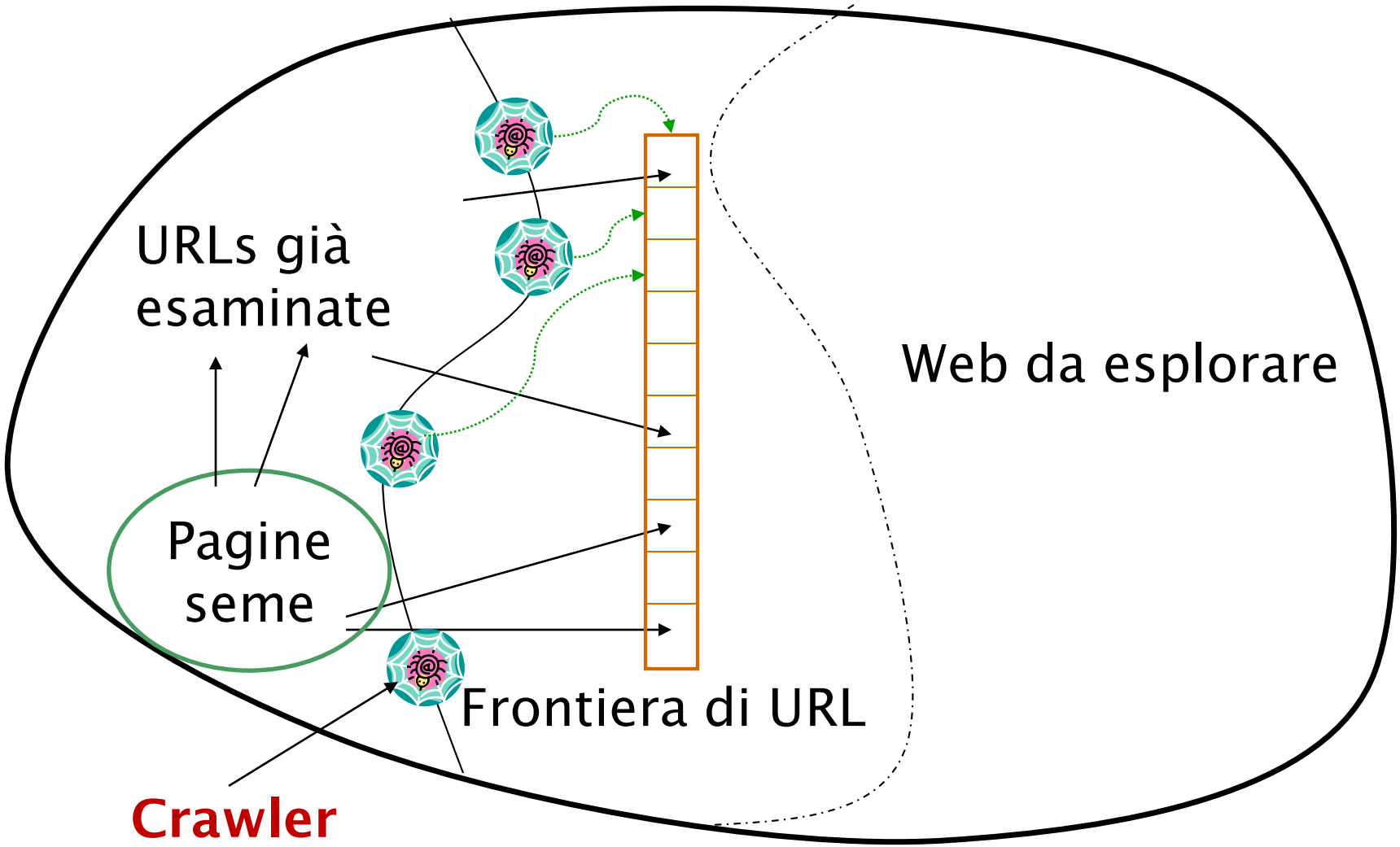
Il **Crawler**

1. **S = «pagine seme»**
2. Finché S non è vuoto
   a) Sia P una pagina di S
   b) Scarica P dal web
   c) Analizza il contenuto di P
   d) Inserisce i link di P in S,

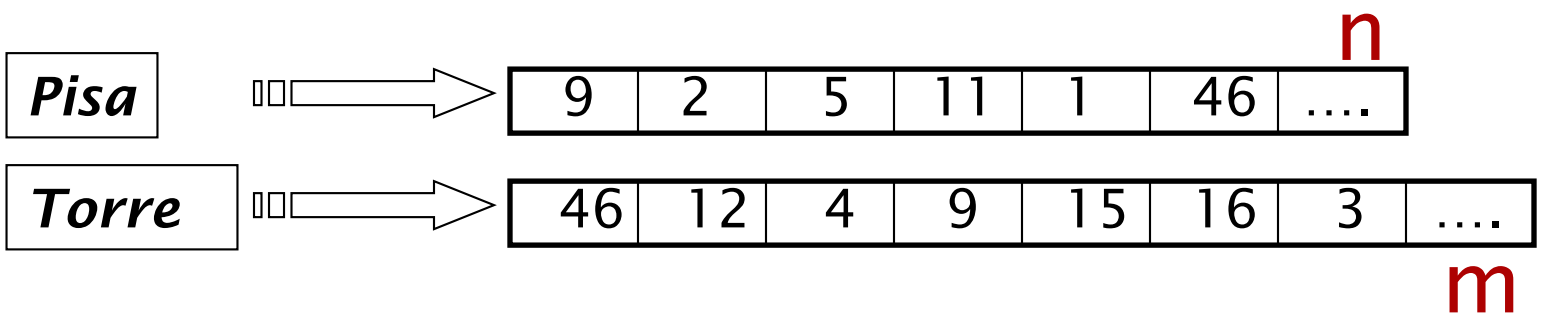
      ma solo se **…??…**

Un esempio realistico

- Quanto esplorare nel Web ?
- Quanto esplorare di un singolo sito ?
- Quanto spesso visitare una pagina ?



Il cuore dei motori di ricerca

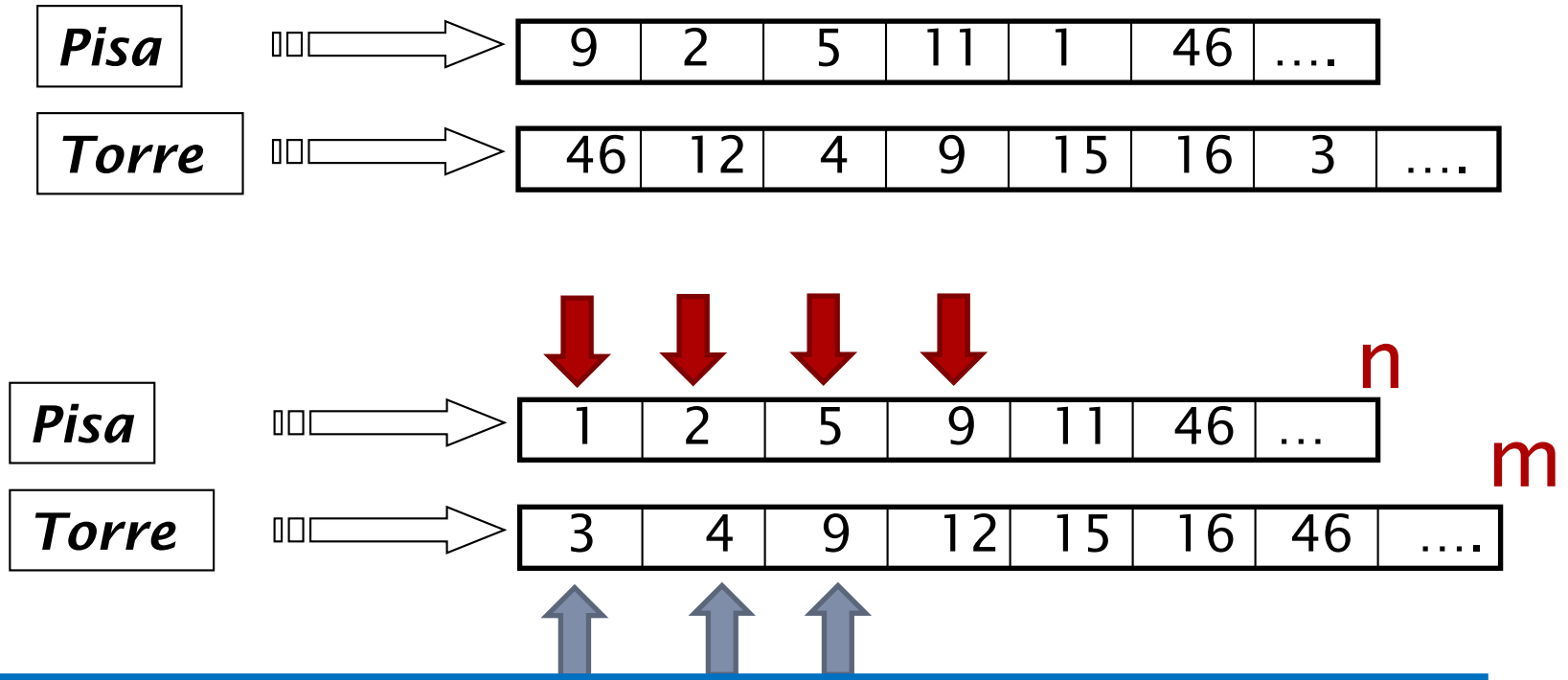


Sono eseguiti n*m confronti (cfr)

Se n,m ≈ 10^6 , si eseguono n * m ≈ 10^{12} cfr

Se CPU esegue 10^9 cfr/sec, intersezione ≈10^3 sec

Una soluzione più efficiente...



Se n,m ≈ 10^6 , si eseguono n + m ≈ 10^6 cfr

Se CPU esegue 10^9 cfr/sec, intersezione ≈ 10^{-3} sec

Esercizio: l'interrogazione per frase !

Q = "pensiero computazionale"

- ***pensiero:***

<2: 1, 17, 74, 222, 551***>***; ***<4***: 8, 16, 190, 429, 433***>***; ***<7***: 13, 23, 191***>***; ...

- ***computazionale:***

<1: 17, 19***>***; ***<4***: 11, 14, 17, 191, 291, 430, 434***>***; ***<5***: 14, 19, 101***>***; ...

Esercizio: l'interrogazione per frase !

Q = "pensiero computazionale"

- ***pensiero:***

<2: 1, 17, 74, 222, 551**>**; **<4**: 8, **16,** 190, **429,** 433**>**; **<7**: 13,23,191**>**; …

- ***computazionale:***

<1: 17, 19**>**; **<4**: 11, 14, **17,** 191, 291, **430,** 435**>**; **<5**: 14,19,101**>**; …

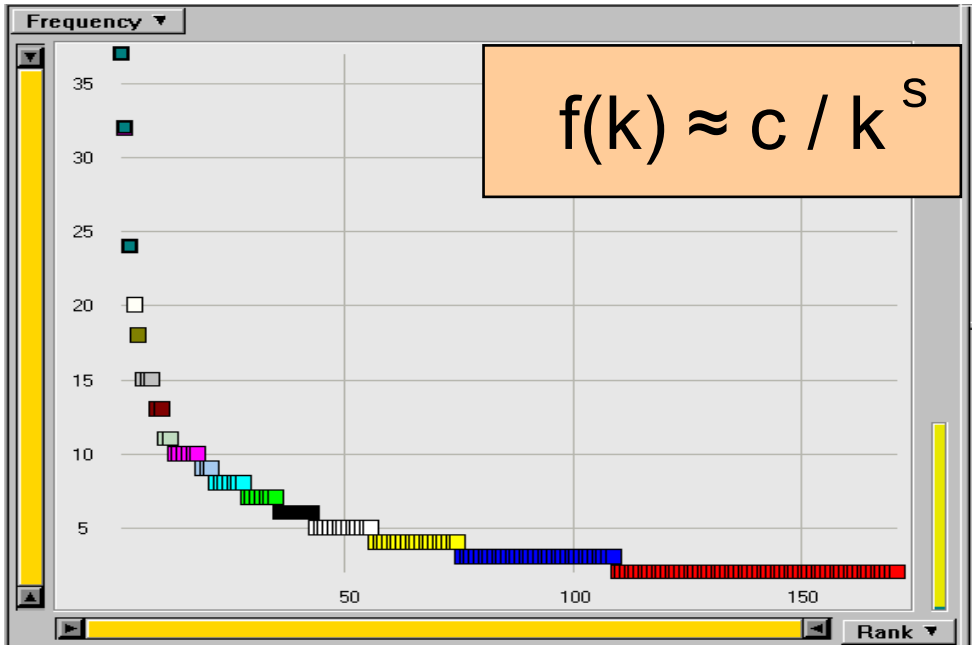
Quali sono i top-10 risultati ?

Possiamo usare la frequenza delle parole?

rango	forma	frequenza	rango	forma	frequenza
1	e	1752	16	ma	290
2	di	1338	17	i	283
3	che	1019	18	come	234
4	a	932	19	da	233
5	il	925	20	io	225
6	la	711	21	mi	219
7	un	708	22	le	211
8	non	507	23	più	210
9	per	481	24	l'	206
10	in	453	25	disse	202
11	Pinocchio	415	26	lo	199
12	si	393	27	burattino	195
13	gli	364	28	se	189
14	una	360	29	con	188
15	è	296	30	era	185

Le 30 parole più frequenti in *Pinocchio*

La legge di Zipf



Parole ordinate per freq decrescente
Posizione k → Frequenza f(k)

- I testi sono vincolati da regole (sintassi, grammatica, estetica, ecc) e quindi si *ribellano* alla legge gaussiana

Una famosa misura di rilevanza

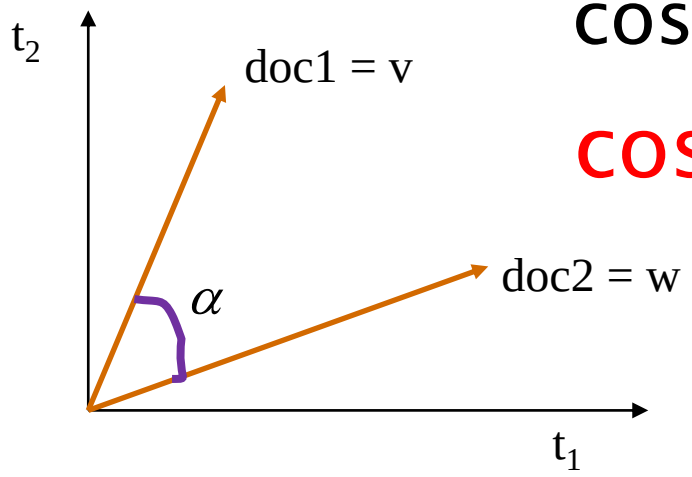
$$rilevanza(t,d) = freq(t,d) \times raritàColl(t)$$

- Rilevanza(e , «Pinocchio») → (freq = 1752) * (rarità ≅ 0)
- Rilevanza(burattino , «Pinocchio») → (freq = 195) * (rarità ≅ 1)
- Ogni documento **d** (p.e. «Pinocchio») è rappresentato quindi con un vettore di pesi di rilevanza, uno per ogni possibile termine **t**

Pinocchio = [0.6, 0.8, 0, 0.1, 0.3, ….]

Siamo in uno Spazio Vettoriale
con milioni di dimensioni, tante quanti sono i termini del dizionario

Un po' di trigonometria



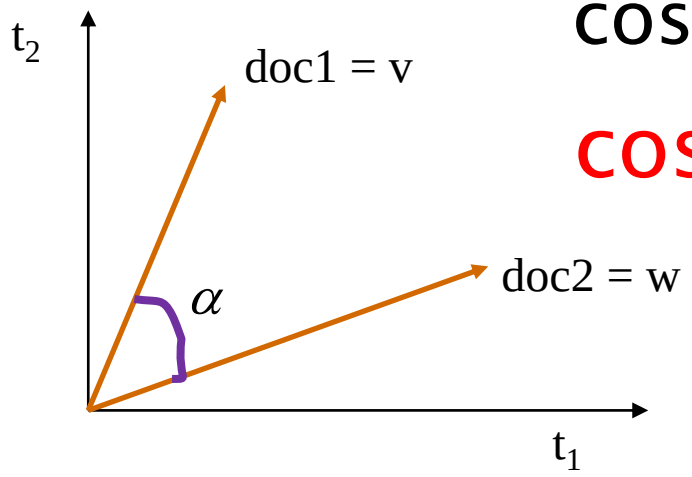
$\cos(\alpha) \in [0$ *(dissimile)*, 1 *(simile)*$]$

$\cos(\alpha) = v \bullet w \,/\, \|v\| * \|w\|$

	Doc1 = v	Doc2 = w
term 1	2	4
term 2	3	1

$\cos(\alpha) = 2*4 + 0*0 + 3*1 \,/\, \text{sqrt}\{ 2^2 + 3^2 \} * \text{sqrt}\{ 4^2 + 1^2 \} \cong 0{,}75 \rightarrow 40°$

Un po' di trigonometria



$\cos(\alpha) \in [0$ *(dissimile)*, 1 *(simile)*$]$

$\cos(\alpha) = v \bullet w \,/\, \|v\| * \|w\|$

SPAM !!!

Problema computazionale

#pagine .it ≈ qualche mld → 10^9

termini ≈ qualche mln → 10^6

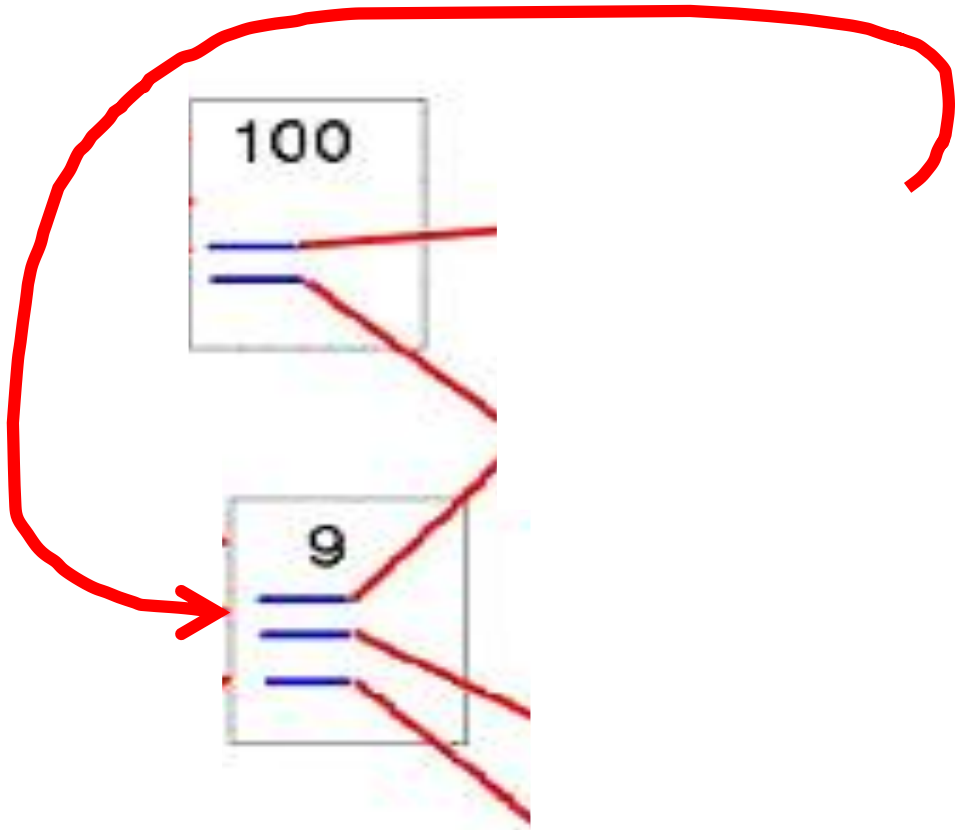
#ops ≈ 10^{15}

10^9 ops/sec ≈ 10^6 sec ≈ 1 settimana

Stavano svolgendo un PhD in Computer Science a Stanford

Il (classico) PageRank

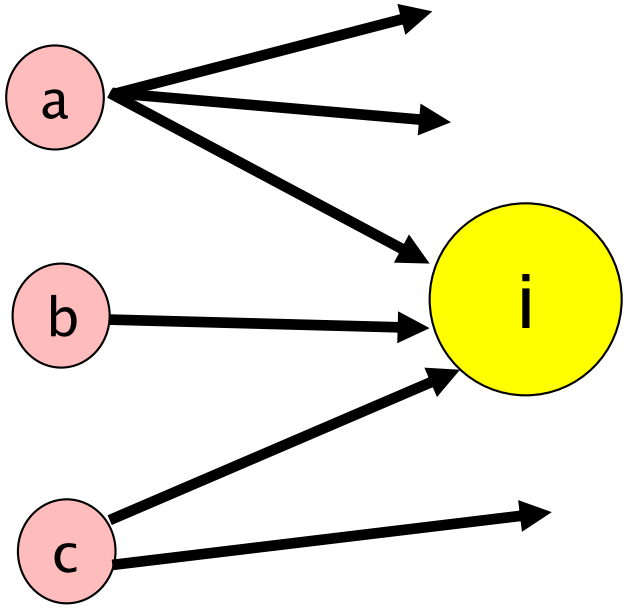
Se lasciamo propagare i valori, questi si "stabilizzano" ?



- Sistema lineare con miliardi di vincoli e variabili, efficienza?
- Varie interpretazioni: Cammini casuali, Catene di Markov

PageRank, come Sistema Lineare

$$r(i) = \alpha \cdot \sum_{j \in B(i)} \frac{r(j)}{\#out(j)} + (1-\alpha) \cdot \frac{1}{N}$$



$\alpha = 0.85$

N = # nodi grafo

r(i) = α (r(a) / 3 +
r(b) / 1 +
r(c) / 2)
+ (1 - α) / N

Il PageRank «legato» agli autovettori della matrice del sistema lineare

PageRank, come Cammino Casuale sul grafo del Web

Il PageRank di un nodo è la «frequenza» con cui si visiterebbe quel nodo muovendosi per sempre

Sorta di «misura di centralità» di un nodo nel grafo

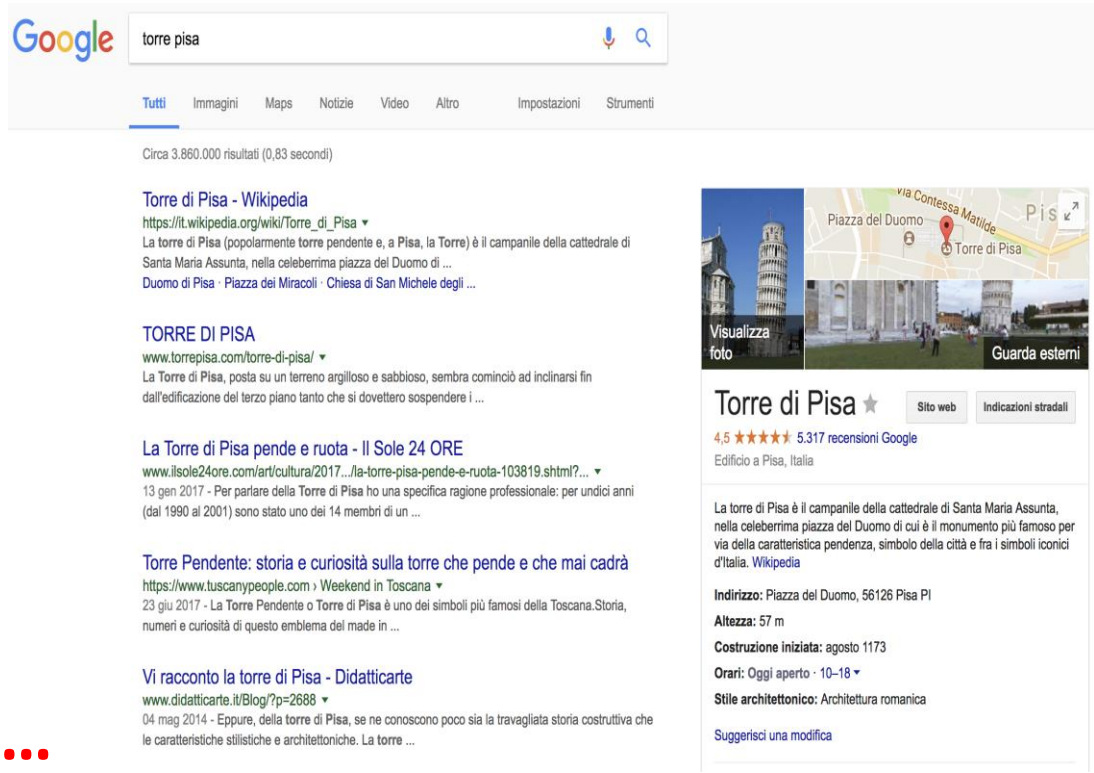
Rilevanza di un documento

È un concetto matematicamente non ben definibile, essendo innanzitutto soggettivo e mutevole nel tempo

Per ogni pagina si calcolano una serie di ***caratteristiche***:

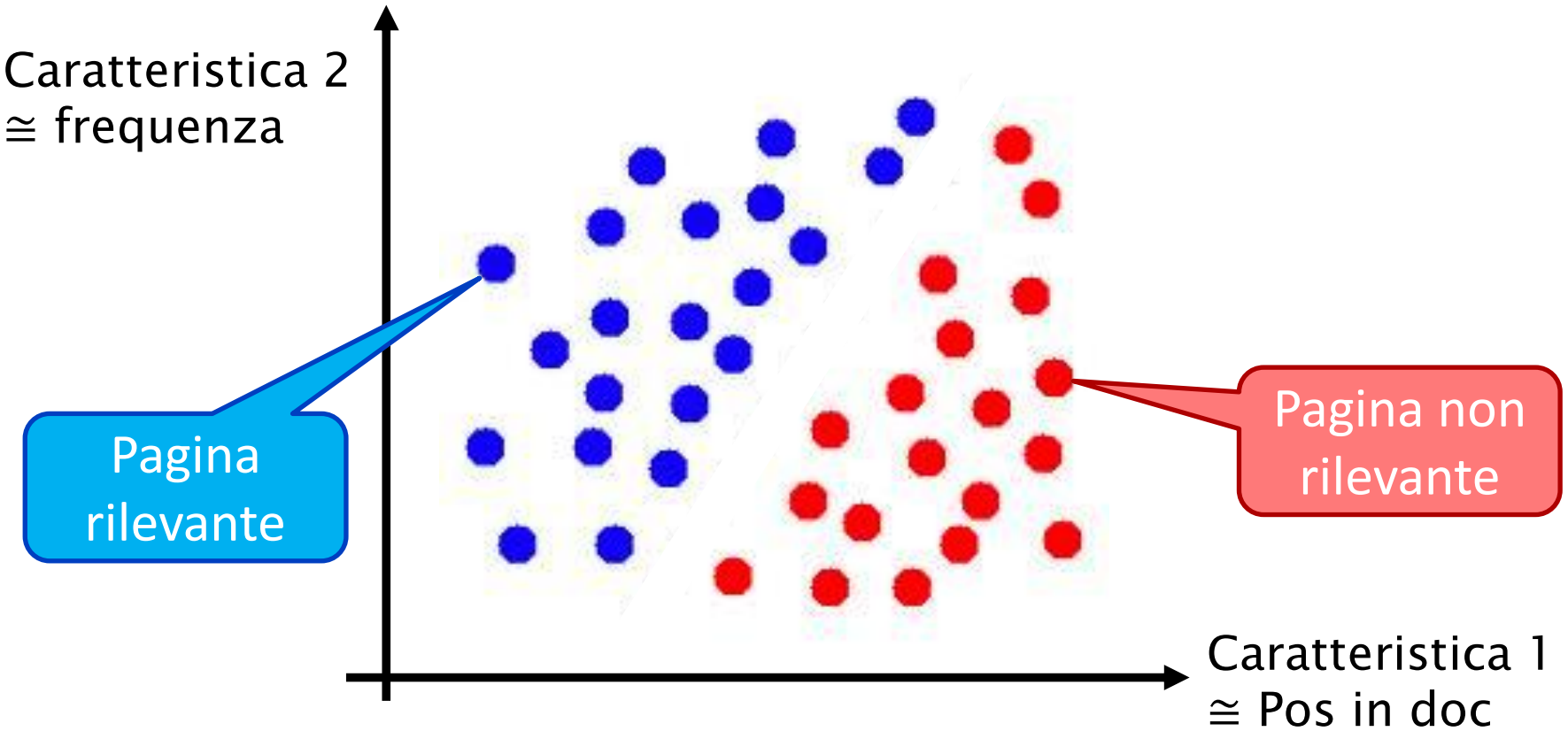
- TF-IDF delle parole
- PageRank
- Loro vicinanza
- Se nel titolo o URL
- Posizione nella pagina
- …

sono oltre 200 «misure»…

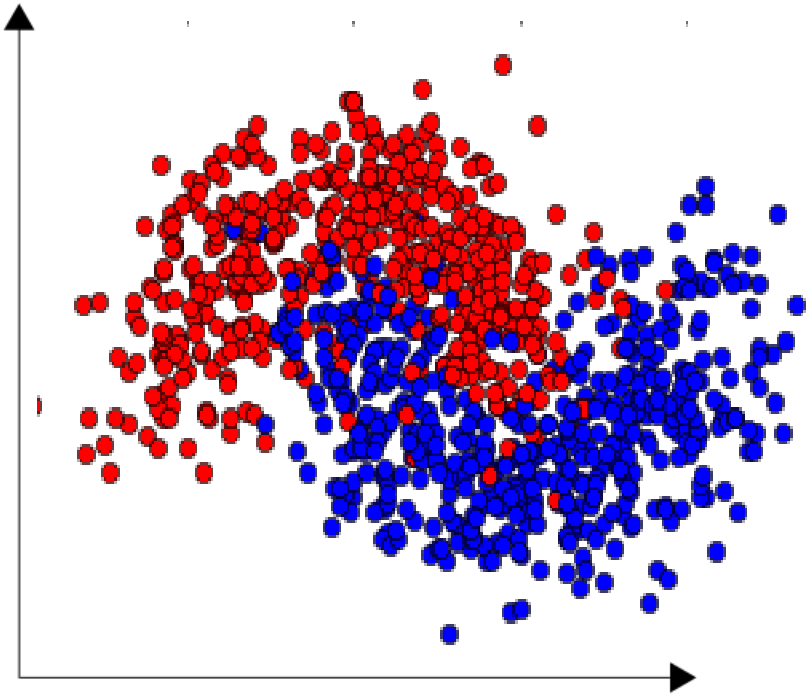


Rilevanza di una Pagina

Legata oggi alle tecniche di **AI e Machine Learning**



IA e motori di ricerca



C’è una ***nuova*** attenzione grazie alla disponibilità di tanti dati (*esempi*), G/CPU molto potenti, e nuovi algoritmi che possono elaborarli (*training*).

L’ultimo passo recente

il paparazzo ha fotografato una stella del cinema

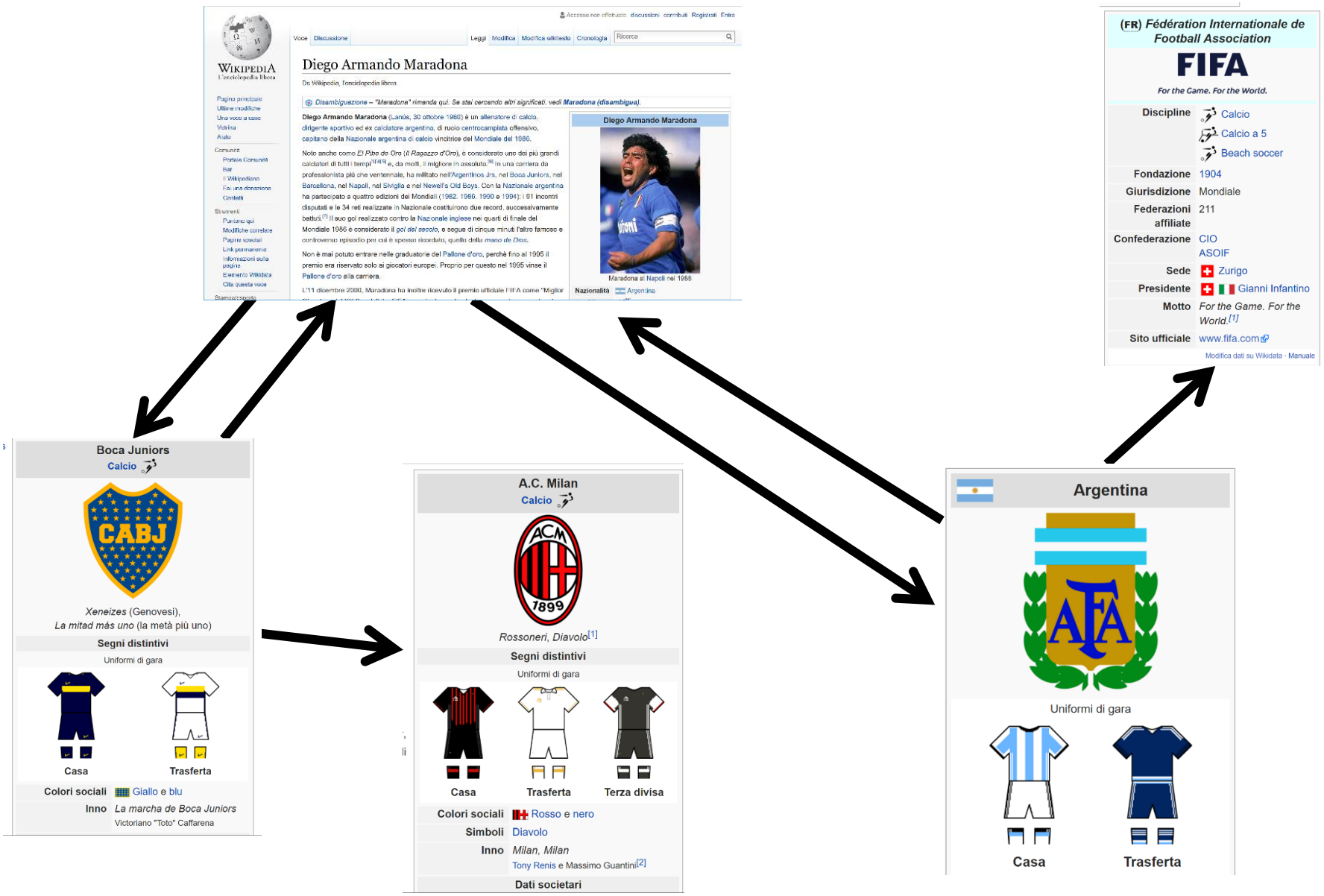
L’ astronomo ha fotografato una stella

Giorgio sta usando il browser di Microsoft

Luca è un fan di Internet Explorer

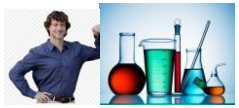
Google | Inside Search
Home
Tips & Tricks
Features
Search Stories
Playground
Blog
Help
Search plus your world
Flight Search
Related Search Previews
Search by Image
Voice Search
Google Instant
Maggio 2012
The Knowledge Graph
Learn more about one of the key breakthroughs behind the future of search.
See it in action
Discover answers to questions you never thought to ask, powered by the Knowledge Graph.

Il Grafo di Wikipedia



Da parole a concetti

Leonardo è lo scienziato che ha dipinto la Mona Lisa



Scienza

Cartografia

Leonardo da Vinci

Italia

Arte

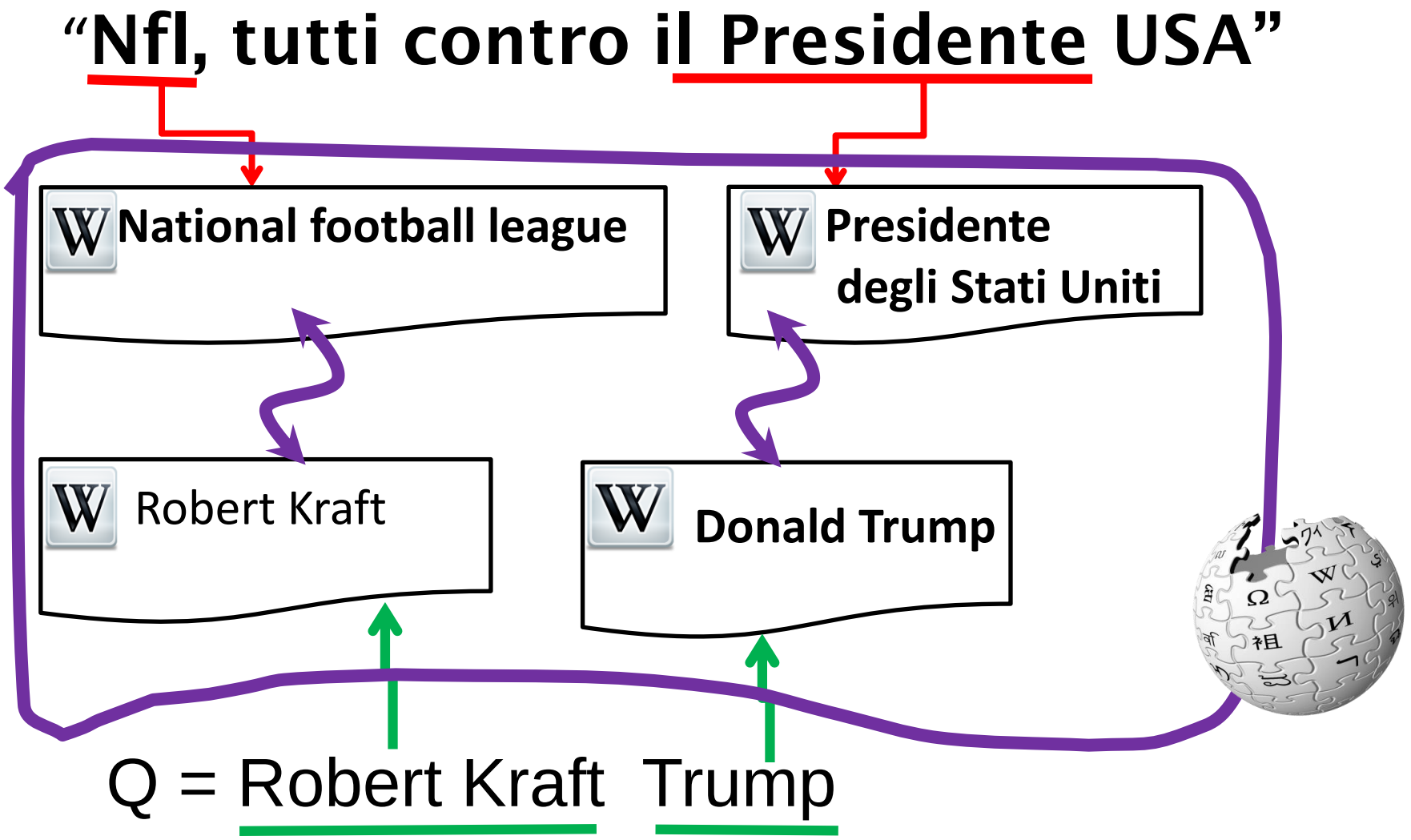
Louvre

Mona Lisa (dipinto)

Rinascimento

Firenze

Ricerche “semantiche”



Einstein, Albert

Enciclopedie on line

Einstein, Albert. - Fisico (Ulma 1879 - Princeton 1955). Trascorsi i primi anni della sua vita a Monaco, dopo un breve soggiorno a Milano **Einstein** si trasferì [...] tedeschi e una sempre più accesa campagna antisemita contro la sua persona e le sue idee, **Einstein**, Albert restò a Berlino fino al 1933. Con l'ascesa al potere di Hitler, abbandonò ... ... Leggi

CATEGORIA: BIOGRAFIE

TAGS: ACCADEMIA PRUSSIANA DELLE SCIENZE – ASSORBIMENTO DELLA RADIAZIONE – INSTITUTE FOR ADVANCED STUDY – SPOSTAMENTO VERSO IL ROSSO – SISTEMA DI RIFERIMENTO

Nascondi altri risultati su EINSTEIN, ALBERT (5)

Einstein, Albert

Dizionario di filosofia (2009)

Fisico tedesco naturalizzato svizzero e statunitense (Ulma 1879 - Princeton, New Jersey, 1955). Vita e carriera scientifica. Di famiglia ebrea, trascorse i primi anni della sua vita a Monaco e, dopo un breve soggiorno a Milano, si trasferì con i suoi in Svizzera. A Zurigo completò gli studi secondari ... Leggi

Einstein, Albert

Enciclopedia dei ragazzi (2005)

Lo scienziato che ha rivoluzionato la fisica del Novecento Lo studio della materia a livello atomico e le teorie cosmologiche che descrivono la struttura e l'evoluzione dell'Universo sono due tra le maggiori conquiste compiute dalla scienza nel Novecento. Entrambe sono nate dalle ipotesi sulla luce ... Leggi

Einstein Albert

Dizionario delle Scienze Fisiche (1996)

Einstein ‹àinstain› Albert (Ulma 1879 - Princeton, Nex Jersey, 1955) Prof. di fisica teorica nell'univ. di Zurigo (1909), poi nell'univ. tedesca di Praga (1910) e nel politecnico di Zurigo (1912); nel 1913, nominato membro dell'Accademia di Prussia, vi ebbe una cattedra di fisica; nel 1921 ricevette ... Leggi

MOSTRA TUTTI

Vocabolario

einstein

einstein ‹àinstain› s. m. [dal nome di A. Einstein (v. einsteiniano)]. – Unità di misura dell'energia raggiante (detta anche mole di fotoni o di quanti), usata spec. in fotochimica, pari – per una radiazione ...

einsteiniano

einsteiniano ‹ainstainiàno› agg. – Che si riferisce al fisico e matematico ted. Albert Einstein (1879-1955), e soprattutto alla teoria della relatività da lui enunciata (v. relatività).

VEDI TUTTI

Tag

ACCADEMIA PRUSSIANA DELLE SCIENZE

ASSORBIMENTO DELLA RADIAZIONE

INSTITUTE FOR ADVANCED STUDY

SPOSTAMENTO VERSO IL ROSSO

SISTEMA DI RIFERIMENTO

tiscali:

Google

qual è il capoluogo della regione calabria

Tutti | Notizie | Immagini | Shopping | Video | Altro | Impostazioni | Strumenti

Circa 77 risultati (0,76 secondi)

Calabria / Capitale

Catanzaro

Ricerche correlate

Visualizza altri 10 elementi

Calabria

Provincia di Catanzaro

Reggio Calabria

Cosenza

Provincia di Cosenza

Crotone

Provincia di Reggio Calabria

Altre informazioni su Catanzaro

Feedback

Google

quanti abitanti ha catanzaro

Tutti | Notizie | Maps | Shopping | Immagini | Altro | Impostazioni | Strumenti

Circa 79 risultati (0,62 secondi)

Provincia di Catanzaro / Popolazione

361.783

30 apr 2017

Feedback

quanti abitanti ha il capoluogo della regione calabria

Tutti | Notizie | Immagini | Shopping | Maps | Altro | Impostazioni | Strumenti

Circa 63 risultati (0,65 secondi)

Calabria - Wikipedia

https://it.wikipedia.org/wiki/Calabria

La Calabria (AFI: /kaˈlabrja/; Calàbbria in napoletano (nelle varianti calabresi), Calàvria in ... Il nome Calabria designava in origine il Salento (i cui **abitanti** erano chiamati .... La pianura più estesa è quella **di** Sibari, situata a nord-est **della** regione. ..... istituita la **Regione Calabria** con Catanzaro **capoluogo di** Regione.

Storia della Calabria · Simboli della Calabria · Comuni della Calabria · Catanzaro

Reggio Calabria - Wikipedia

https://it.wikipedia.org/wiki/Reggio_Calabria

Reggio **di** Calabria (AFI: [ˈrɛdʒdʒodikaˈlabrja] ascolta; Rìggiu in dialetto reggino; Ρήγι, Rìghi in .... Il comune **di** Reggio è il più occidentale **della Regione Calabria**, e la sua ..... Nell'ultimo censimento **del** 2011, grazie al fenomeno **dell'**immigrazione che **ha** compensato la diminuzione **delle** nascite, il numero degli **abitanti ha** ...

Inizio 2019

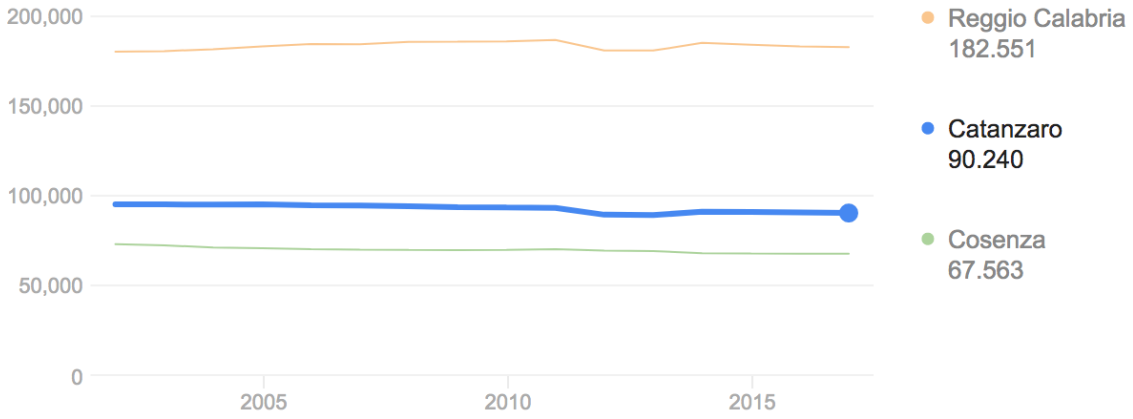
quanti abitanti ha il capoluogo della calabria

All Images News Shopping Maps More Settings Tools

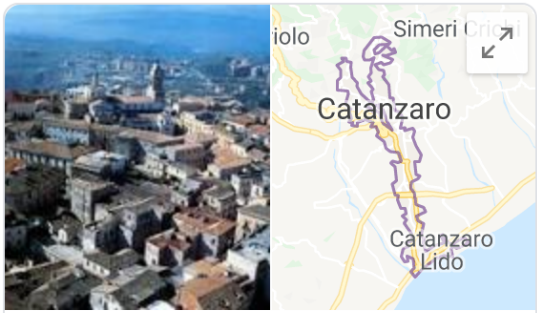
About 91 results (0.96 seconds)

Catanzaro / Popolazione

90.240 (2017)



Ulteriori informazioni



Catanzaro

Comune italiano

Catanzaro è un comune italiano di 88 845 abitanti, capoluogo dell'omonima provincia e della regione Calabria. È il secondo comune della regione per popolazione. Wikipedia

Fine 2019

In conclusione

- Matrici e vettori (rappresentazione di documenti)
- Trigonometria (similarità tra documenti)
- Sistemi di equazioni (PageRank)
- Grafi (annotazione semantica)

Gli algoritmi che sottostanno al progetto dei Motori di Ricerca «usano» tanta Matematica, con un occhio alla «risoluzione efficiente» di problemi sempre più complessi posti dai Big .

Sistemi di Raccomandazione

- Supponiamo di avere una lista di ristoranti
  - Con valutazioni ↑ e ↓ per *alcuni di essi*

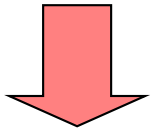
	Brahma Bull	Spaghetti House	Mango	Il Fornaio	Zao	Ming's	Ramona's	Straits	Homma's
Alice		Yes	No	Yes				No	
Bob		Yes				No		No	
Cindy				Yes	No			No	
Dave	No			No	Yes	Yes			Yes
Estie				No	Yes	Yes		Yes	
Fred	No						No		

Quale ristorante consiglieresti a Dave?

Algoritmo #1

>> Raccomanda il ristorante più popolare

- Contando # voti positivi - # voti negativi



	Brahma Bull	Spaghetti House	Mango	Il Fornaio	Zao	Ming's	Ramona's	Straits	Homma's
Alice		1	-1	1				-1	
Bob		1				-1		-1	
Cindy				1	-1			-1	
Dave	-1			-1	1	1			1
Estie				-1	1	1		1	
Fred	-1						-1		

E se a Dave non piacessero gli spaghetti ?

Algoritmo #2

- **Idea**: trova la persona "più simile" a Dave in accordo alla similarità-del-coseno (quindi Estie), e poi suggerisce il ristorante che Estie preferisce (quindi Straits Café).

	Brahma Bull	Spaghetti House	Mango	Il Fornaio	Zao	Ming's	Ramona's	Straits	Homma's
Alice		1	-1	1				-1	
Bob		1				-1		-1	
Cindy				1	-1			-1	
Dave	-1			-1	1	1			1
Estie				-1	1	1		1	
Fred	-1						-1		

☝ Vogliamo basarci sono sul parere di Estie ?