

Sistemi per il recupero delle informazioni

Gabriele Pozzani

A.A. 2014/2015

**Corso di Laurea Magistrale in
Editoria e Giornalismo**

Ricerca semantica

LA STAMPA.it PANORAMA

TORINO - CUNEO - ASTI - NOVARA - VCO - VERCELL - BIE
ATTUALITÀ | OPINIONI | ECONOMIA | SPORT | TORINO | CULTURA
HOME | POLITICA | ESTERI | CRONACHE | COSTUME | TECNOLOGIA
TECNOLOGIA | 16/03/2012 - L'IDEA DEL COLOSSO DI MOUNTAIN VIEW
Google, il futuro è nella ricerca semantica

Il nuovo motore funzionerà come il cervello: metterà in relazione le richieste degli utenti con un database contenente centinaia di milioni

Google "soggiorna" il suo sistema di ri

Web. Nei prossimi mesi, chi digiterà chiave troverà infatti molto di più di

che si trova oggi su Google.

Non solo link, i motori forniranno informazioni e dati funzionando come il cervello umano

colossi web

» Corriere della Sera > Scienze > Ricerca semantica, per Google è il futuro

Ricerca semantica

Non solo link, i motori forniranno informazioni e dati

funzionando come il cervello umano

Big G ha modificato il suo algoritmo. Ed è scoppiata la polemica

"semantica"

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Facebook, i messaggi,

"Facebook", "tastiera",

Zuckerberg urla sulla chat segreta

Sai che cosa è?

Agenda digitale, il

Problema, esempio

A screenshot of a Google search results page for the query "earth". The results are categorized by type: Web, Images, Maps, Shopping, News, More, and Search tools. A green callout bubble labeled "Software?" points to the first result, which is a link to Google Earth's website. A blue callout bubble labeled "Pianeta?" points to the second result, which is a link to Wikipedia's Earth article. An orange callout bubble labeled "Giornata?" points to the third result, which is a link to a NASA Earth Observatory article. The search bar at the top contains the word "earth".

Google earth

Web Images Maps Shopping News More Search tools

About 1,220,000,000 results (0.24 seconds)

Google Earth earth.google.com/
Google Earth lets you fly anywhere on Earth to view satellite imagery, maps, terrain, 3D buildings, from galaxies in outer space to the canyons of the ocean. Download Google Earth - Earth View in Maps - Google Панорама Земли - Connect

Earth - Wikipedia, the free encyclopedia en.wikipedia.org/wiki/Earth
Earth is the third planet from the Sun, and the densest and the largest of the eight planets in the Solar System. It is also the largest of the Solar System's four ... History of the Earth - Structure of the Earth - Age of the Earth - Earth's orbit

Earth | Earth facts, pictures and information nineplanets.org/earth.htm
Earth is the third planet from the Sun and the fifth largest. Find out more...

Google Earth for mobile www.google.com/mobile/earth/
Features. Explore in 3D. View the same 3D imagery, terrain, and buildings* available in the desktop version of Google Earth. Local search. Search for cities ...

NASA Earth Observatory : Home earthobservatory.nasa.gov/
Explore the causes and effects of climatic and environmental change through the use of real satellite data.

Flash Earth - Zoom into satellite and aerial imagery of the Earth i... www.flashearth.com/
Satellite and aerial imagery of the Earth in Flash. Based on different data that can be chosen by a click.

Earth www.universetoday.com/14367/earth/
by Fraser Cain - In: 677,906 Google+ circles - More by Fraser Cain 12 Mar 2012 – The Earth is the third planet from the Sun and the fifth largest of the eight planets in our Solar System. It is located between Mars and Venus, ...

Earth Day Network | Earth Day 2013 www.earthday.org/
You can make a difference! Help Earth Day Network – that works with over

5

Problema (III)

- “earth” è solo una parola, una stringa, una sequenza di caratteri
 - I risultati dei diversi significati sono mescolati tra loro
 - Non sono organizzati in base al significato

6

La ricerca semantica

- Obiettivo della ricerca semantica è di
 - aggiungere semanticità ai termini
 - Non più stringhe, ma concetti/entità
 - Correlare i documenti non a parole ma a concetti
 - Organizzare i risultati in base ai concetti
 - Cercare di comprendere quale concetto era sottinteso dall'utente nella propria ricerca (ed esplicitarlo)

7

Attuale evoluzione

The screenshot illustrates the evolution of semantic search from simple keyword matching to complex entity-based results.

Top Result: "Earth - Wikipedia, the free encyclopedia" (Web) - This is the primary result, showing a detailed summary of Earth as a planet, its size, and its position in the solar system.

Second Result: "Google Earth" (Web) - A link to the official Google Earth website, which allows users to view satellite imagery, maps, terrain, 3D buildings, and other features.

Third Result: "Earth (band), White Noise, the free encyclopedia" (Web) - A link to a page on the free encyclopedia website, which provides information about the band Earth.

Fourth Result: "Earth (Earth, Blindfolded, 1982)" (Web) - A link to a page about the song "Earth" by the band Earth, featuring lyrics and a link to a video.

Fifth Result: "Earth facts, statistics and information – The Nine Planets" (Web) - A link to a page on the Nine Planets website, which provides facts and statistics about Earth.

Annotations:

- A green arrow points from the "Earth - Wikipedia" result down to the "Earth (band)" result, indicating the shift from a general entity to a specific entity.
- A blue arrow points from the "Earth (band)" result down to the "Earth facts" result, further illustrating the narrowing of the search focus.
- A red arrow points from the "Earth (band)" result to the right side of the page, where a list of related songs and albums is displayed.

Non solo google

The figure displays four separate search results from WolframAlpha:

- Search term:** earth
- Input interpretation:** Earth (planet)
- Basic properties:**
 - current distance from Sun: 0.9835 au
 - 8.179 light minutes
 - largest distance from orbit center: 1.5209701 × 10⁸ Km
 - 1.0167033 au
 - nearest distance from orbit center: 1.470680724 × 10⁸ Km
- Definitions:**
 1. noun the 3rd planet from the sun; the planet we live on
 2. noun the loose soft material that makes up a large part of the land surface
 3. noun the solid part of the earth's surface
 4. noun the abode of mortals (as contrasted with Heaven or Hell)
- Assuming "earth" is a general material or a book or a movie countries instead**
- Assuming "earth" is a planet | Use as a general material or a book or a movie countries instead**
- Input interpretation:** Parola (English word)
- Basic properties:**
 - author: Jon Stewart | David Javerbaum | Rory Albanese | Steve Bodow | Josh Leib | Kevin Bleyer | Rich Blomquist | Tim Carvell | Kevin Wyatt Cenac | Hallie Hartland | J. R. Mcond
- Definitions:**
 - 1. noun the 3rd planet from the sun; the planet we live on
 - 2. noun the loose soft material that makes up a large part of the land surface
 - 3. noun the solid part of the earth's surface
 - 4. noun the abode of mortals (as contrasted with Heaven or Hell)
- Assuming "earth" is all countries | Use as a planet or a general material or a word or a book or a movie instead**
- Input interpretation:** Planeta
- Basic properties:**
 - Input interpretation: Earth (planet)
 - Orbital properties:
 - current distance from Sun: 0.9835 au
 - 8.179 light minutes
 - largest distance from orbit center: 1.5209701 × 10⁸ Km
 - 1.0167033 au
 - nearest distance from orbit center: 1.470680724 × 10⁸ Km
 - Physical properties:
 - angle of repose: 35° (Fuller's)
 - 37.5° (moist)
 - 37.5° (dry loam)
 - 50° (black loam)
- Definitions:**
 1. noun the 3rd planet from the sun; the planet we live on
 2. noun the loose soft material that makes up a large part of the land surface
 3. noun the solid part of the earth's surface
 4. noun the abode of mortals (as contrasted with Heaven or Hell)
- Assuming "earth" is a general material or a book or a movie countries instead**
- Assuming any type of soil | Use soil, fuller's, or movie instead**
- Input interpretation:** Materiale
- Physical properties:**
 - angle of repose: 35° (Fuller's)
 - 37.5° (moist)
 - 37.5° (dry loam)
 - 50° (black loam)
- Definitions:**
 1. noun the 3rd planet from the sun; the planet we live on
 2. noun the loose soft material that makes up a large part of the land surface
 3. noun the solid part of the earth's surface
 4. noun the abode of mortals (as contrasted with Heaven or Hell)
- Assuming "earth" is a general material or a book or a movie countries instead**
- Assuming any type of soil | Use soil, fuller's, or movie instead**

Altri motori di ricerca

- Esistono altri motori di ricerca semanticici, anche meno conosciuti

- Bing
- GoPubMed
 - Motore di ricerca semantico con capacità di annotazione/arricchimento semantico dei siti web in realtime
- Igloo
 - Motore di ricerca semantico con capacità di annotazione/arricchimento semantico per ricette e cibo
- Yummly
 - Motore di ricerca semantico per ricette e cibo

Come funziona (I)

- L'idea di base è simile a quella vista per i thesauri
 - I termini riconducibili allo stesso significato e concetto vengono raggruppati in un unico insieme (Synset)
 - I synset vengono collegati tra loro in base alle relazioni semantiche (le più disparate) esistenti tra i concetti che rappresentano
 - Si costruisce così un grafo (rete semantica) in cui
 - I nodi rappresentano i concetti/synset
 - Gli archi rappresentano le relazioni semantiche tra i concetti
 - L'informazione non strutturata (classiche pagine web) deve essere strutturata (grafo) per essere fruita al meglio

11

Come funziona (II)

- Si entra nell'ambito della “rappresentazione della conoscenza” e dei “sistemi esperti”
- In questo campo non si parla di “thesauro” ma di “ontologia” e “knowledge base”

Come funziona (III)

- I termini ambigui di una ricerca vengono disambiguati vedendo se appartengono a synset/concetti legati tra loro
 - Termini utilizzati in una stessa ricerca, molto probabilmente, sono correlati tra loro nel significato
 - “red apple” è molto più probabile che abbia il significato di “mela rossa” che non di “Apple® rosso”

13

Google Knowledge Graph (I)

- Il nome deriva proprio dalla sua rappresentazione a grafo dei concetti
 - Una ricerca
 - oltre ad essere eseguita sulle pagine web viene effettuata anche sul knowledge graph
 - Quando un termine non può essere disambiguato, tutti i concetti ad esso riconducibili nel grafo vengono mostrati

14

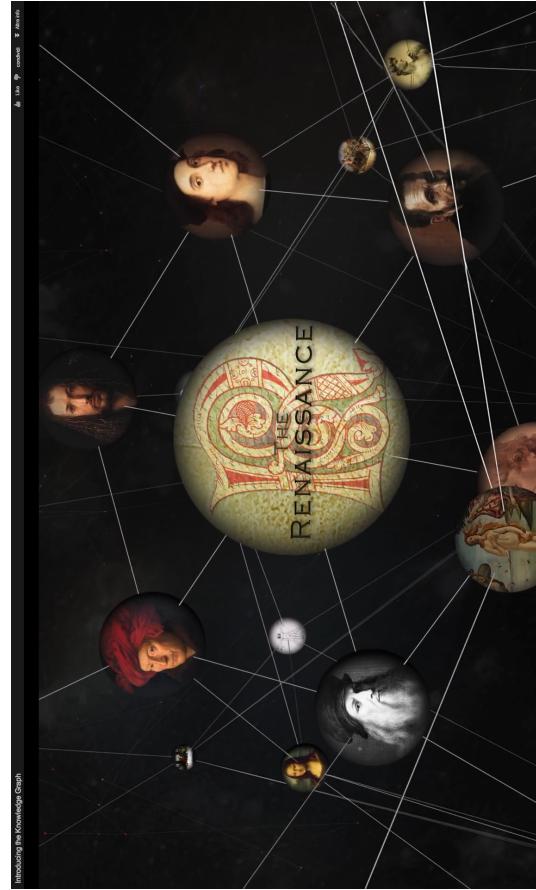
Google Knowledge Graph, esempio (I)

The screenshot shows a Google search results page for the query "sun". At the top, there's a navigation bar with "Google", a search input field containing "sun", and a blue "Search" button. Below the bar, it says "About 2,620,000,000 results (0.24 seconds)". The main content area starts with a large result for "The Sun" (www.thesun.co.uk/), which is described as "The Best for News, Sport, Showbiz, Celebrities | The Sun". It includes a thumbnail of the newspaper, a brief summary, and links to "News", "Football", and "Oracle and Sun". A callout box highlights the link to "Oracle and Sun" with the text: "Cliccando su un concetto, Si inizia a ‘navigare’ il grafo". Below this, there are several other search results, including "The Sun NewsVoice of The Nation" (sunnewsonline.com/), "The Sun (Micros)" (www.oracle.com/us/sun/index.htm), and a "Recent posts" section with a link to "£62' lino axed".

15

Google Knowledge Graph (II)

- La struttura a grafo permette di navigare tra i concetti
 - Ci si sposta da un nodo/synset/concetto ad altro ad esso collegato
 - La ricerca (Web e/o su KG) corrispondente viene eseguita



Google Knowledge Graph, esempio (III)

Google search results for "sun life cycle". The results include:

- A large image of the Sun.
- A summary card for the Sun with details: "The Sun is the star at the center of the Solar System. It is almost perfectly spherical and consists of hot plasma interwoven with magnetic fields.", a link to "Sun - Wikipedia, the free encyclopedia", and a video thumbnail for "Life cycle of our Sun - YouTube".
- A snippet about the Sun's life cycle: "The Sun by the Atmospheric Imaging Assembly of NASA's Solar Dynamics Observatory".
- A snippet about the Sun's path: "Sun path - Faint young Sun paradox - Sun dog Sun (disambiguation) - Sun path - Sun dog".
- A snippet about the Sun's orbit: "Orbits: Galactic Center".
- A section titled "People also search for" with links to "Moon", "Earth", "Mars", "Venus", "Jupiter", and "Saturn".
- A blue callout box with the text: "Cliccando su un altro concetto, lo si ‘esplora’".
- A link to "More videos for sun life cycle".

17

Google Knowledge Graph, esempio (III)

Google search results for "earth's moon". The results include:

- A large image of the Moon.
- A summary card for the Moon with details: "Jump to Appearance from Earth. The Moon is in synchronous rotation: it rotates about its axis in about the same time it takes to orbit the Earth. This results ... Orbit of the Moon - Moon landing - Moon-landing conspiracy theory - Natural satellite Moon - Wikipedia, the free encyclopedia".
- A snippet about the Moon's appearance: "Our Moon makes Earth a more livable planet by moderating our home planet's wobble on its axis, leading to a relatively stable climate, and ...".
- A snippet about the Moon's orbit: "Orbit of the Moon - Moon landing - Moon-landing conspiracy theory - Natural satellite Moon - Facts and Information about the Earth's Moon | Space.com".
- A snippet about the Moon's composition: "Scientists say a giant impact knocked off the raw ingredients for the moon off the primitive molten Earth and into orbit. SPACE.com has an overview of Earth's ...".
- A snippet about the Moon's phases: "Earth's Moon Phases | Lunar Curves | Infographic | Share.com".
- A section titled "People also search for" with links to "Moon", "Earth", "Mars", "Venus", "Jupiter", and "Mercury".
- A grid of images for other planets: "Uranus", "Neptune", "Saturn", "Mercury", "Venus", "Jupiter", "Mars", "Earth", and "Moon".
- A section titled "More images" with a grid of four images of the Moon.

18

Google Knowledge Graph (III)

- Knowledge Graph ha attualmente affiancato e integrato la “classica” ricerca web
- Si può
 - Navigare nel grafo
 - Navigare i risultati della ricerca testuale web
- Inoltre eventuali termini “base” nella ricerca effettuata sul KG vengono ricondotti ad apposite ricerche web classiche, prestatabilité anche dai termini che identificano il concetto “in cui ci si trova”

19

Come viene costruito il KG? (I)

- Informazioni sui concetti vengono tratte da database pubblici, inclusi:
 - Freebase
 - database gratuito e aperto di oltre 24 milioni di “cose”, tra cui film, libri, spettacoli televisivi, personaggi famosi, luoghi, aziende
 - Wikipedia
 - CIA World Factbook
 - pubblicazione annuale della CIA che riporta i dati statistici fondamentali e una sintesi di informazioni riguardanti tutti i paesi del mondo
 - Risorse specifiche per argomento
 - e.g., Weather Underground per informazioni meteo e la Banca Mondiale per statistiche economiche
 - I dati di ricerca di Google
 - usati per misurare la popolarità di un argomento e contribuire a decidere quali sono le informazioni più richieste

20

Come viene costruito il KG? (III)

- I concetti e le loro relazioni vengono inferiti
 - Utilizzando thesauri e ontologie già esistenti
 - Automaticamente analizzando i termini chiave delle pagine web e come le pagine web sono collegate dai link
- Secondo i dati rilasciati nel 2012 e ribaditi nel 2014, Knowledge Graph contiene dati su
 - 570M di entità
 - 18G di fatti

21

Google Knowledge Graph (IV)

- Il motto del progetto è “Things, not strings” (“cose, non parole”)
 - Il focus delle ricerche si sposta dai termini/keyword agli oggetti/entità che esse rappresentano
- Google: non più “motore di ricerca”, ma “motore di conoscenza”

22

Google Knowledge Vault (I)

- Sembra Google stia già lavorando all'evoluzione di KG:
 - Google Knowledge Vault
- Data l’“esiguità” delle fonti da cui trarre concetti e fatti con semantica strutturata, KG è abbastanza limitato
 - Usare l’intero Web come fonte per concetti semantici
- Knowledge Vault è un algoritmo capace
 - di estrarre le informazioni dal Web e
 - di comporre i dati grezzi in vere e proprie porzioni di conoscenza
 - utilizzando strumenti di machine learning e di intelligenza artificiale
 - Annunciato nell’agosto 2014

23

Google Knowledge Vault (II)

- KV estrae automaticamente la conoscenza dalle pagine web analizzando
 - Testo delle pagine
 - La struttura delle pagine HTML
 - Le tabelle contenute nelle pagine HTML
 - Annotazioni aggiunte alle pagine HTML dagli autori stessi
- Essendo l'estrazione automatica, restituisce risultati “probabilistici”
- Secondo dati del 2014, KV ha estratto:
 - 1,6G fatti, di cui
 - 324M corretti con una probabilità di almeno il 70%
 - 271M corretti con una probabilità di almeno il 90%

24

Annotazioni HTML

- Le annotazioni HTML sono metadati che possono essere aggiunti dagli autori delle pagine stesse
- Questi metadati aggiungono **semantica**, descrivendo oggetti e la loro natura
 - Ad es.: persone, luoghi, aziende, eventi
- Questa informazione deve essere strutturata utilizzando ontologie già esistenti
 - Microdata from schema.org
 - Microformats.org

25

Microdata

- Estensione di HTML
- Permette di annotare/arricchire le pagine web (HTML) con nuovi marcatori che includono informazioni semantiche

26

Microdata, esempio (1)

```
<section> Hello, my name is John Doe, I am a graduate research  
assistant at the University of Dreams. My friends call me Johnny.  
You can visit my homepage at <a  
href="http://www.JohnnyD.com">www.JohnnyD.com</a>.  
  
I live at 1234 Peach Drive Warner Robins, Georgia.</section>
```



```
<section itemscope itemtype="http://schema.org/Person">  
  
Hello, my name is <span itemprop="name">John Doe</span>,  
I am a <span itemprop="jobTitle">graduate research assistant</span>  
at the <span itemprop="affiliation">University of Dreams</span>.  
My friends call me <span itemprop="additionalName">Johnny</span>.  
You can visit my homepage at  
<a href="http://www.JohnnyD.com" itemprop="url">www.JohnnyD.com</a>.  
  
<section itemprop="address" itemscope  
itemtype="http://schema.org/PostalAddress">  
  
I live at <span itemprop="streetAddress">1234 Peach Drive</span>  
<span itemprop="addressLocality">Warner Robins</span>,  
<span itemprop="addressRegion">Georgia</span>.  
  
</section>  
</section>
```

27

Microdata, esempio (1)

Un web engine comprenderà il microdata come

Item

Type: http://schema.org/Person

name = John Doe

jobTitle = graduate research assistant

affiliation = University of Dreams

additionalName = Johnny

url = http://www.johnnyd.com/

address = Item(1)

Item 1

Type: http://schema.org/PostalAddress

streetAddress = 1234 Peach Drive

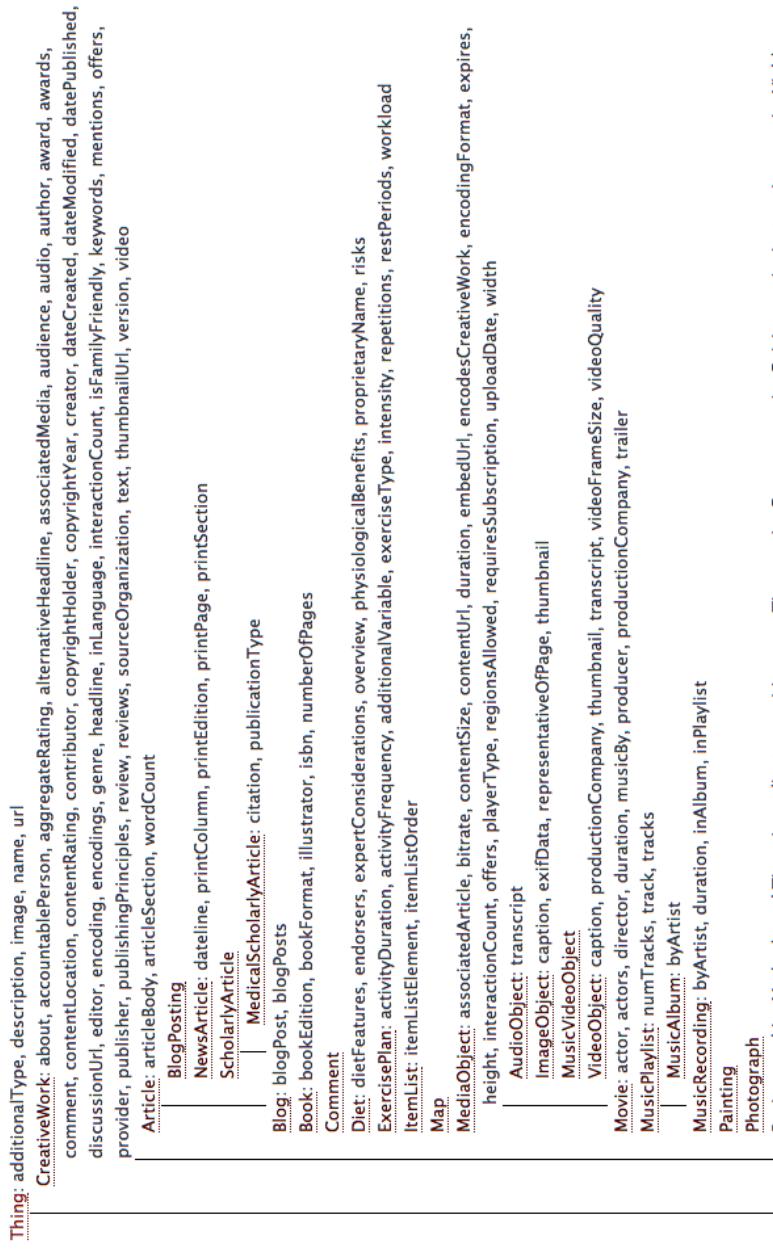
addressLocality = Warner Robins

addressRegion = Georgia

28

Semantica di microdata

- Gli oggetti, con i loro attributi, definibili in <http://schema.org>



29

Riferimenti

- Ricerca semantica
 - http://en.wikipedia.org/wiki/Semantic_search
 - <http://searchenginewatch.com/>
- Google Knowledge Graph e Knowledge Vault
 - <http://www.google.com/intl/it/insidesearch/features/search/knowledge.html>
 - <http://googleblog.blogspot.it/2012/05/introducing-knowledge-graph-things-not.html>
 - <http://support.google.com/websearch/bin/answer.py?hl=it&answer=2620861>
 - <http://searchenginewatch.com/search?q=%22knowledge+graph%22>
 - <http://www.cs.cmu.edu/~nlao/publication/2014.kdd.pdf>
- Ricerca semantica e SEO
 - <http://searchenginewatch.com/article/2214849/Googles-Knowledge-Graph-Implications-for-Search-SEO>
 - <http://searchenginewatch.com/article/2234448/Keywords-Are-Dead-Long-Live-User-Intent>
- Annotazioni:
 - <https://support.google.com/webmasters/answer/99170>
 - <http://schema.org/>
 - <http://getschema.org/>
 - <http://schema-creator.org/>
 - <http://www.microdatagenerator.com/>

30