



DEEP
LEARNING
INSTITUTE



DLI Accelerated Data Science Teaching Kit

Lecture 5.1 - Knowledge Graph



The Accelerated Data Science Teaching Kit is licensed by NVIDIA, Georgia Institute of Technology, and Prairie View A&M University under the [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

What is **Data Integration**?

Combining data from **multiple sources** to provide the user with a **unified view**.

Why is it **Important**?

Think about the apps, websites, and services that you use every day.

Businesses **derive value**
through data integration.

About 467,000,000 results (0.64 seconds)

Atlanta, GA : Home

<https://www.atlantaga.gov/>

It is a great time to be in the City of Atlanta! Whether you're a native, first time visitor, business traveler who makes regular stops here, or one of the thousands of ...

Atlanta - Wikipedia

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

Atlanta is the capital of and the most populous city in the U.S. state of Georgia, with an estimated 2016 population of 472,522. Atlanta is the cultural and ...

[Atlanta \(disambiguation\)](#) · [Hartsfield–Jackson Atlanta](#) · [Demographics of Atlanta](#)

Top stories



Crash kills brothers returning to Georgia Southern after Thanksgiving br...

AJC.com

18 hours ago



Postal worker shot in the head, killed outside Atlanta

Fox News

20 hours ago



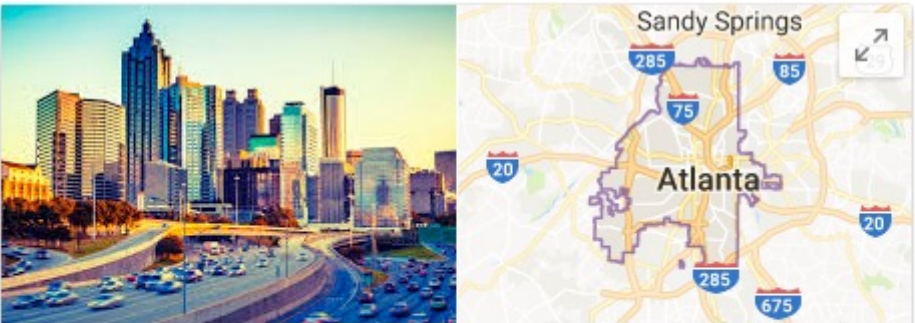
New FOX 5 poll: Virtual dead heat in Atlanta mayor's race

FOX 5 Atlanta

9 hours ago

More for atlanta

Things to do in Atlanta



Atlanta

City in Georgia

Atlanta is the capital of the U.S. state of Georgia. It played an important part in both the Civil War and the 1960s Civil Rights Movement. Atlanta History Center chronicles the city's past, and the Martin Luther King Jr. National Historic Site is dedicated to the African-American leader's life and times. Downtown, Centennial Olympic Park, built for the 1996 Olympics, encompasses the massive Georgia Aquarium.

Weather: 51°F (11°C), Wind NE at 3 mph (5 km/h), 96% Humidity

Local time: Wednesday 6:14 AM

Population: 472,522 (2016)

Area code: 404

Plan a trip

Atlanta travel guide

3-star hotel averaging \$128, 5-star averaging \$366

Upcoming Events

Colleges and Universities: [Emory University](#), [MORE](#)

Did you know: Atlanta has the world's fourth-busiest city airport system by passenger traffic (101,491,106 total passengers). [wikipedia.org](https://www.wikipedia.org)

People also search for

View 15+ more



Apple Siri

Getting Answers



"Do I need an umbrella today?"



"How is the Nikkei doing?"



"When is daylight saving time?"

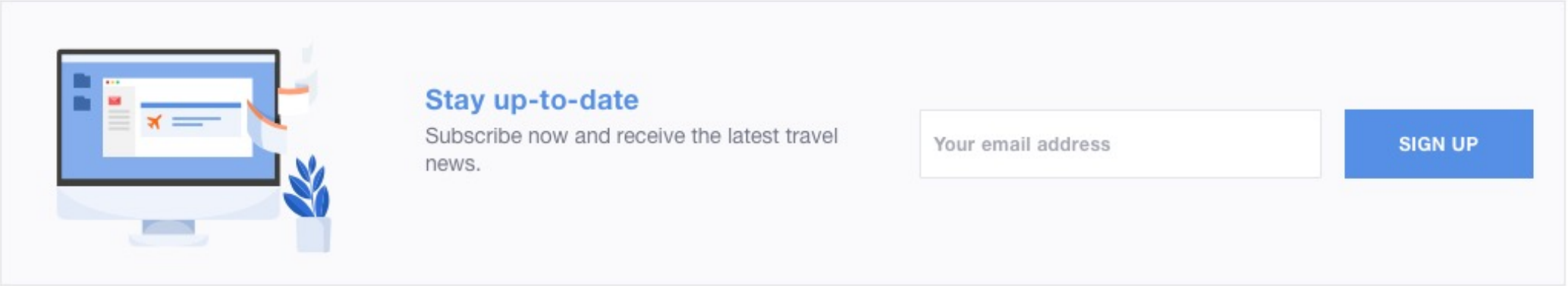
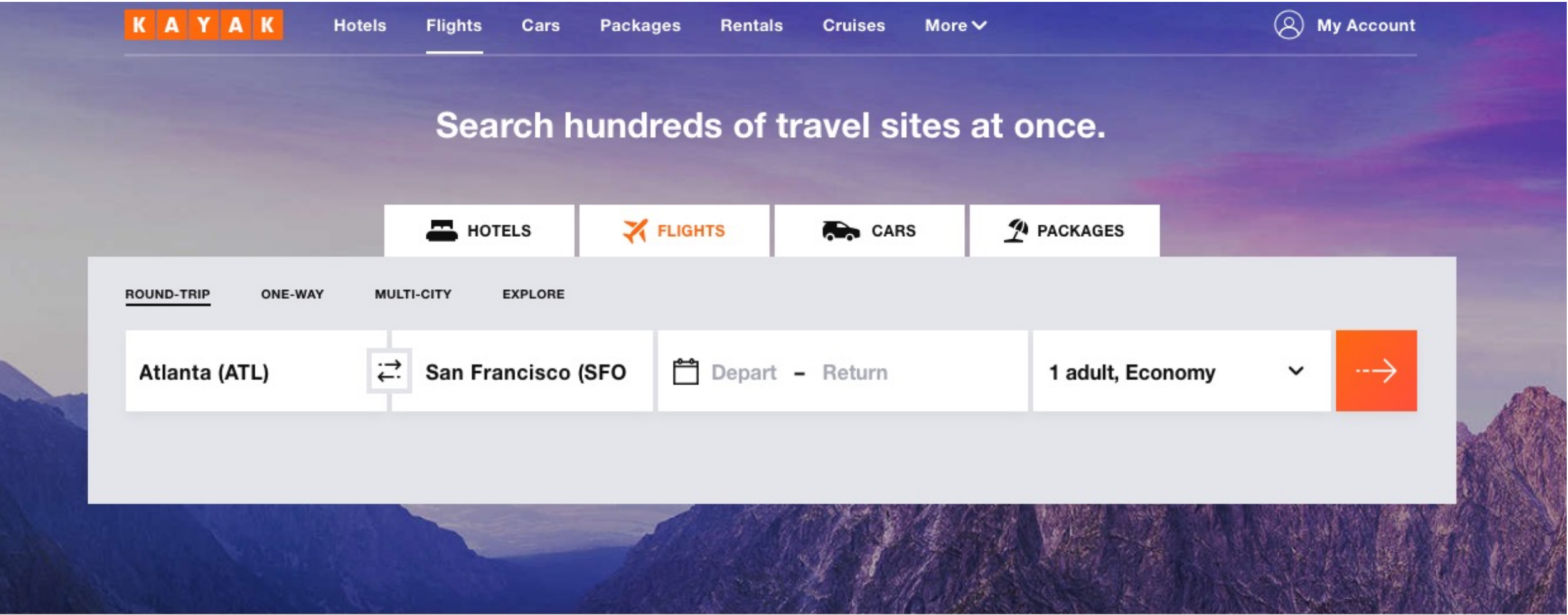


"What's the latest in San Francisco?"

See what people are saying on social media about a place or event.



"Was that an earthquake?"



Recommended for you



More Examples?

- Social media (data from users, businesses)
 - Facebook: your posts, advertisements, review
- Search engine: Google, Bing, Yahoo, etc.
- Smart assistants: Siri, Cortana, Alexa
- Price comparison: Kayak
- Uber, Lyft: drivers, traffic data, customers
- Google maps: users, restaurants, traffic....

How to do data integration?

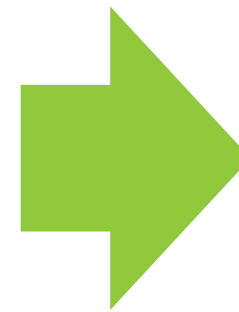
“Low” Effort Approaches

1. Use database’s “Join”! (e.g., SQLite)

When does this approach work? (Or, when does it NOT work?)

| id | name |
|-----|---------|
| 111 | Smith |
| 222 | Johnson |
| 333 | Lee |

| id | salary |
|-----|--------|
| 111 | \$40k |
| 222 | \$60k |
| 333 | \$50k |



| id | name | salary |
|-----|---------|--------|
| 111 | Smith | \$40k |
| 222 | Johnson | \$60k |
| 333 | Lee | \$50k |

2. Open Refine

<http://openrefine.org> (Video #3 “Reconcile and Match Data”)

IDs are really important, and
can simplify data integration!

But who creates the IDs?

Crowd-sourcing Approaches: Freebase

The screenshot shows the Freebase website interface. At the top, there's a search bar with the text "Find..." and navigation links for "Browse", "Query", and "Help". On the right, there are links for "Sign In or Sign Up" and a language dropdown set to "English". A prominent orange banner across the top states: "Important! Freebase is read-only and will be shut-down. More." Below this, a large dark box displays the number "3,179,263,202" with the text "Facts (and counting)" next to it. Underneath, a blue banner reads "A community-curated database of well-known people, places, and things". A navigation bar below the banner includes tabs for "Data", "Schema", "Queries", "Apps", "Loads", "Review Tasks", and "Users", with "Data" being the active tab. The main content area is titled "Explore Freebase Data" and features a table with the following data:

| Domain | ID | Topics | Facts |
|---------------------|---------------------|--------|-------|
| Music | /music | 33M | 240M |
| Books | /book | 6M | 15M |
| Media | /media_common | 6M | 17M |
| People | /people | 4M | 20M |
| Film | /film | 2M | 22M |
| Location | /location | 2M | 20M |
| TV | /tv | 2M | 19M |
| Business | /business | 1M | 4M |
| Fictional Universes | /fictional_universe | 1M | 1M |
| Organization | /organization | 996K | 4M |
| Biology | /biology | 966K | 5M |

To the right of the table, there's a sidebar with a yellow header "How can you get started?". It contains two sections: "Learn how it works" with a paragraph about the database and a "Keep reading »" link, and "Use Freebase data" with a paragraph about the open license and two bullet points: "Query Freebase using our Search, Topic, or MQL APIs" and "Download our weekly data dumps". At the bottom of the sidebar is a "Join the Community" section with a bullet point: "Follow Freebase on G+".

Freebase intro video: <https://youtu.be/TJfrNo3Z-DU>

Learn more about Freebase at <https://en.wikipedia.org/wiki/Freebase>

Freebase

(a graph of entities)

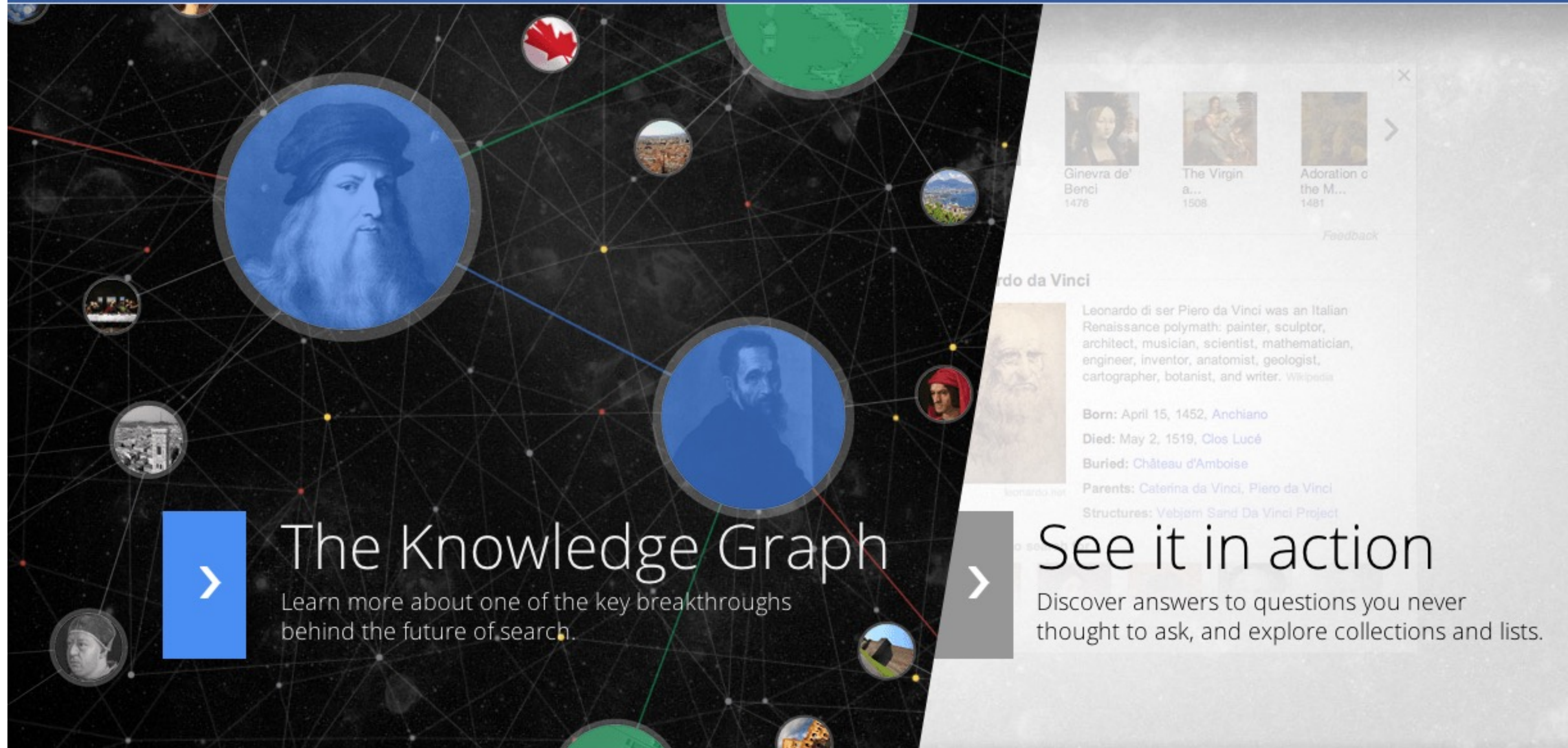
“...a large collaborative knowledge base consisting of metadata composed mainly by its **community members**...”

Wikipedia.

So what?

What can you do with the Freebase knowledge graph?

Hint: Google acquired it in 2010.



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, and explore collections and lists.

Leonardo da Vinci

Leonardo di ser Piero da Vinci was an Italian Renaissance polymath: painter, sculptor, architect, musician, scientist, mathematician, engineer, inventor, anatomist, geologist, cartographer, botanist, and writer. Wikipedia

Born: April 15, 1452, Anchiano
Died: May 2, 1519, Clos Lucé
Buried: Château d'Amboise
Parents: Caterina da Vinci, Piero da Vinci
Structures: Vebjem Sand Da Vinci Project

Ginevra de' Benci 1478
The Virgin & Child 1508
Adoration of the M... 1481

Feedback

Freebase replaced by **Google Knowledge Graph API**



Example:
**What does Google know
about Taylor Swift?**

<https://developers.google.com/knowledge-graph/>



What does Google know about Taylor Swift?

<https://developers.google.com/knowledge-graph/>

```
"@type": "ItemList",
"itemListElement": [
  {
    "@type": "EntitySearchResult",
    "result": {
      "@id": "kg:/m/0dl567",
      "name": "Taylor Swift",
      "@type": [
        "Thing",
        "Person"
      ],
      "description": "Singer-songwriter",
      "image": {
        "contentUrl": "https://t1.gstatic.com/images?q=tbn:ANd9GcQmVDAhjhWnN2OWys2ZM03PGAhu",
        "url": "https://en.wikipedia.org/wiki/Taylor_Swift",
        "license": "http://creativecommons.org/licenses/by-sa/4.0"
      },
      "detailedDescription": {
        "articleBody": "Taylor Alison Swift is an American singer-songwriter and actress. R",
        "url": "http://en.wikipedia.org/wiki/Taylor_Swift",
        "license": "https://en.wikipedia.org/wiki/Wikipedia:Text_of_Creative_Commons_Attrib",
        "url": "http://taylorswift.com/"
      }
    }
  }
]
```

Google has the Knowledge Graph.

Facebook has...

facebook

Email or Phone

Password

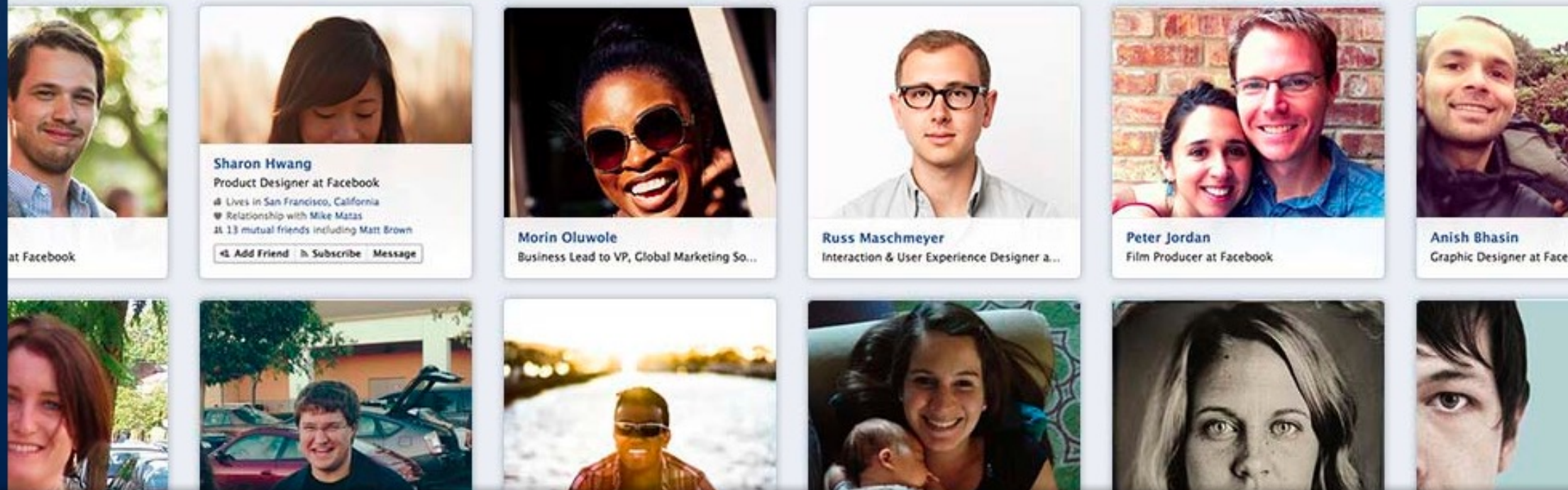
Log In

Sign Up

Connect and share with the people in your life.

Introducing Graph Search

Q People who like **Cycling** and are from my hometown



Find people who share your interests

Want to start a book club or find a gym buddy? Connect with friends who like the same activities—and meet new people, too.

Graph Search intro video: <https://youtu.be/W3k1USQbq80>



DEEP
LEARNING
INSTITUTE



DLI Accelerated Data Science Teaching Kit

Thank You