

A new era:
Topic-based Annotators

“Diego Maradona won against Mexico”

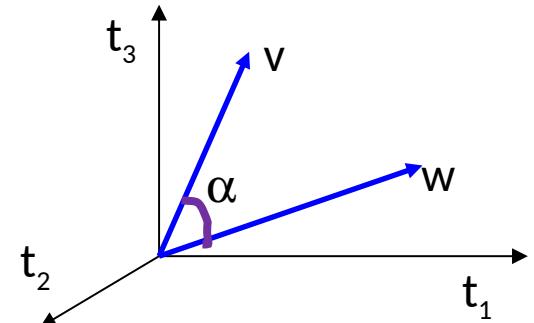
Dictionary of terms

against
Diego
Maradona
Mexico
won

Term Vector

2.2
5.1
9.1
1.0
0.1

Vector Space model



$$\text{Similarity}(v, w) \approx \cos(\alpha)$$

Mainly term-based:
polysemy and synonymy issues

He is using Microsoft's browser

He is a fan of Internet Explorer

the paparazzi photographed the star

the astronomer photographed the star

A new approach: Massive graphs of entities and relations



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

The screenshot shows the Google Inside Search interface with a focus on the Knowledge Graph. On the left, there's a large network graph visualization featuring circular nodes with images of historical figures like Leonardo da Vinci and other entities. A prominent blue node in the center represents Leonardo da Vinci. A red arrow points from this central node to a detailed card on the right. The card for Leonardo da Vinci includes his portrait, a brief biography, birth and death dates, burial location, inventions, and parents. It also lists five of his famous works with small thumbnail images: Mona Lisa, The Last Supper, Virgin of the Rocks, Lady with an Ermine, and The Battle of Anghiari. At the bottom of the card, there's a link to read more on wikipedia.org. Below the graph, a large red text overlay reads "May 2012". At the bottom of the page, there are two calls-to-action: "The Knowledge Graph" with a blue arrow pointing right, and "See it in action" with a grey arrow pointing right.

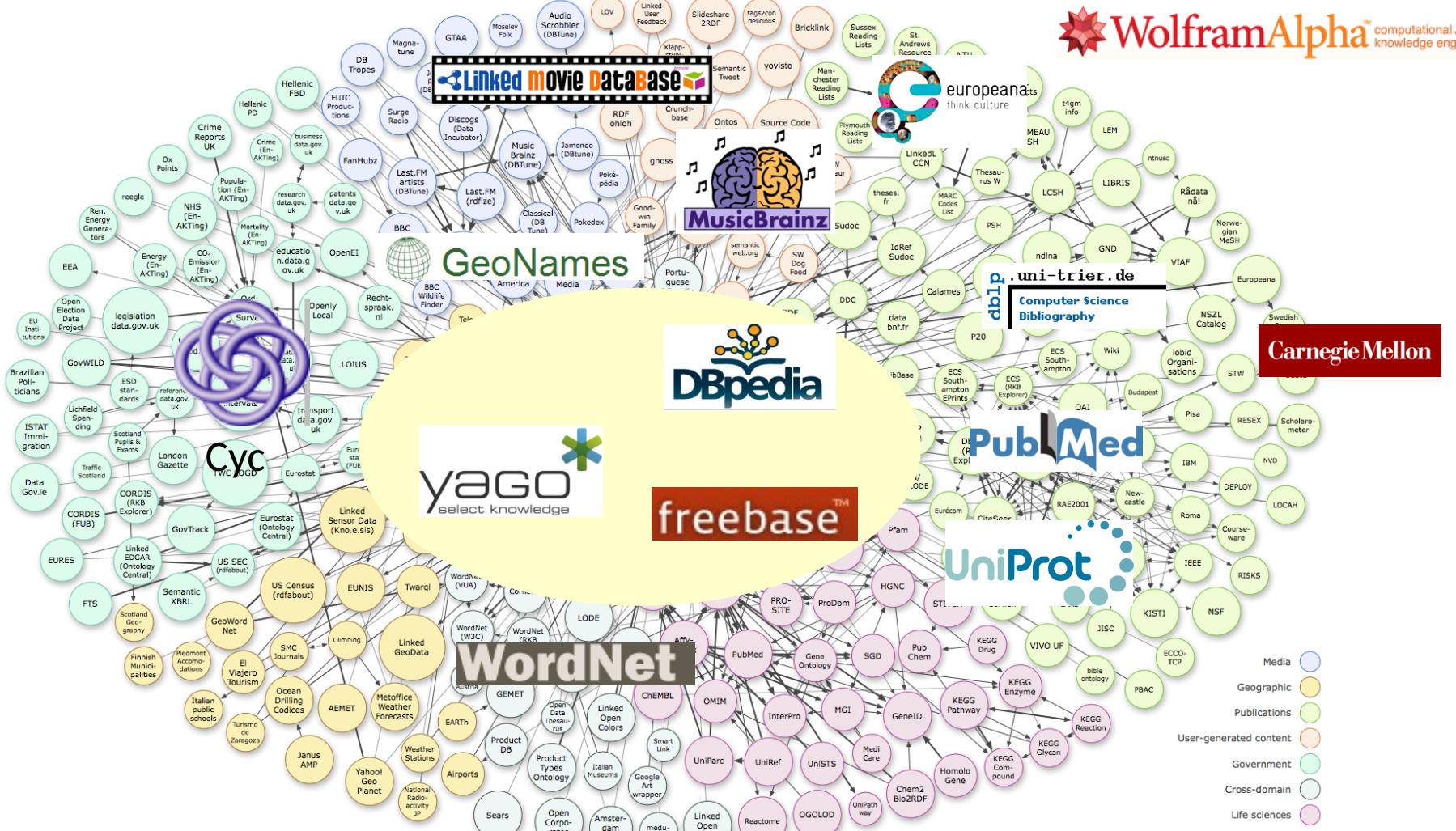
May 2012

The Knowledge Graph

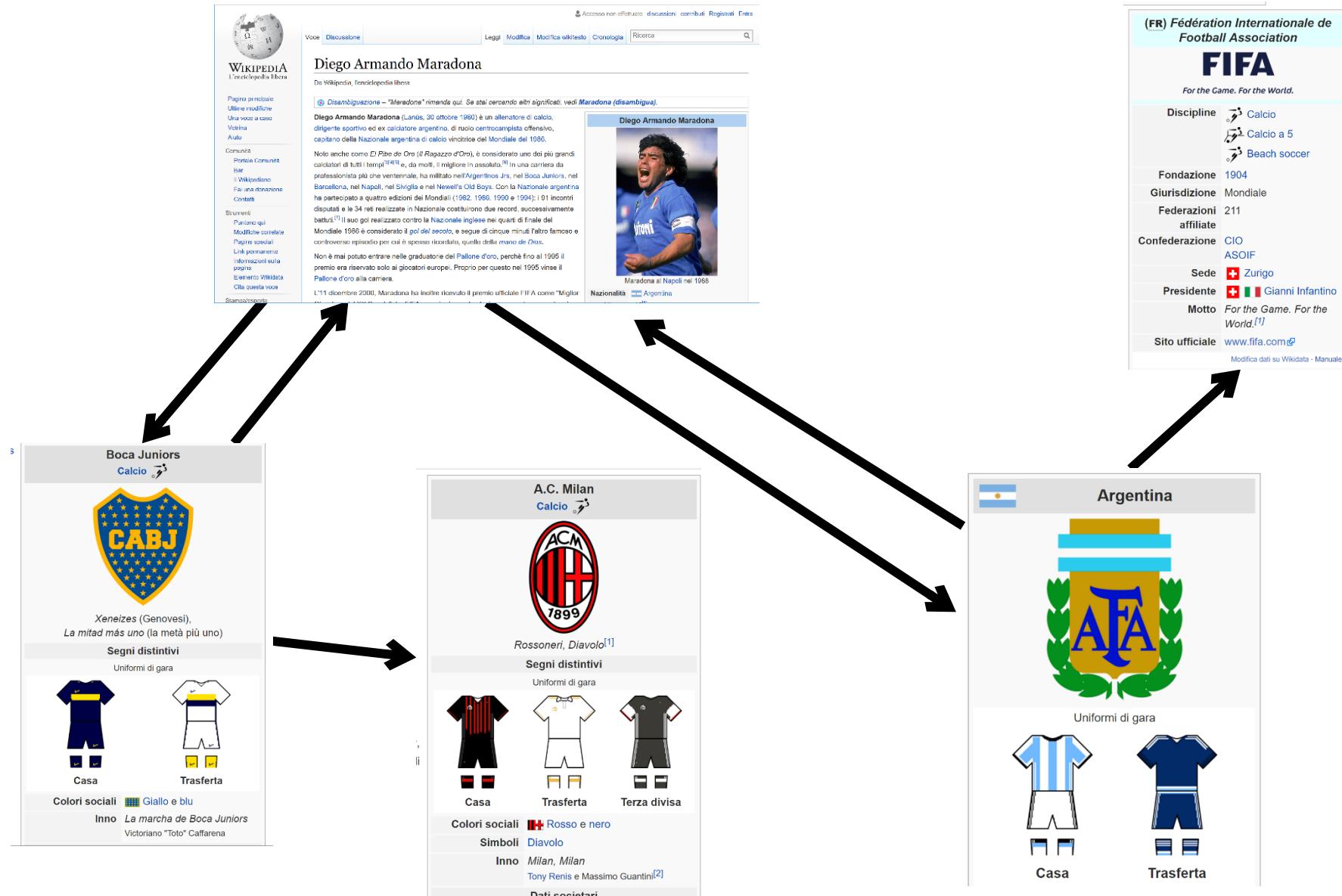
Learn more about one of the key breakthroughs behind the future of search.

See it in action

Discover answers to questions you never thought to ask, powered by the Knowledge Graph.



The Wikipedia Graph





Steve Jobs

From Wikipedia, the free encyclopedia

For the biography, see Steve Jobs (book).

Steven Paul "Steve" Jobs (/dʒɒbz/; February 24, 1955 – October 5, 2011^{[5][6]}) was an Arab-American^[7] entrepreneur^[8] and inventor,^[9] who was the co-founder, chairman, and CEO of Apple Inc. Through Apple, he was widely recognized as a charismatic pioneer of the personal computer revolution^[10] and for his influential career in the computer and consumer electronics fields, transforming "one industry after another, from computers and smartphones to music and movies...".^[12] Jobs also co-founded and served as chief executive of Pixar Animation Studios; he became a member of the board of directors of The Walt Disney Company in 2006, when Disney acquired Pixar. Jobs was among the first to see the commercial potential of Xerox PARC mouse-driven graphical user interface, which led to the creation of the Apple Lisa and, one year later, the Macintosh. He also played a role in introducing the LaserWriter, one of the first widely available laser printers, to the market.^[13]

After a power struggle with the board of directors in 1985, Jobs left Apple and founded NeXT, a computer platform development company specializing in the higher-education and business markets. In 1986, he acquired the computer graphics division of Lucasfilm, which was spun off as Pixar.^[14] He was credited in *Toy Story* (1995) as an executive producer. He served as CEO and majority shareholder until Disney's purchase of Pixar in 2006.^[15] In 1996, after Apple had failed to deliver its operating system, Copland, Gil Amelio turned to NeXT Computer, and the NeXTSTEP platform became the foundation for the Mac OS X.^[16] Jobs returned to Apple as an advisor, and took control of the company as an interim CEO. Jobs brought Apple from near bankruptcy to profitability by 1998.^{[17][18][19]}



PARC (company)

From Wikipedia, the free encyclopedia
(Redirected from PARC User Interface)

Steve Jobs

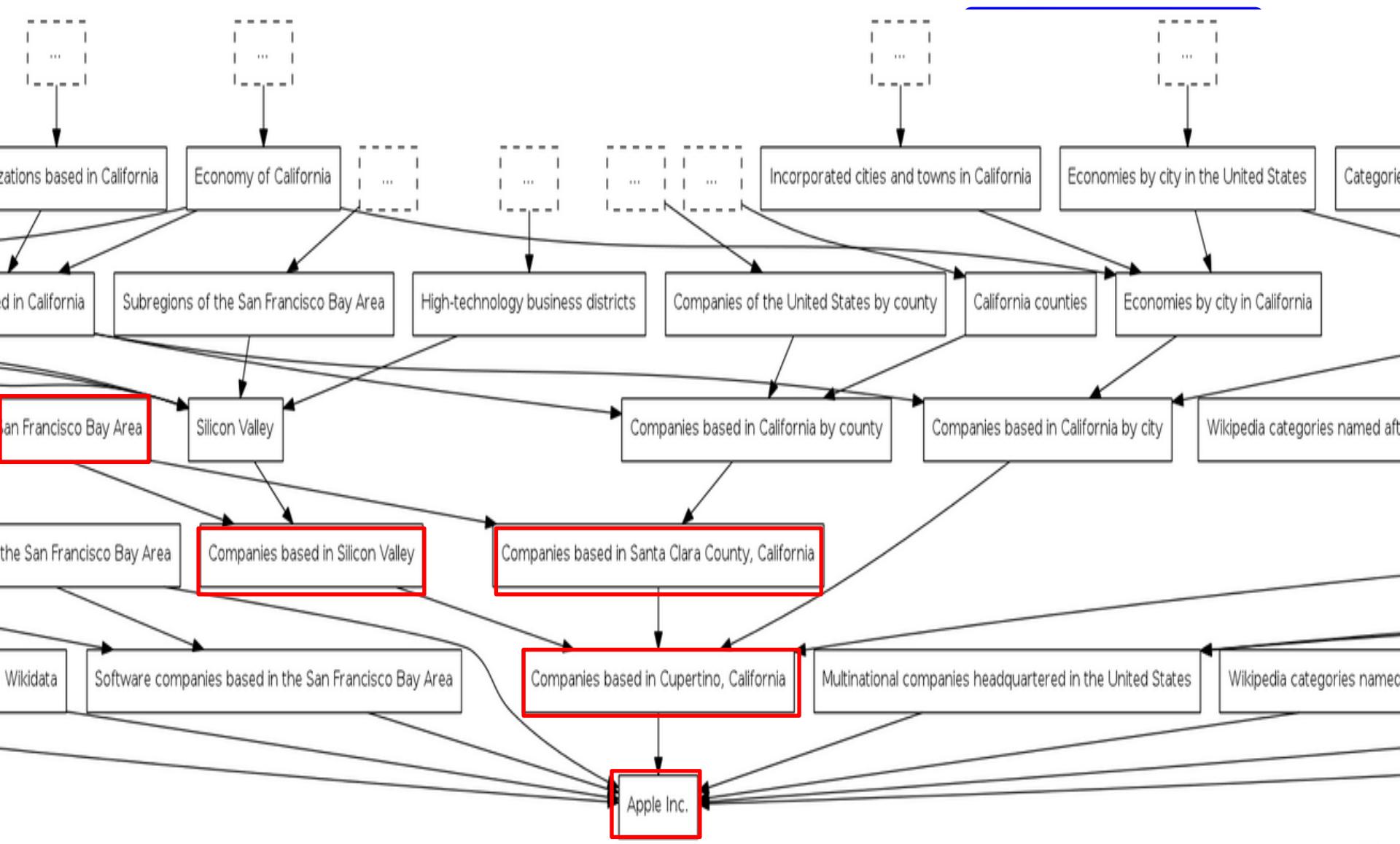


Jobs holding a white iPhone 4 at Worldwide Developers Conference 2010

Born Steven Paul Jobs

Categories: Steve Jobs | 1955 births | 2011 deaths | American adoptees | American billionaires
| American chief executives | American computer businesspeople | American industrial designers
| American inventors | American people of German descent | American people of Swiss descent
| American people of Syrian descent | American technology company founders | American Zen Buddhists
| Apple Inc. | Apple Inc. employees | Businesspeople from California | Businesspeople in software
| Cancer deaths in California | Computer designers | Computer pioneers | Deaths from pancreatic cancer
| Disney people | Internet pioneers | National Medal of Technology recipients | NeXT
| Organ transplant recipients | People from the San Francisco Bay Area | Pescetarians
| Reed College alumni

Categories form a taxonomic DAG (not really...)



Topic-based annotation

“Diego Maradona” won against “Mexico”

Ex-Argentina’s player

Mexico’s football team

Find anchors and annotate them
with articles drawn from Wikipedia!



Celebrity is a person who is famously recognized ...

the paparazzi photographed the star

the astronomer photographed the star



Star is a massive, luminous ball of plasma ...



Internet Explorer (IE o MSIE), oggi noto anche con il nome Windows Internet Explorer (WIE), è un browser web grafico proprietario sviluppato da Microsoft ...

He is using Microsoft's browser

She plays with Internet Explorer

“Obama asks Iran for RQ-170 sentinel drone back”

W Barack Obama

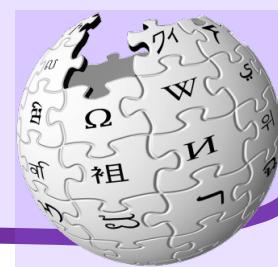
W Iran

W Lockheed Martin RQ-170 Sentinel

W President of
the
United
States

W Mahmoud Ahmadinejad

Query = USA president Ahmadinejad



“Diego Maradona won against Mexico”

- Ex-Argentina's coach
- His nephew
- Maradona Stadium
- Maradona Movie
- ...

Mexico

- ~~Mexico nation~~
- ~~Mexico state~~
- Mexico football team
- ~~Mexico baseball team~~
- ...

Don't annotate!

Commonness of a page
p wrt an anchor a

$$\Pr(p | a) = \frac{\# a \text{ linked to } p}{\# a \text{ as anchor}}$$

Context of a around the mention

$$T = \dots w_1 w_2 w_3 \underline{a} w_4 w_5 w_6 \dots$$

and the content of a *page/entity*

$$Page p = z_1 z_2 z_3 z_4 z_5 z_6 \dots$$

Link probability of an anchor a

$$lp(a) = \frac{\text{freq of } a \text{ as anchor}}{\text{freq of } a \text{ in the text}}$$

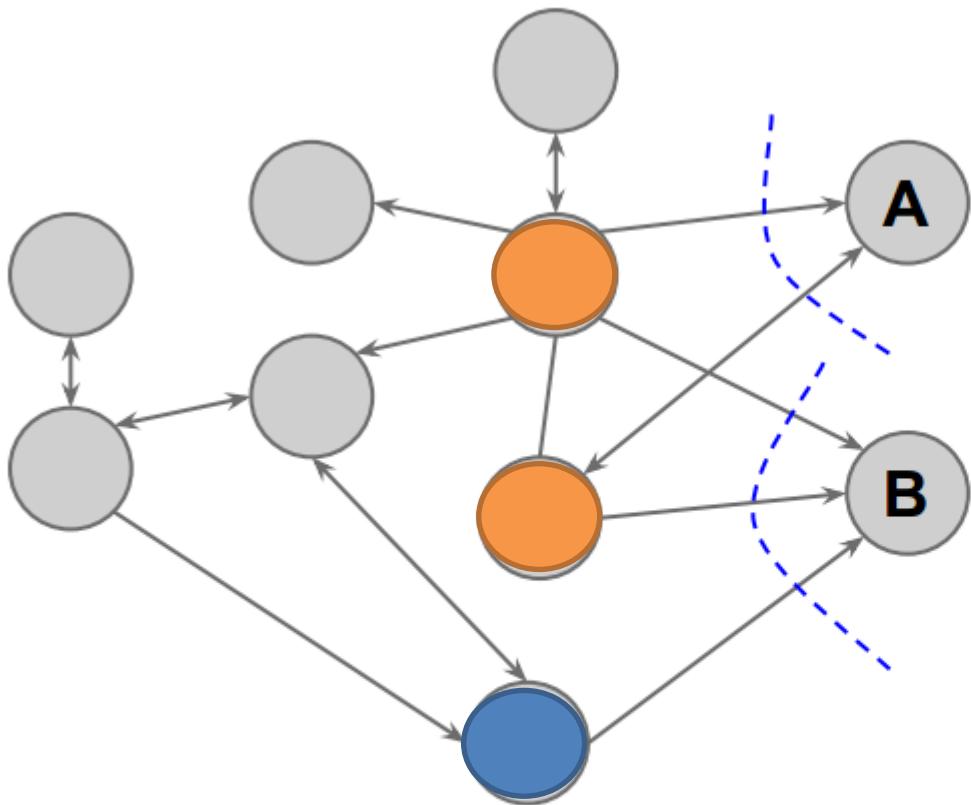
Graph-based features

a is a mention-node

p is an entity-node

Links $a \rightarrow p$ and paths
between pages





$$sr(a, b) = \frac{\log(\max(|A|, |B|)) - \log(|A \cap B|)}{\log(|W|) - \log(\min(|A|, |B|))}$$

“Diego Maradona won against Mexico”

- Diego A. Maradona
- ~~Diego Maradona jr.~~
- ~~Maradona Stadium~~
- ~~Maradona Film~~
- ...

• ...
• ...

- ~~Mexico nation~~
- ~~Mexico state~~
- Mexico football team
- ~~Mexico baseball team~~
- ...

No Annotation

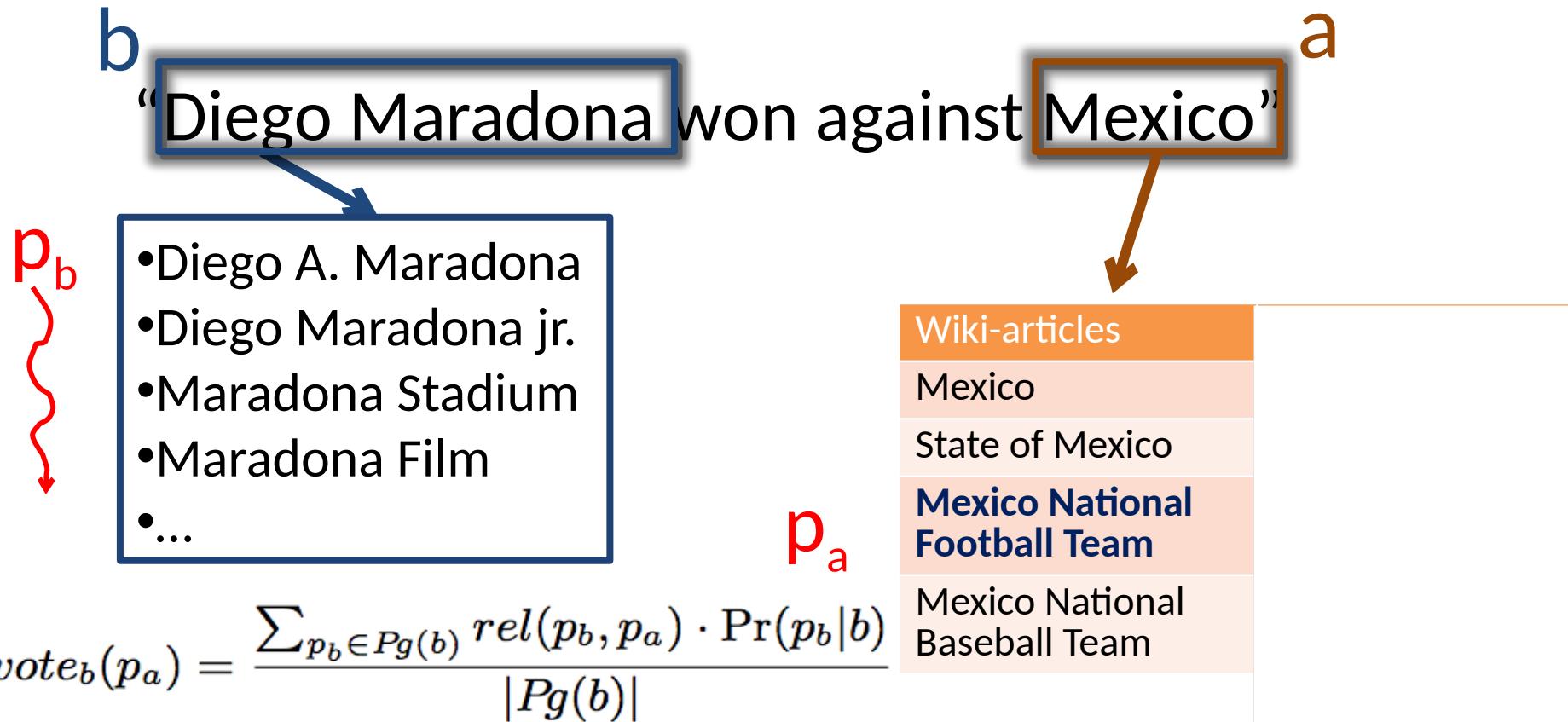
PARSING

DISAMBIGUATION
by a voting scheme

PRUNING
2 simple features

Disambiguation: The voting scheme

Collective agreement among topics via voting



Disambiguation: all steps

τ = trade-off
speed vs. recall

**Pruning by
commonness $< \tau$**

Commonness of a page
p wrt an anchor a

$$\Pr(p | a) = \frac{\# a \text{ linked to } p}{\# a \text{ as anchor}}$$

Voting scheme

Select top- ϵ pages

**Select the best in
commonness**

$$\begin{aligned} t &= 2\% \\ \epsilon &= 30\% \end{aligned}$$



TAGME is a powerful tool that is able to identify *on-the-fly* meaningful short-phrases (called "spots") in an unstructured text and link them to a pertinent [Wikipedia page](#) in a fast and effective way. This annotation process has implications which go far beyond the enrichment of the text with explanatory links because it concerns with the *contextualization* and, in some way, the *understanding* of the text.

Try **TAGME** now!

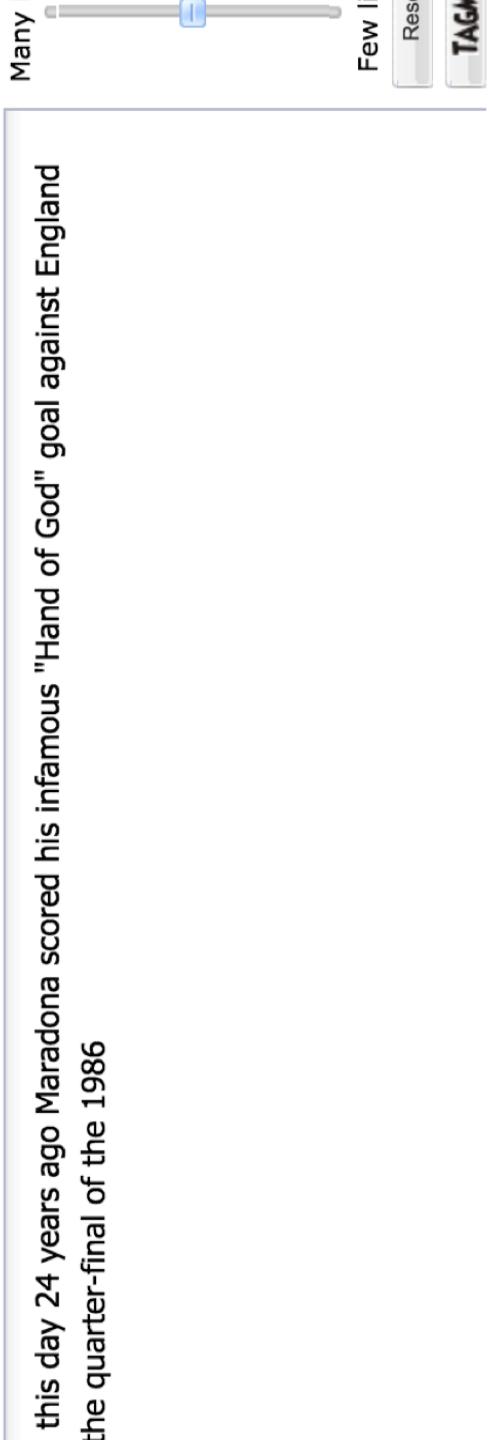
can play with the demo interface below or check the [TAGME RESTful API](#) we are currently supporting. Currently **TAGME** is available in English, German and in Italian and it is based on Wikipedia snapshots of April 6.

NEWS! TAGME is now hosted by the D4Science infrastructure. Check the [RESTful API page](#) for details.

Developed by [Paolo Ferragina](#) and Ugo Scaiella at [A³ Lab Dipartimento di Informatica, University of Pisa](#).

Text

Italiano English Deutsche



Tagged text

Topics

On this day 24 years ago Maradona scored his infamous "[Hand of God](#)" goal against [England](#) in the quarter-final of the 1986

Less links

Tagged text

Topics

On this day 24 years ago [Maradona](#) scored his infamous "[Hand of God](#)" goal against [England](#) in the [quarter-final](#) of the 1986

More links

```
    {
      "timestamp": "2013-07-12T09:11:27",
      "time": 0,
      "api": "tag",
      "annotations": [
        {
          "id": 8485,
          "title": "Diego Maradona",
          "start": 0,
          "rho": "0.38134",
          "end": 8,
          "spot": "maradona"
        },
        {
          "id": 2326217,
          "title": "Football at the 1996 Summer Olympics",
          "start": 9,
          "rho": "0.29498",
          "end": 12,
          "spot": "won"
        },
        {
          "id": 808402,
          "title": "Mexico national football team",
          "start": 21,
          "rho": "0.52456",
          "end": 27,
          "spot": "mexico"
        }
      ],
      "lang": "en"
    }
```

Useful for relating topics

NLP ?



Service Overview Demo TagMe API WAT Api SWAT Api

The Entity Linking tools by Acube lab

Welcome to the Tagme Virtual Research Environment. From here, you can access all Entity Linking tools provided by the [Acube lab](#) at the University of Pisa.



Entity linker, ideal for annotating noisy text.

Available languages: [en](#), [de](#), [it](#)



Entity linker, ideal for annotating well-formed text.

More accurate than TagMe, but still experimental.

Available languages: [en](#)



Entity linker for web search queries.

Available languages: [en](#)



Entity Salience service: assigns a relevance score to the entities mentioned by a document.

Available languages: [en](#)

Credits

Current and former members of this lab, who contributed to the development and deployment of these services, include **Paolo Ferragina**, **Marco Cornolti**, **Francesco Piccinno**, **Marco Ponza**, **Ugo Scaiella**, **Daniele Vitale**.

Powered by D4SCIENCE INFRASTRUCTURE

Access TagMe VRE

Access the TagMe VRE with your [SoBigData](#) or [D4Science](#) Gateway credentials.

[access the VRE](#)

Create an account

If you don't have an account you should first create one on the [SoBigData](#) Infrastructure Gateway to access the VRE.

[create account](#)

A glimpse on current “users”



tiscali: Cerved®



Carnegie
Mellon
University

RubyStar

The Alexa Prize

Over \$3.5 Million to Advance Conversational Artificial Intelligence

December 2017 - November 2018

