Knowledge Extraction in Text

Marco Ponza

Who am I? Last-year PhD student...

Advanced Algorithms & Applications Lab

supervisor: Prof. Paolo Ferragina

A3 Lab

RESEARCH

1. Data Compression
2. Searching & Mining Web & Social Media
3. Natural Language Understanding

Natural Language Understanding

Easy for humans

Hard for machines

But, machines need today to access, read and understand information stored in very large data archives

...and this will get to be more and more crucial with Conversational AI systems!

Natural Language Understanding

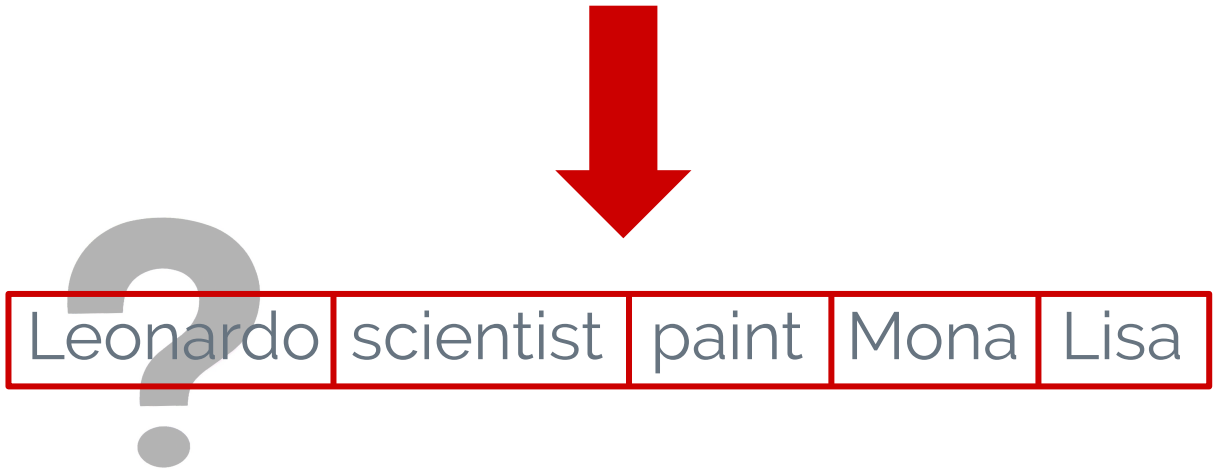
- Machines represent texts by their *(possibly, ambiguous) words*

Leonardo is the scientist who painted Mona Lisa

Natural Language Understanding

- Machines represent texts by their *(possibly, ambiguous) words*

Leonardo is the scientist who painted Mona Lisa



Natural Language Understanding

- Machines represent texts by their *(possibly, ambiguous) words*

Leonardo is the scientist who painted Mona Lisa

Leonardo DiCaprio

Leonardo (Town)

Leonardo (Ninja Turtle)

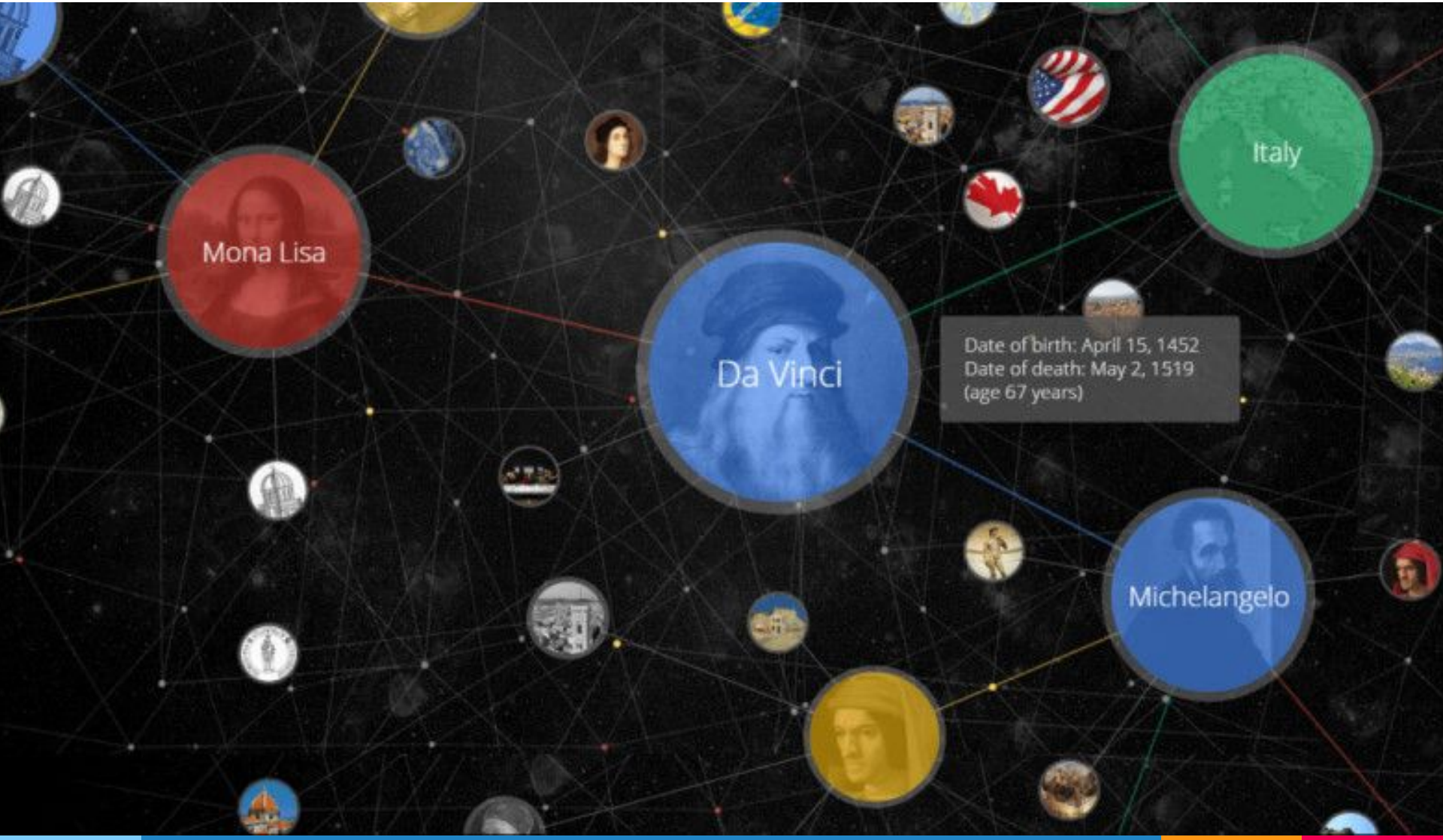
Leonardo	scientist	paint	Mona	Lisa

Leonardo da Vinci

Google Knowledge Graph

May 2012

https://www.blog.google/products/search/introducing-knowledge-graph-things-not



Google leonardo da vinci

All Images Videos News Books More Settings Tools

About 156,000,000 results (0.68 seconds)

Leonardo da Vinci - Wikipedia
https://en.wikipedia.org/wiki/Leonardo_da_Vinci
Leonardo di ser Piero da Vinci more commonly **Leonardo da Vinci** or simply Leonardo, was an Italian polymath of the Renaissance whose areas of interest ...

Works: Mona Lisa; The Last Supper; Salvator ... **Known for:** Art, science
Movement: High Renaissance

List of works by Leonardo da ... · Science and inventions · Portal:Leonardo da Vinci

Leonardo da Vinci - Wikipedia
https://it.wikipedia.org/wiki/Leonardo_da_Vinci ▾ Translate this page
Leonardo di ser Piero da Vinci (Anchiano, 15 aprile 1452 – Amboise, 2 maggio 1519) è stato un inventore, artista e scienziato italiano. Uomo d'ingegno e ...
Codici di Leonardo da Vinci · Vita personale di Leonardo da ... · Uomo vitruviano

Top stories

Leonardo da Vinci's Geological Secrets Revealed By Forensic Science
Forbes · 1 day ago

I miti da sfatare su Leonardo da Vinci
Wired · 2 days ago

Leonardo da Vinci masterpieces to be displayed in Derby
Derbyshire Live · 2 days ago

More for leonardo da vinci

People also ask

What was Leonardo da Vinci's greatest achievement?
What is so great about da Vinci?
Who was Leonardo da Vinci apprenticed to?
Why was Leonardo da Vinci famous?

Feedback

Leonardo da Vinci - Paintings, Drawings, Quotes, Facts, & Biography
https://www.leonardodavinci.net/ ▾
Leonardo da Vinci was a true genius who graced this world with his presence from April 15, 1452 to May 2, 1519. He is among the most influential artists in ...

Leonardo da Vinci

Polymath

Leonardo di ser Piero da Vinci, more commonly Leonardo da Vinci or simply Leonardo, was an Italian polymath of the Renaissance whose areas of interest included invention, drawing, painting, sculpting, ... Wikipedia

Born: April 15, 1452, Anchiano

Died: May 2, 1519, Château du Clos Lucé, Amboise, France

On view: Louvre Museum, Royal Collection Trust, Uffizi Gallery, MORE

Periods: High Renaissance, Early renaissance, Renaissance, Italian Renaissance, Florentine painting

Known for: Art, science

Buried: Chapel of Saint-Hubert, Amboise, France

Quotes View 7+ more

Simplicity is the ultimate sophistication.

Art is never finished, only abandoned.

When once you have tasted flight, you will forever walk the earth with your eyes turned skyward, for there you have been, and there you will always long to return.

Artworks View 15+ more

Salvator Mundi
1500

Mona Lisa
1503

The Last Supper
1495

Lady with an Ermine
1490

Vitruvian Man

People also search for View 15+ more

Nikola Tesla
Trending

Michelan...

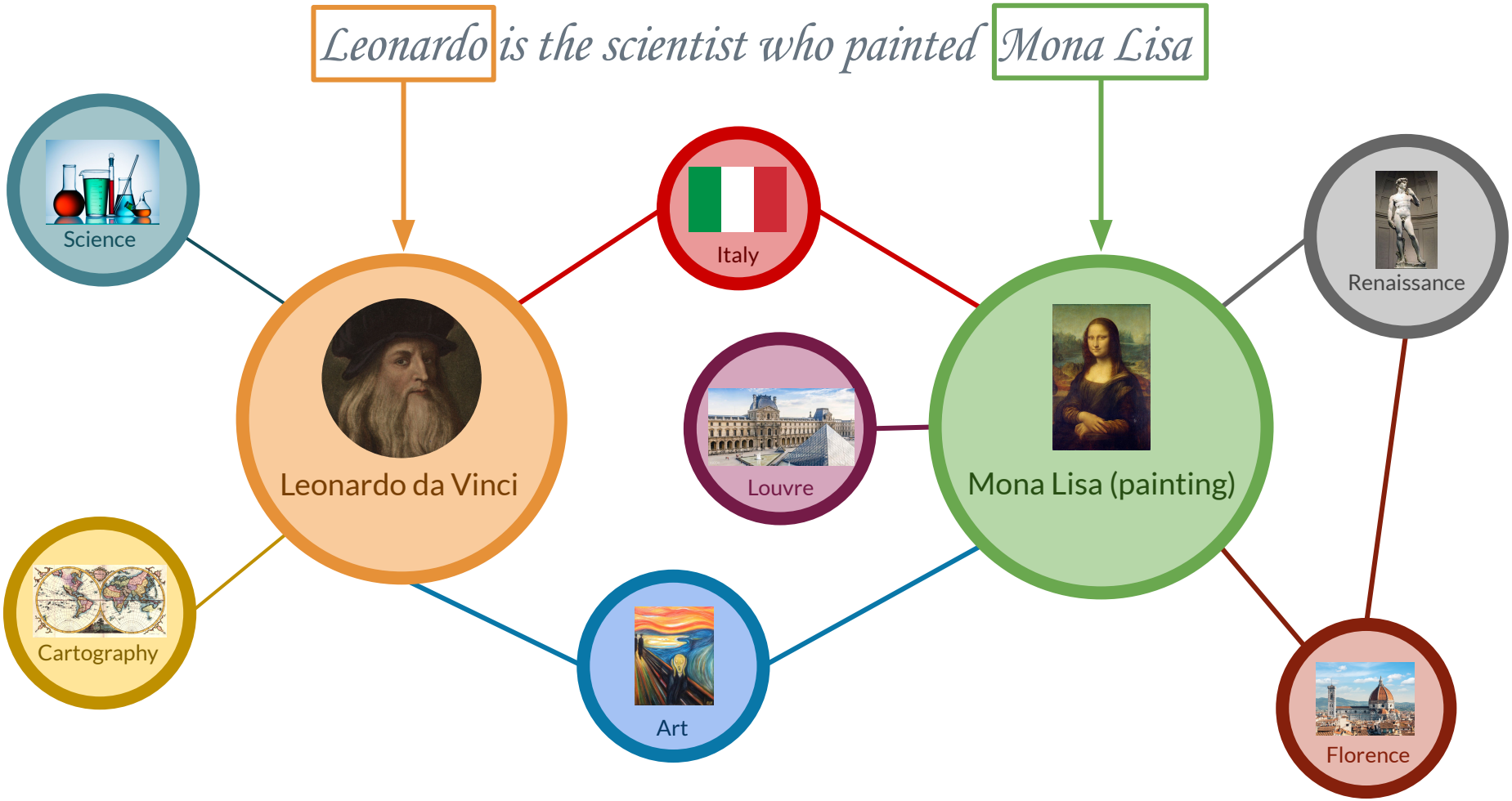
Raphael

Vincent van Gogh

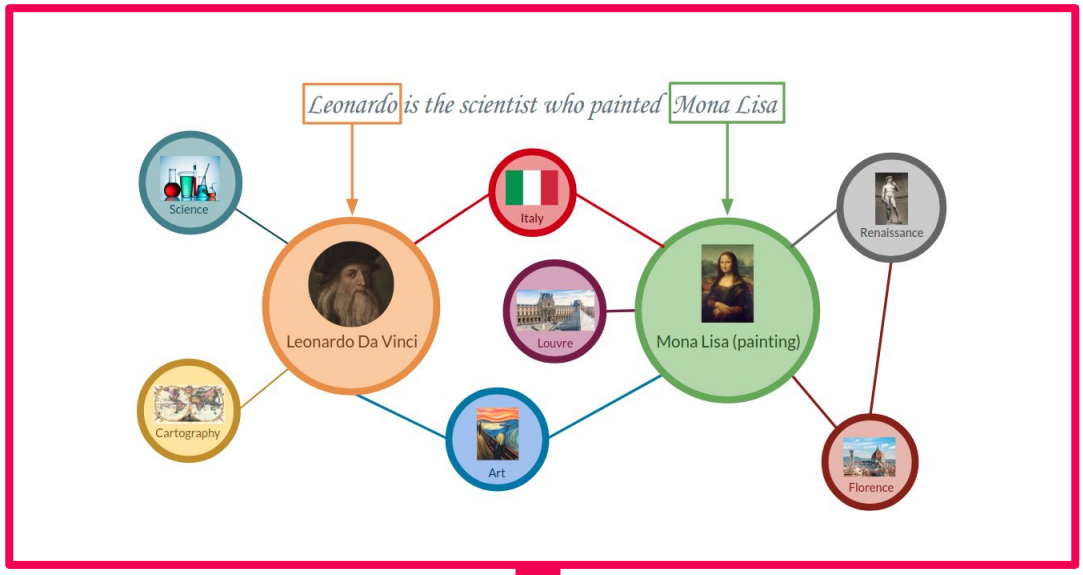
Pablo Picasso

Feedback

Understanding the Text by Entities, not Strings



Understanding the Text by Entities, not Strings



Since 2010: Efficient & effective solutions for this problem!

A3 Lab

TAG·ME

SMAPH

Google Faculty Research Awards
Two Editions: 2010 & 2013

Understanding the Text by **Salient Entities**

The New York Times

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPINION

POLITICAL ACTION; Decisions on the Horizon

By JEFF ZELENY and PATRICK HEALY
Published: January 9, 2007

FACEBOOK
TWITTER
GOOGLE+
EMAIL
SHARE
PRINT
REPRINTS

Don't look for presidential announcements from Senators Barack Obama and Hillary Rodham Clinton anytime soon, but stay tuned.

At least that is the word from their associates. Mr. Obama, Democrat of Illinois, is not likely to say whether he intends to seek the party's presidential nomination until after President Bush's State of the Union address on Jan. 23. As he walked out of the Capitol on a recent afternoon, Mr. Obama only smiled when asked about his timing. Then, he rushed to change the subject.

Initially, Mr. Obama said he intended to announce his decision after returning from a holiday vacation in Hawaii, where he was visiting his grandmother and other relatives. Now, several people close to the senator say, he needs a little more time to make up his mind.

Still, Mr. Obama has been busy telephoning crucial Democrats in Iowa, New Hampshire and other states. There is, of course, only one reason for him to be making such inquiries.

Last week on Capitol Hill, Mr. Obama bumped into Ethel Kennedy, who has been a big admirer. When asked about him, she said, "He can't run soon enough."

Mrs. Clinton, meanwhile, plans to announce her decision in the next several weeks, her advisers say. According to several Democrats who have spoken to her, as well as advisers, Mrs. Clinton has given every indication that she is running, short of saying so, and no signals that she is not.

She is making phone calls to Democratic officials, labor leaders and supporters in early nominating states. And she continues to talk to possible consultants and donors, yet she has not made any travel plans to kick off a campaign.

...and more and more entities!

Understanding the Text by **Salient Entities**

- Entity Salience Problem

  Relevant vs **Non-Relevant** Entities

...and more and more entities!

Understanding the Text by **Salient Entities**

- Entity Salience Problem

  Relevant vs **Non-Relevant** Entities

- Our Solution

  SWAT

  - Improvements of +12%

    wrt CMU/Google system

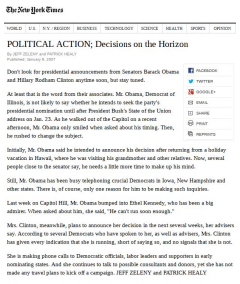
  - Published at NLdb 2017 Liège, Belgium

Bloomberg Research Grant 2017

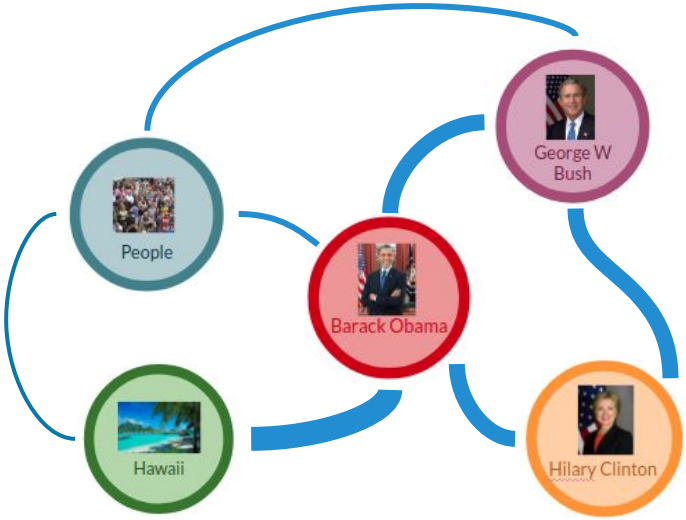
...and more and more entities!

Understanding the Text by **Salient Entities**

- How? Applying Graph Theory and Algorithms!

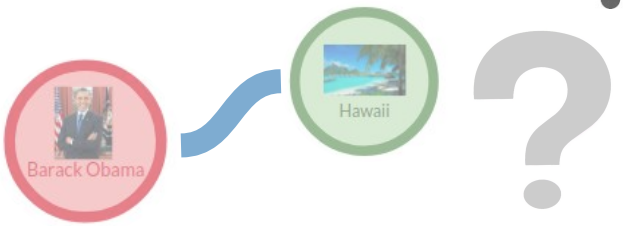


Text



1. Used to draw new features
2. Classify entities into salient/non-salient via ML

- Problem: How can we **weight** the **edges**?

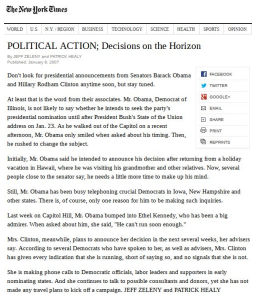


CIKM 2017 SINGAPORE

Understanding the Text by **Extraction of Facts**

- How to **enrich** the Knowledge Graph with **information** that comes from a **text**?

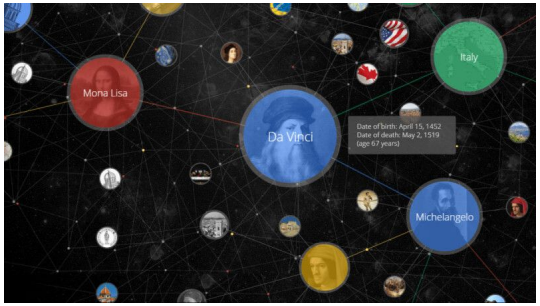
1. Identify **salient entities** ✓
2. Extract **facts** connecting them ✗



The New York Times

POLITICAL ACTION; Decisions on the Horizon

Text



Knowledge Graph

Leonardo is the scientist who painted Mona Lisa

Facts

- *(“Leonardo”, “is”, “scientist”)*
- *(“Leonardo”, “painted”, “Mona Lisa”)*

max planck institut informatik

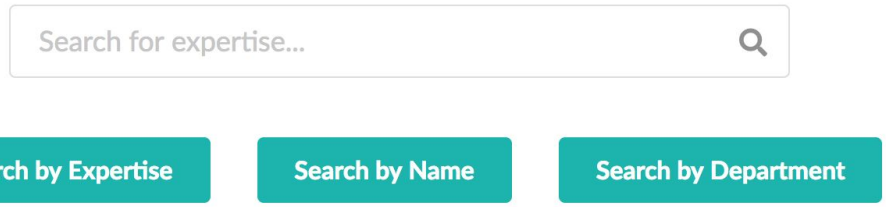
EMNLP 2018
Brussels

Expert Finding & Profiling

SoBigData Research Infrastructure

IN SUPREMÆ DIGNITATIS 1343

WISER
Wikipedia Expertise Ranking



- ~1.5K Authors
- ~65K Documents (papers' abstracts)
- ~35K Research Topics
- More than 1K queries and ~2K profiles view in few months
- Currently used by UniPi's Technology Transfer Office

WISER

Main Topics | Main Areas | Stream Graph | Tag Cloud | Publications | Survey

Select the range of years to analyze.

1998 2000 2002 2004 2006 2008 2010 2012 2014 2016 2018

Ra...	Entity	Count	Doc. count	Years	Wiser score
1	Recurrent neural network	26	15		
2	Artificial neural network	44	24		
3	Tree (data structure)	24	17		
4	Machine learning	16	14		
5	Recursive neural network	16	16		
6	Quantitative structure–activi...	30	16		
7	Echo state network	14	13		
8	Mathematical model	39	28		
9	Prediction	19	17		
10	Generative model	9	7		
11	Group representation	22	14		
12	Empiricism	15	15		
13	Data model	8	8		
14	Data set	20	14		
15	Reservoir computing	8	8		

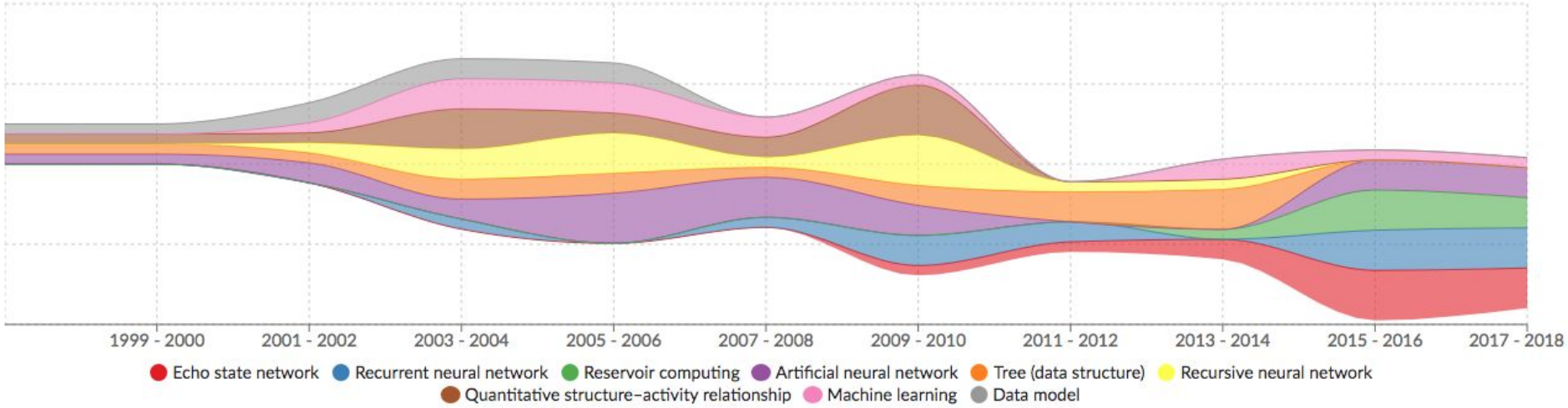
Previous | Page 1 of 24 | 15 rows | Next

WISER

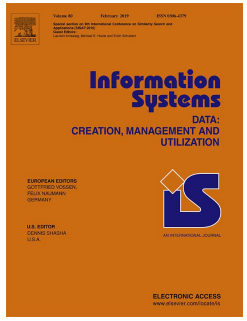
Main Topics | Main Areas | **Stream Graph** | Tag Cloud | Publications | Survey

Select range of years to examine. 6

Select how many topics for each range 4



ELSEVIER



2019

Future Directions

Moving the paradigms: from *text* to *conversations*!

SYSTEMS

http://swat.d4science.org

WISER
Wikipedia Expertise Ranking

http://wiser.d4science.org

So Big Data Research Infrastructure

https://sobigdata.d4science.org/web/tagme

Thanks!