A Software Suite for the Understanding of Natural Language

Marco Ponza Paolo Ferragina

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

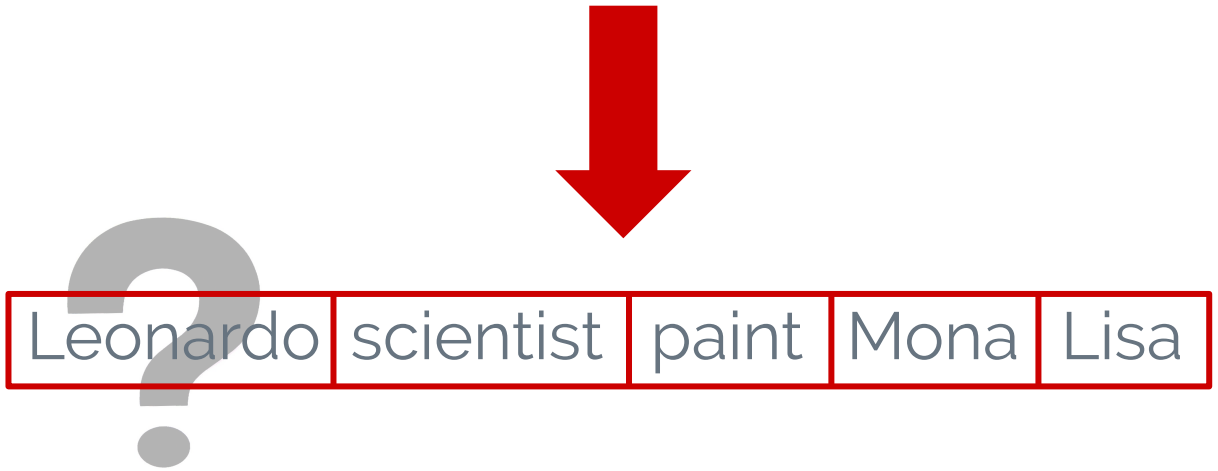
...and this will get to be more and more crucial

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

Leonardo DiCaprio

Leonardo (Town)

Leonardo (Ninja Turtle)

Leonardo da Vinci

Leonardo	scientist	paint	Mona	Lisa

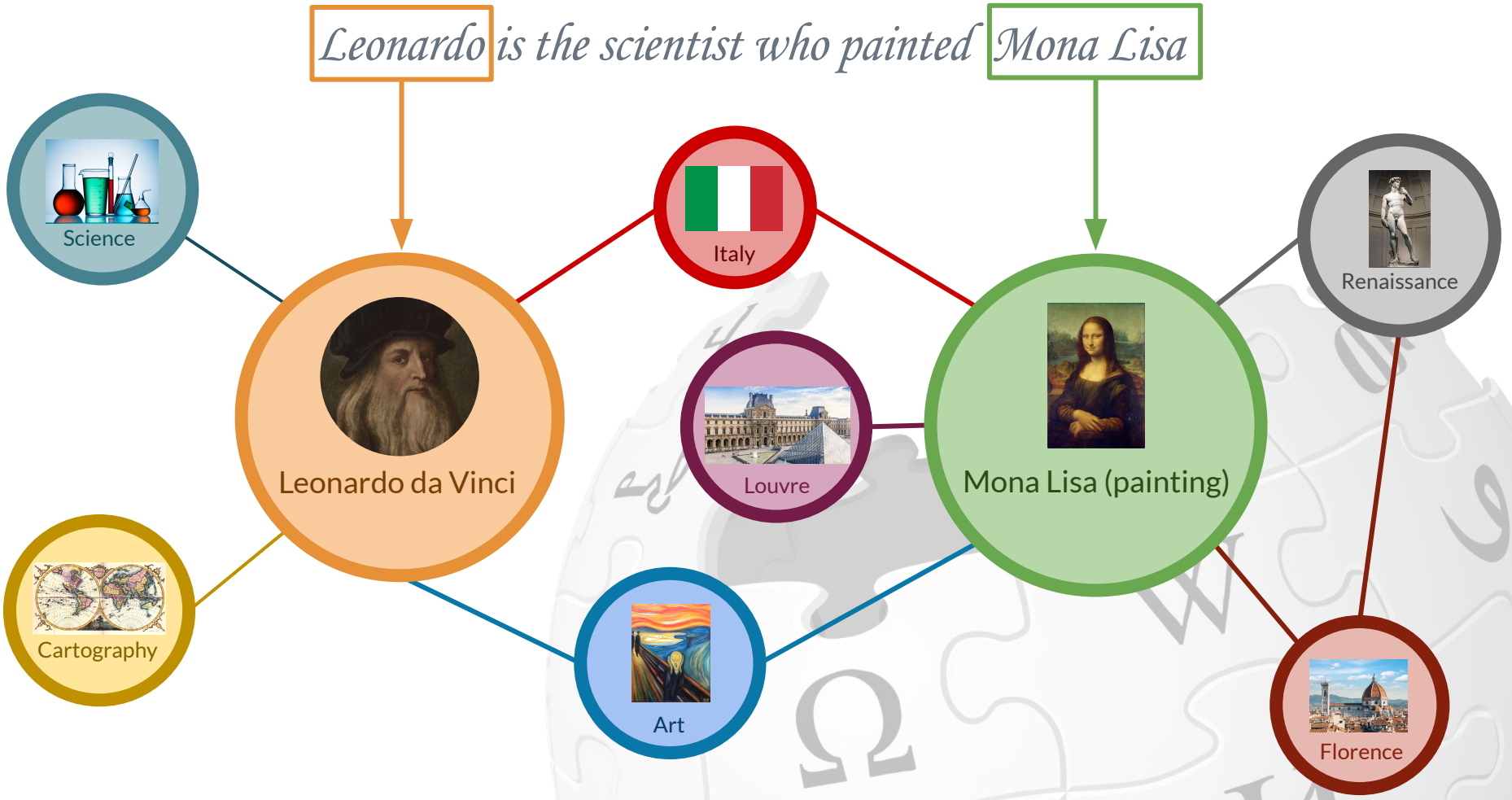
Understanding the Text by Entities, not Strings

Leonardo is the scientist who painted Mona Lisa

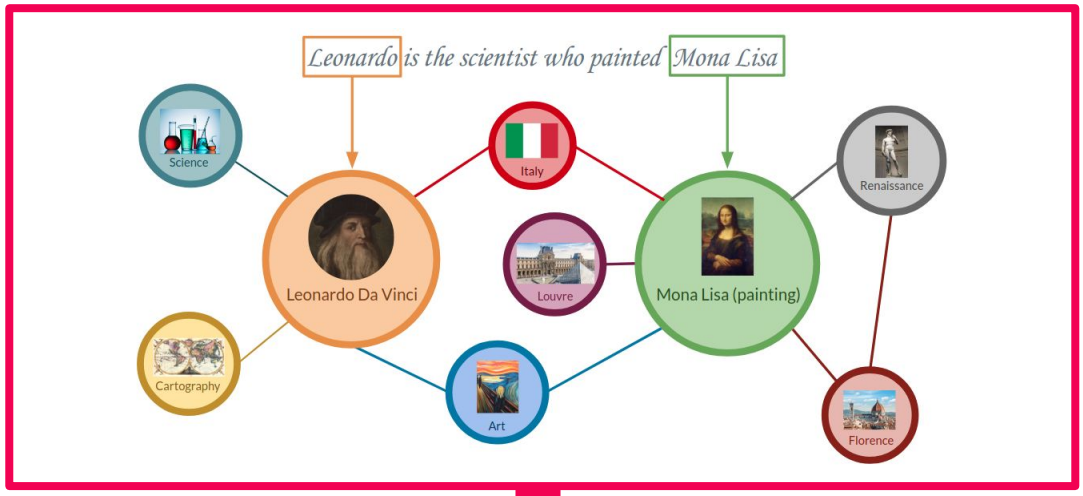
Understanding the Text by Entities, not Strings

Leonardo *is the scientist who painted* *Mona Lisa*

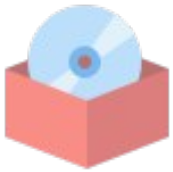
Understanding the Text by Entities, not Strings



Understanding the Text by Entities, not Strings

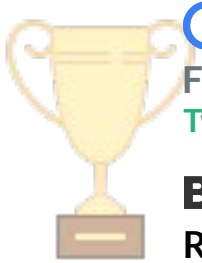


Since 2010: Efficient & effective solutions for this problem!



Software Suite for Natural Language Understanding

TAG-ME WAT SWAT SMAPH



Google Faculty Research Awards
Two Editions: 2010 & 2013

Bloomberg Research Grant 2017

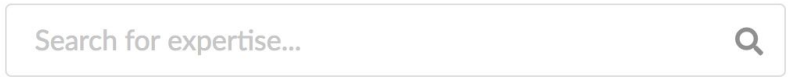
Expert Finding & Profiling

So Big Data 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

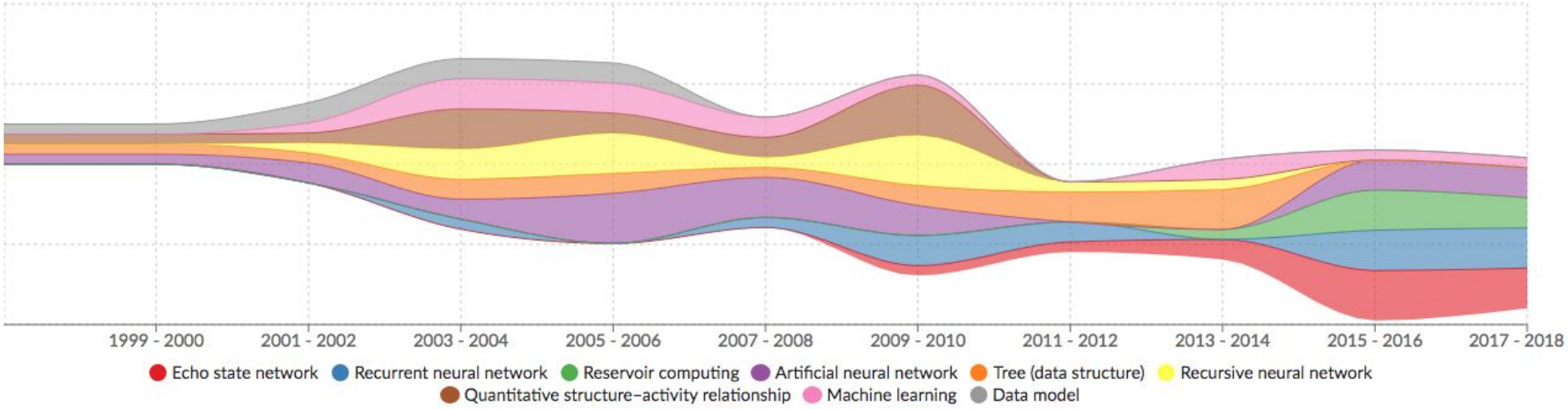
Main Topics | Main Areas | **Stream Graph** | Tag Cloud | Publications | Survey

Select range of years to examine.

6

Select how many topics for each range

4



SOFTWARE SYSTEMS

Software Suite for Natural Language Understanding

TAG·ME WAT SWAT SAMAPH

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

WISER
Wikipedia Expertise Ranking

https://wiser.d4science.org

A3Lab

http://acube.di.unipi.it