

Vers l'analyse à la demande des connaissances de Wikidata

Arnaud Soulet

Université de Tours, LIFAT, Blois



Journée Intelligence Artificielle
Et Humanités Numériques

From Wikipedia to...

The image shows a screenshot of the Wikipedia article for 'Painting'. The page title is 'Painting' and it includes a search bar at the top. The article text defines painting as a visual art involving the application of pigment to a solid surface. It mentions the Mona Lisa as a famous example and discusses the history of painting, including ancient cave art and modern techniques. There are three images: the Mona Lisa, a cave painting of a rhino, and a prehistoric cave painting of a bison.

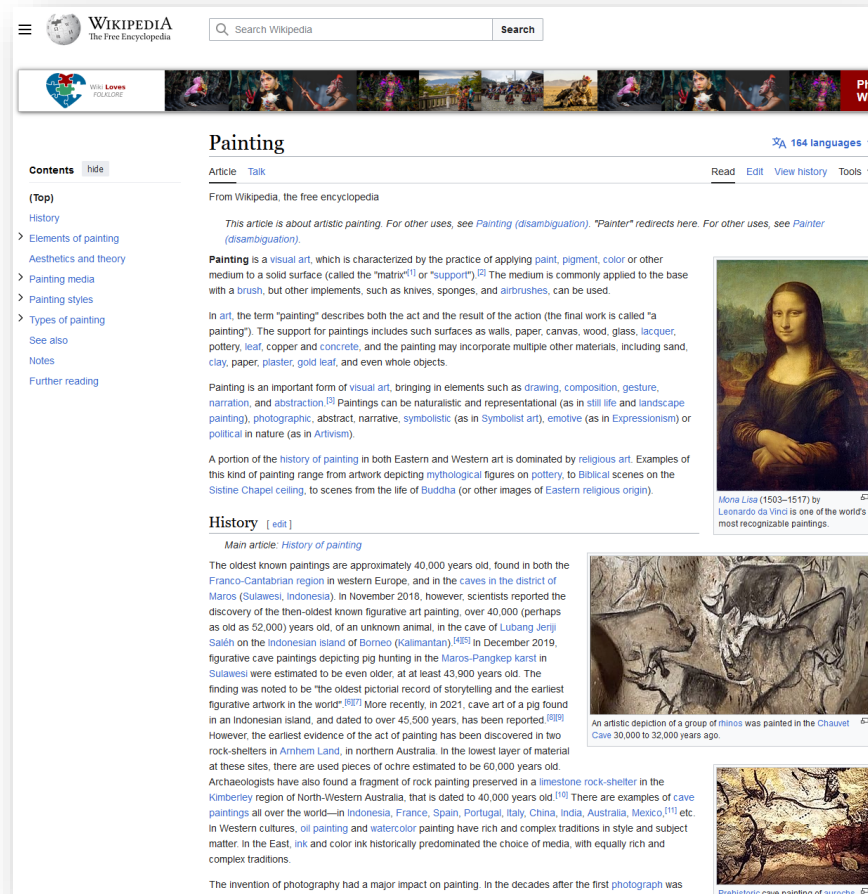
<https://en.wikipedia.org/wiki/Painting>

❑ Is painting as important an art as literature?

❑ Who is the most famous woman painter?

❑ What are the most popular painting techniques?

From Wikipedia to...



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

Content limitations

- No comparison with other types of art
- No information about famous painters

Editorial limitations

- No quantitative analysis
- No generic template

Design an analytical tool of the Web of Data for knowledge workers

Ferré, S. Analytical Queries on Vanilla RDF Graphs with a Guided Query Builder Approach. In *FQAS 2021*.
Weikum, G. Knowledge graphs 2021: A data odyssey. In *VLDB 2021*.
Lissandrini, M. & al. Knowledge graph exploration systems: are we lost. In *CIDR 2022*.

From Wikipedia to Wikidata

Vrandečić, D., & Krötzsch, M. (2014). Wikidata: a free collaborative knowledgebase. CACM.

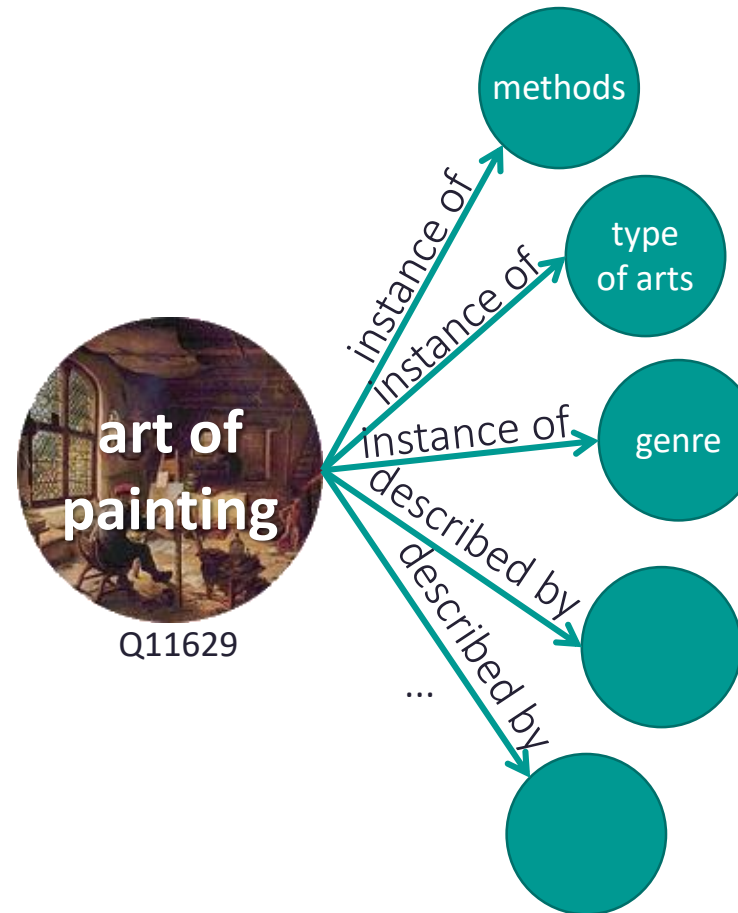
The screenshot shows the Wikipedia article for 'Painting'. The title is 'Painting' with a sub-header '164 languages'. The article text describes painting as a visual art involving paint, pigment, or color on a solid surface. It includes a section on 'History' and a main image of the Mona Lisa. The URL at the bottom is <https://en.wikipedia.org/wiki/Painting>.

The screenshot shows the Wikidata page for 'art of painting' (Q11629). It displays the item's label in multiple languages (English, French, Spanish, German) and its description: 'practice of applying paint, pigment, color or other medium to a surface'. Below this, there are sections for 'Statements' and 'subclass of', showing relationships like 'instance of' and 'subclass of' with their respective reference counts.

Wikidata
= machine-readable encyclopedia

<http://www.wikidata.org/entity/Q11629>

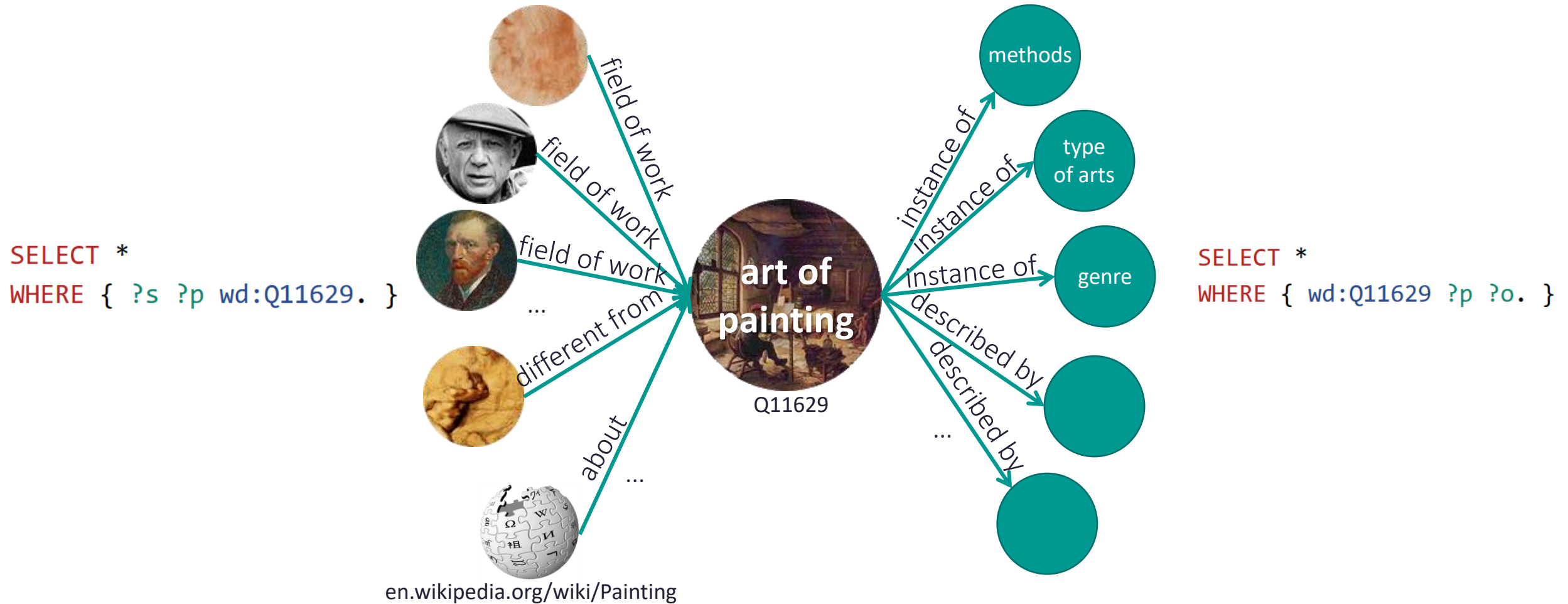
Wikidata: a crowdsourced knowledge graph



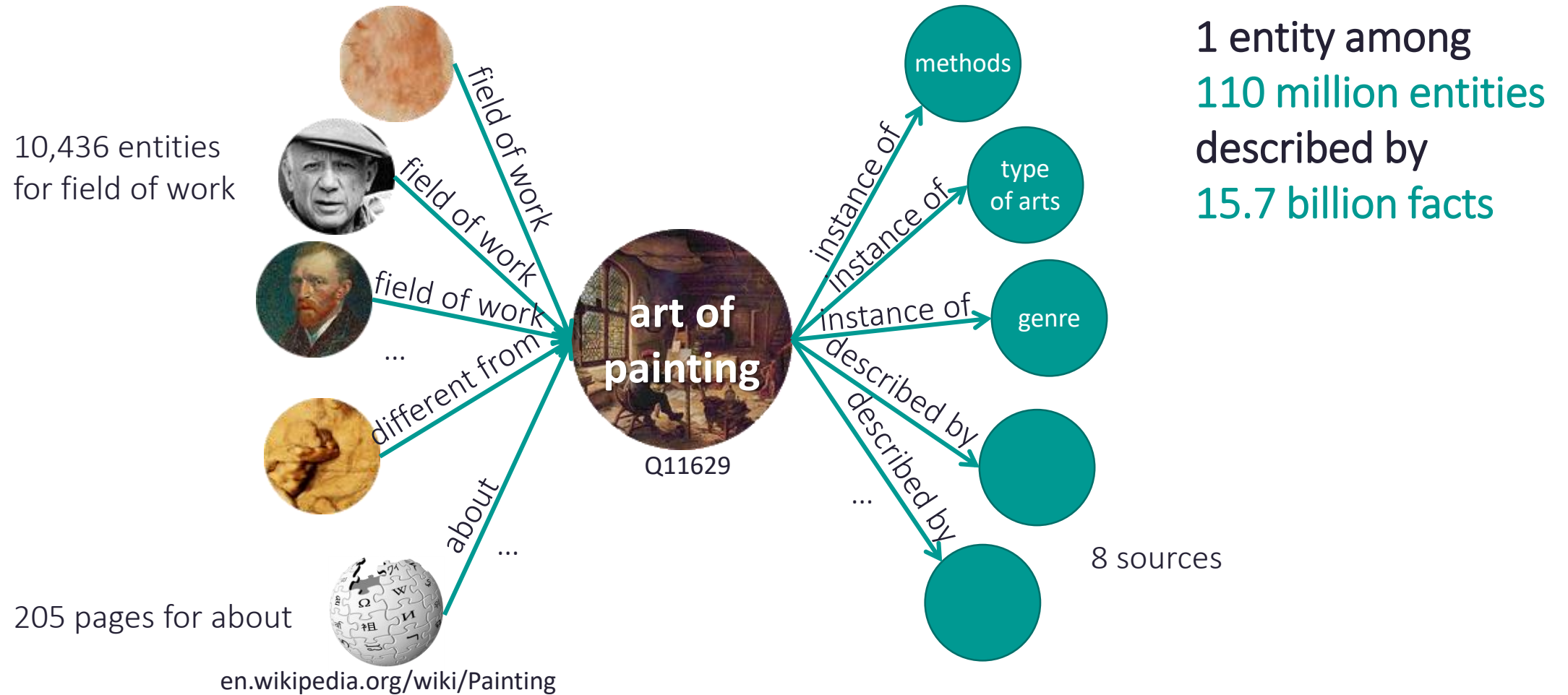
SPARQL language for querying Wikidata:

```
SELECT *  
WHERE { wd:Q11629 ?p ?o. }
```

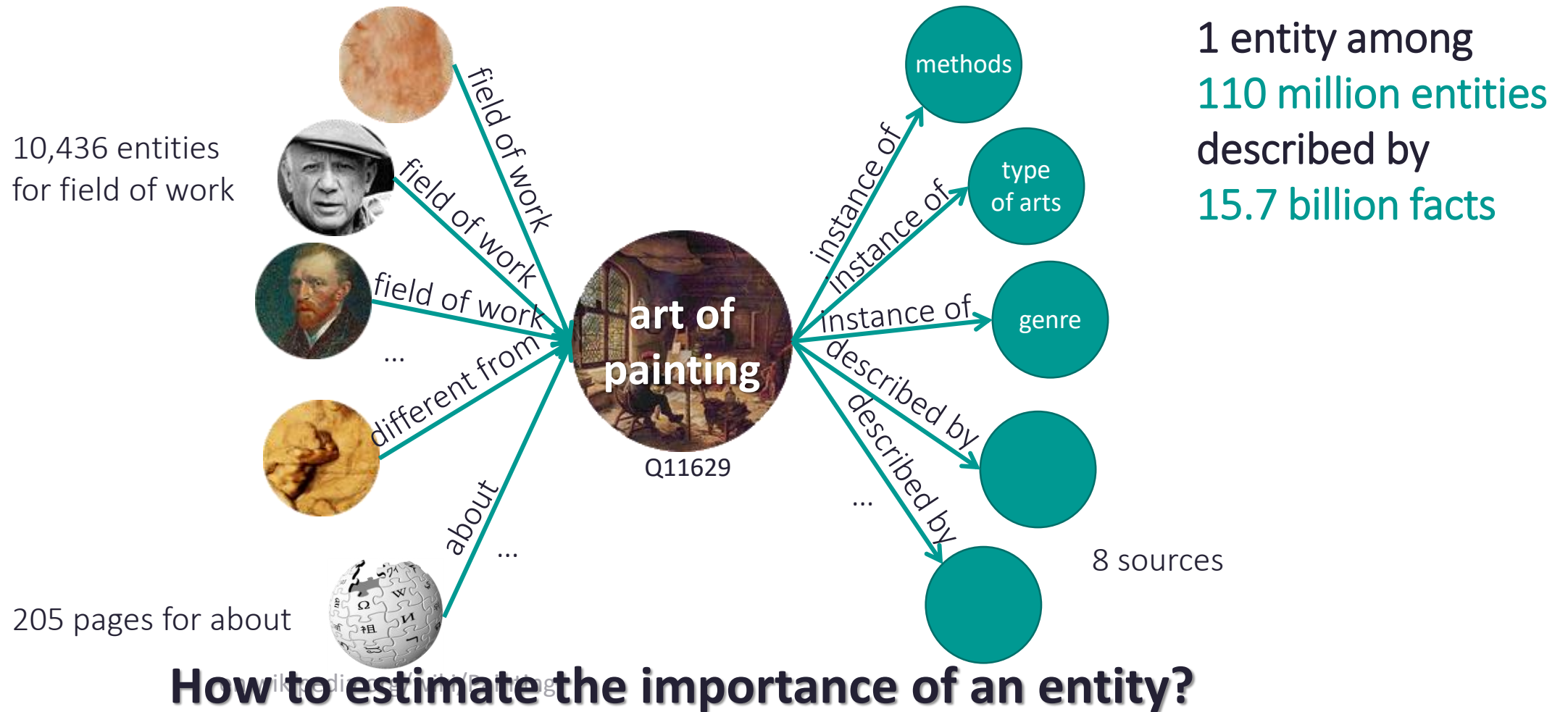
Wikidata: a crowdsourced knowledge graph



Wikidata: a crowdsourced knowledge graph



Wikidata: a crowdsourced knowledge graph



From Wikidata to Rankingdom

Item Discussion

art of painting (Q11629)

practice of applying paint, pigment, color or other medium to a surface
painting | to paint

In more languages

Language	Label	Description	Also known as
English	art of painting	practice of applying paint, pigment, color or other medium to a surface	painting to paint
French	peinture	pratique artistique	
Spanish	pintura	objeto bidimensional que parte de la práctica de aplicar pigmentos, colores u otros objetos a una superficie	pintura artística
German	Malerei	klassische Gattung der bildenden Kunst	

All entered languages

Statements

instance of

- method
0 references
- type of arts
0 references
- genre
0 references

subclass of

- activity
0 references
- visual arts
0 references

image

<http://www.wikidata.org/entity/Q11629>

rankingdom

art of painting

practice of applying paint, pigment, color or other medium to a surface

described by source Metropolitan Museum Of Art Tagging Vocabulary

Quadrant analysis

Two-Dimensional Visual Artwork

- calligraphy: visual art related to writing (114)
- graffiti: writing or drawing etched, scratched, scribbled, or sprayed (often, but not always illicitly) on a wall or other surface in a public place (90)
- mosaic: image made from an assemblage of small pieces of colored glass, stone, or other materials (88)
- manuscript
- relief sculpture
- tapestry

<https://www.rankingdom.org/entity/Q11629>

rankingdom^{beta} Quadrant Ranking

places

Blois
 commune of France city
 French commune and city in Loir-et-Cher, Centre-Val de Loire
 Disambiguation blois

capital of **County Of Blois**
 historical province in the Kingdom of France

Quadrant analysis

rankingdom^{beta} Quadrant Ranking Impact

organizations

Tours University
 university in France
 French university located in Tours, France. Founded in 1971.
 Disambiguation university of tours

instance of **University In France**
 public university in France

Quadrant analysis

rankingdom^{beta} Quadrant Ranking Impact

occupations

computer scientist
 profession
 one who studies or practises computer science
 Disambiguation computer scientist

industry **Service Sector**
 third of the three economic sectors of the three-sector theory

Scope: 380
 Influence: 43,435
 About: 35
 a-index: 56
 Impact: 43,876
 What are these metrics?

Quadrant analysis

rankingdom^{beta} Quadrant Ranking

people

Jean Jaurès
 human journalist writer reporter politician
 French / Occitan Socialist leader (1859-1914)
 Disambiguation jean jaurès

described by source **New International Encyclopedia**
 American encyclopedia first published in 1902

Quadrant analysis

Independent Social Democratic Party Of Germany

rankingdom^{beta} Quadrant Ranking Impact

concepts

knowledge
 philosophical concept mental state
 mental possession of information or skills, contributing to understanding
 Disambiguation knowledge

union of **List Of Values As Qualifiers**
 dummy value for "union of" P2737 and "disjoint union of" P2738 to specify the list of items this class can. Use "list item" P11260 as a qualifier to specify the items.

Scope: 451
 Influence: 882
 About: 176
 a-index: 28
 Impact: 3,087
 What are these metrics?

Quadrant analysis

Philosophical Concept

rankingdom^{beta} Quadrant Ranking Impact

objects

chair
 piece of furniture for sitting on
 Disambiguation chair

has characteristic **Height**
 distance between the lowest end and highest end of an object

Scope: 398
 Influence: 5,626
 About: 147
 a-index: 19
 Impact: 2,100
 What are these metrics?

Quadrant analysis

Furniture

...and more with **110 million** searchable entities

How to deal with such heterogeneous entities?

Overview

art of painting

practice of applying paint, pigment, color or other medium to a surface

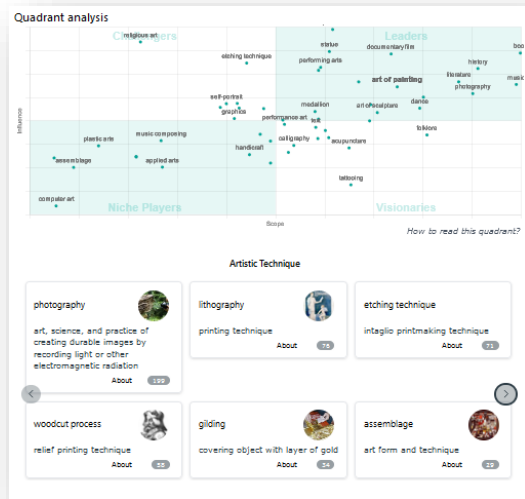
Disambiguation painter

described by source Metropolitan Museum Of Art Tagging Vocabulary

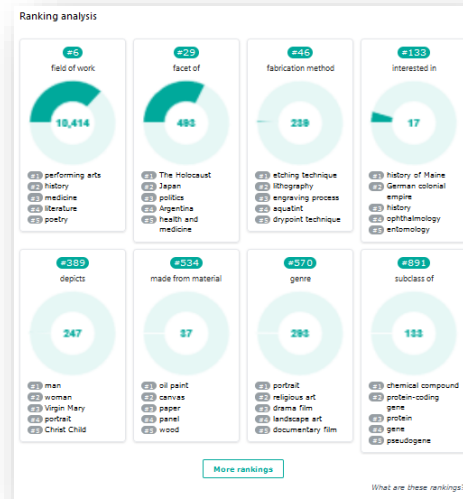
Scope: 489
Influence: 23,418
About: 759
a-index: 69
Impact: 72,129

What are these metrics? 2/4

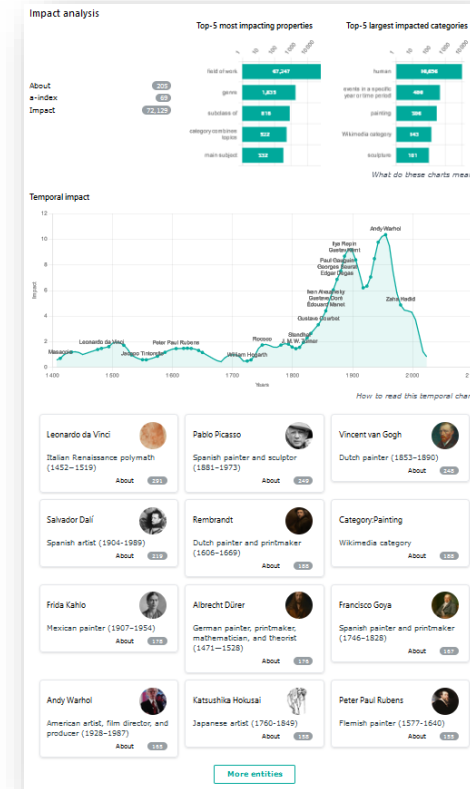
Summary



1 Quadrant analysis

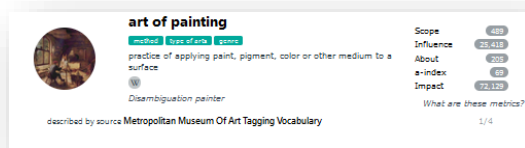


2 Ranking analysis



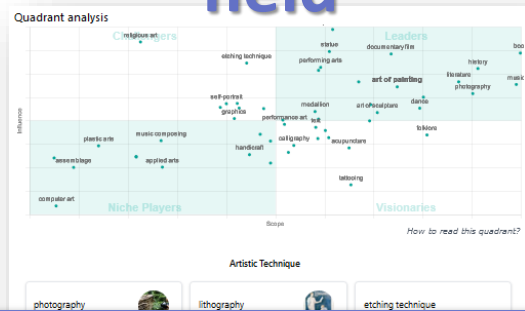
3 Impact analysis

Overview: data narrative



Summary

field



properties



Impacted entities



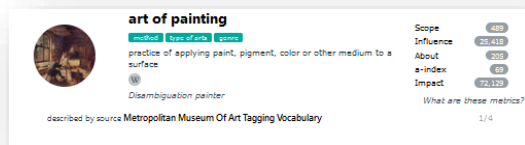
① Quadrant analysis

② Ranking analysis

③ Impact analysis

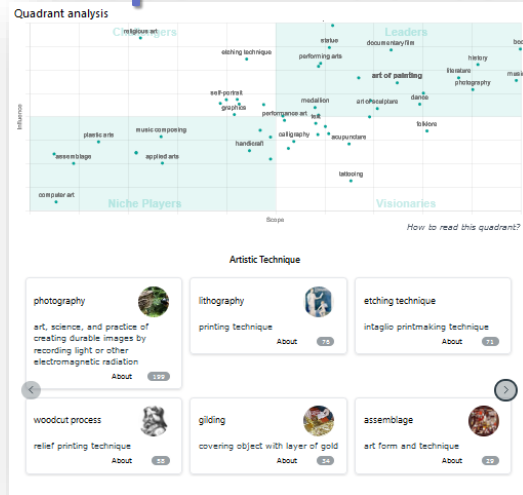
from general to specific

Overview: measure



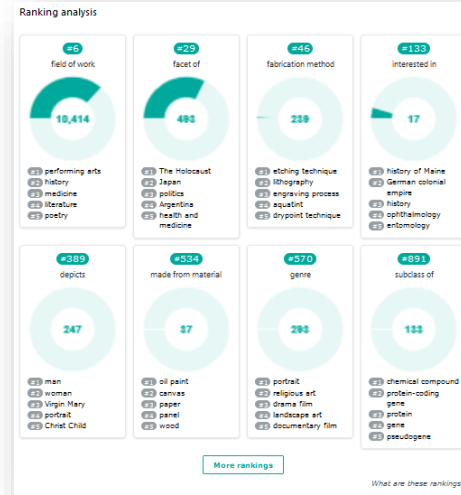
Summary

scope + influence



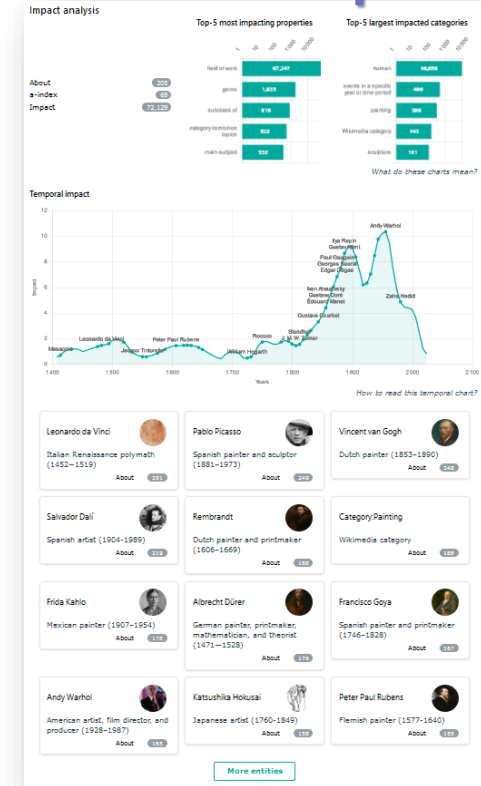
1 Quadrant analysis

semantic scores



2 Ranking analysis

about + impact

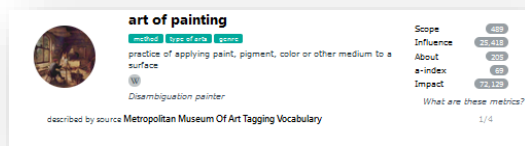


a-index

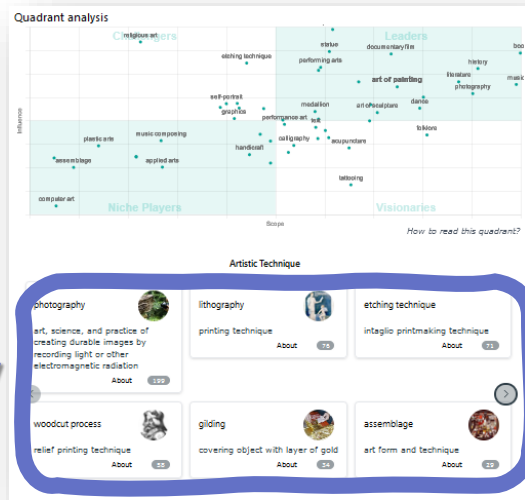
3 Impact analysis

Nacke, O. (1979). "Informetrie: eine neuer Name für eine neue Disziplin". Nachrichten für Dokumentation. 30 (6): 219–226.

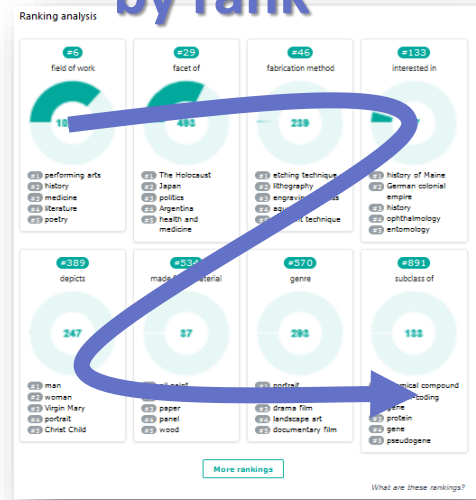
Overview: order



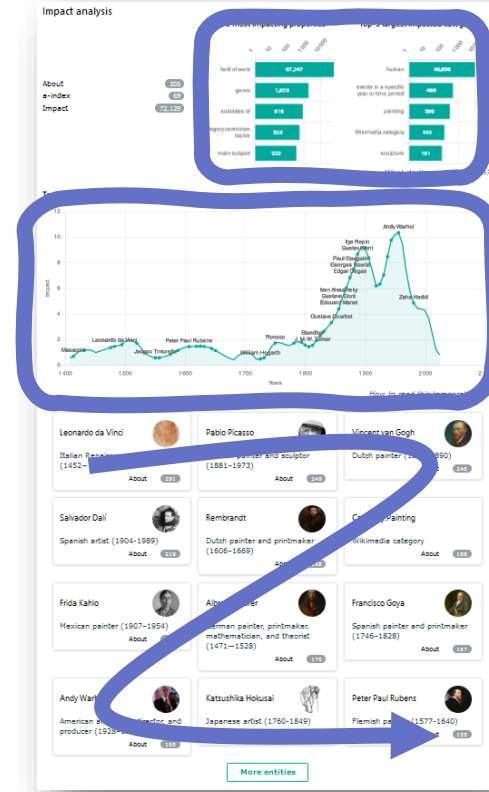
Summary



1 Quadrant analysis



2 Ranking analysis



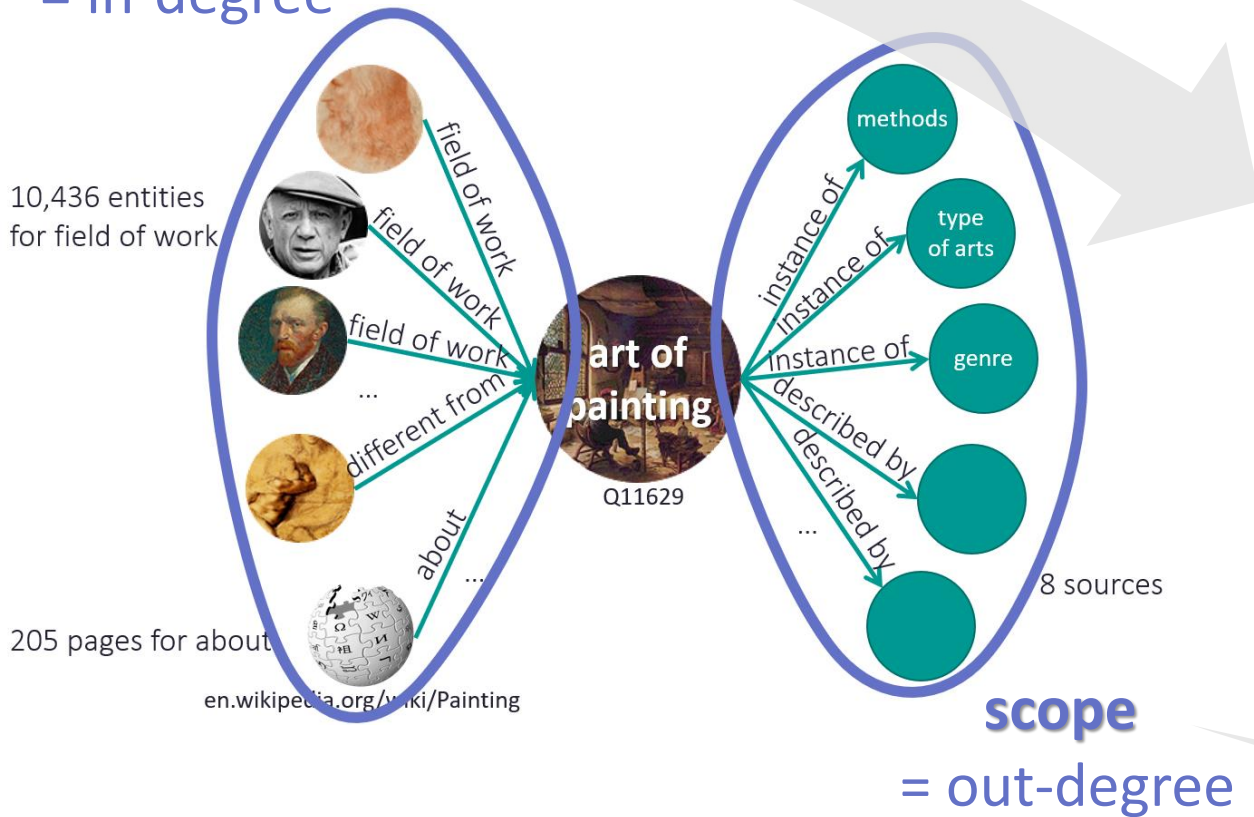
3 Impact analysis

by category

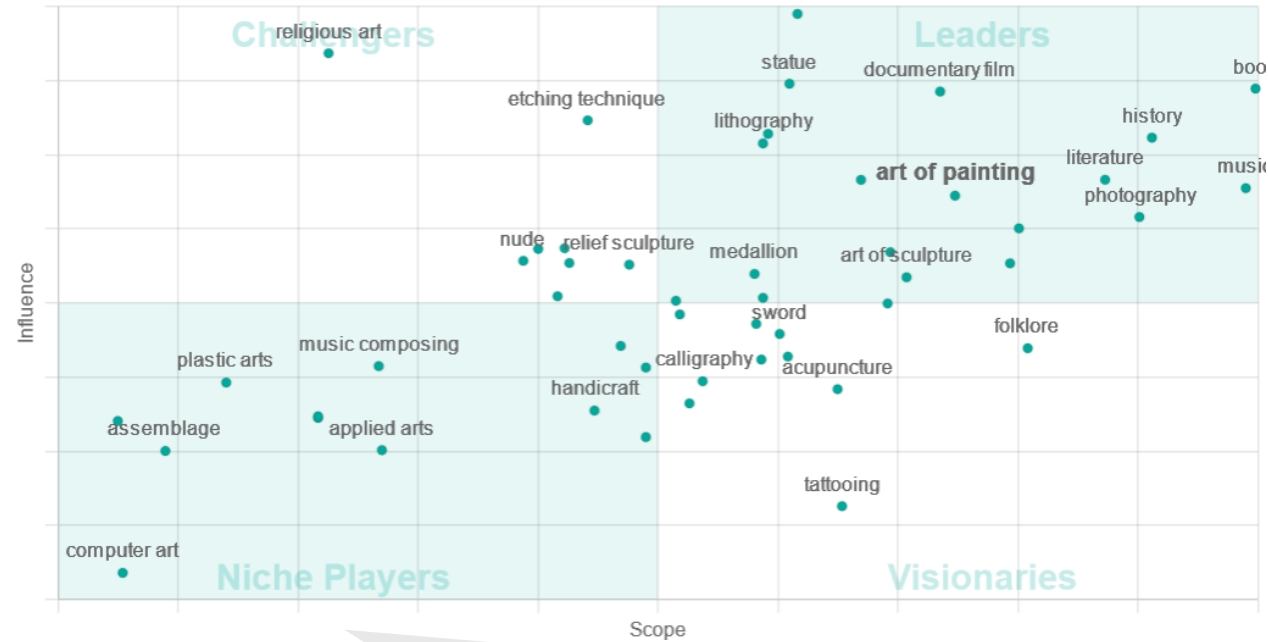
Nacke, O. (1979). "Informetrie: eine neuer Name für eine neue Disziplin". Nachrichten für Dokumentation. 30 (6): 219–226.

1 Quadrant analysis

influence
= in-degree



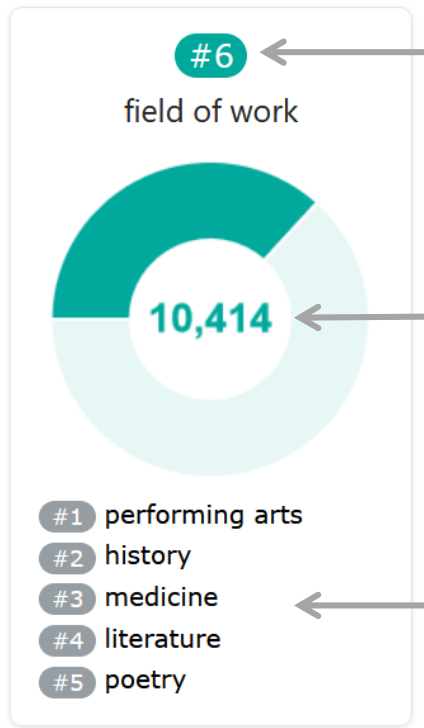
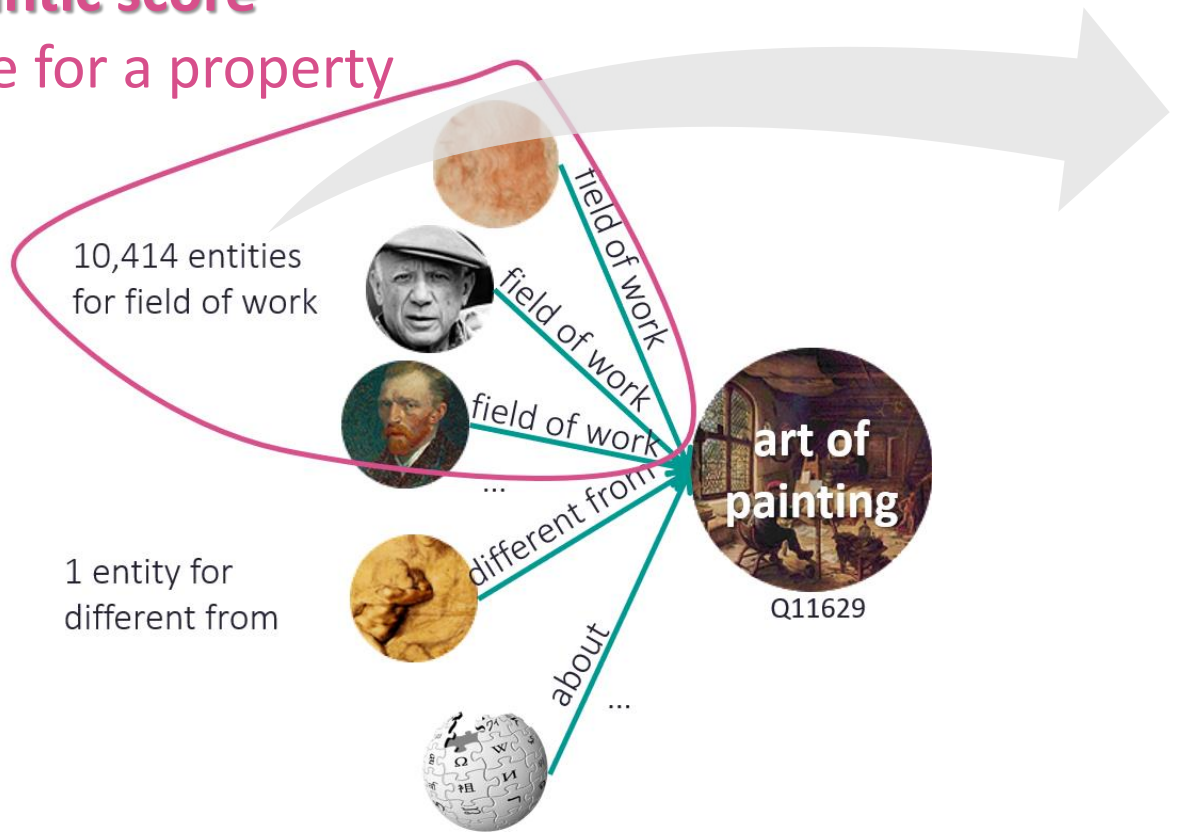
Comparison with similar entities :



2 Ranking analysis

semantic score

= in-degree for a property



Ranking of the entity for this property

Number of impacted entities for this property

Top-5 entities for this property

Good ranking = inequal distribution of facts

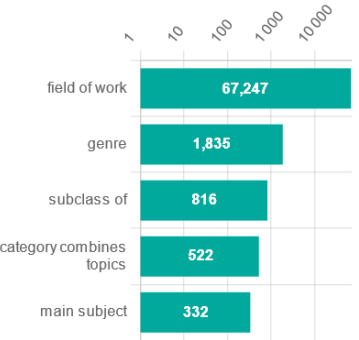
Thèse de Hassan Abdallah

3

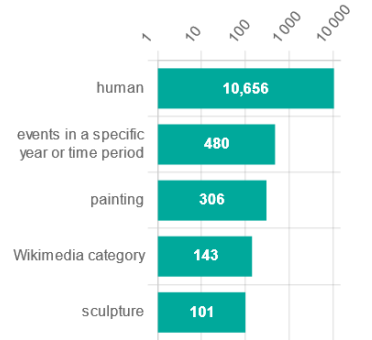
Impact analysis

Categorical analysis

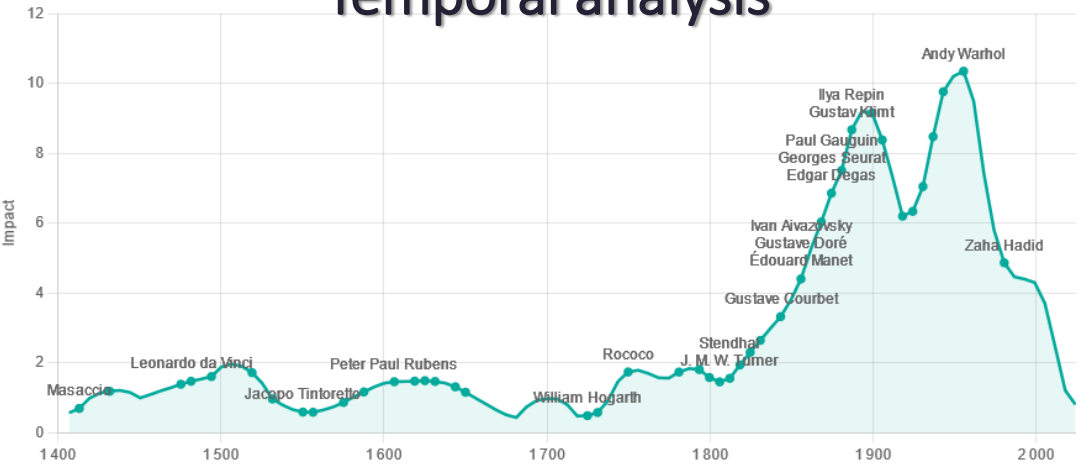
Top-5 most impacting properties



Top-5 largest impacted categories



Temporal analysis



Impacted entities

Leonardo da Vinci
Italian Renaissance polymath (1452–1519)
About 291

Pablo Picasso
Spanish painter and sculptor (1881–1973)
About 249

Vincent van Gogh
Dutch painter (1853–1890)
About 248

Salvador Dalí
Spanish artist (1904–1989)
About 219

Rembrandt
Dutch painter and printmaker (1606–1669)
About 188

Category:Painting
Wikimedia category
About 188

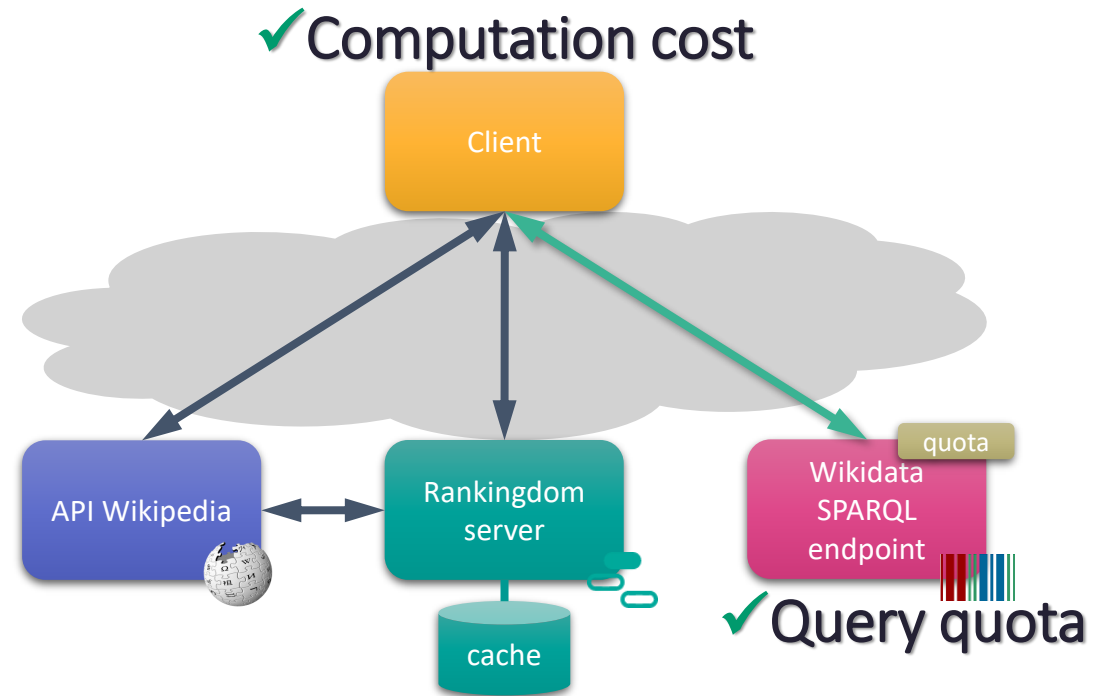
Frida Kahlo
Mexican painter (1907–1954)
About 178

Albrecht Dürer
German painter, printmaker, mathematician, and theorist (1471–1528)
About 176

Francisco Goya
Spanish painter and printmaker (1746–1828)
About 167

On-demand computation

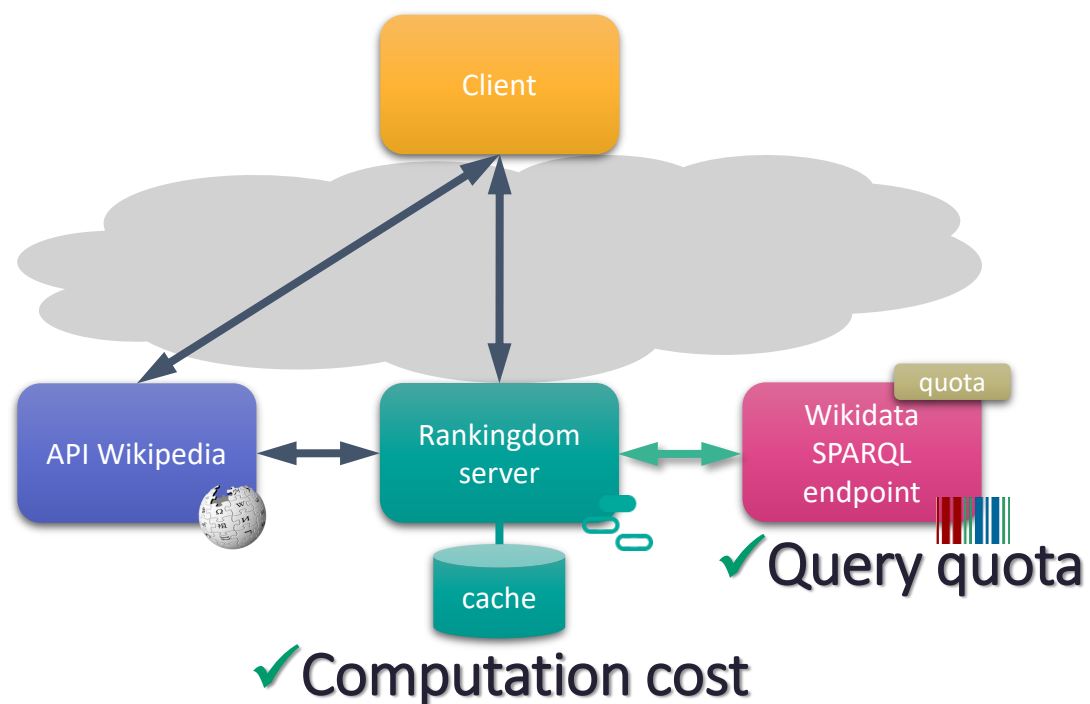
1. Entity search
2. Cache retrieval
3. **On-demand computation**
4. Image downloading
5. Cache storing



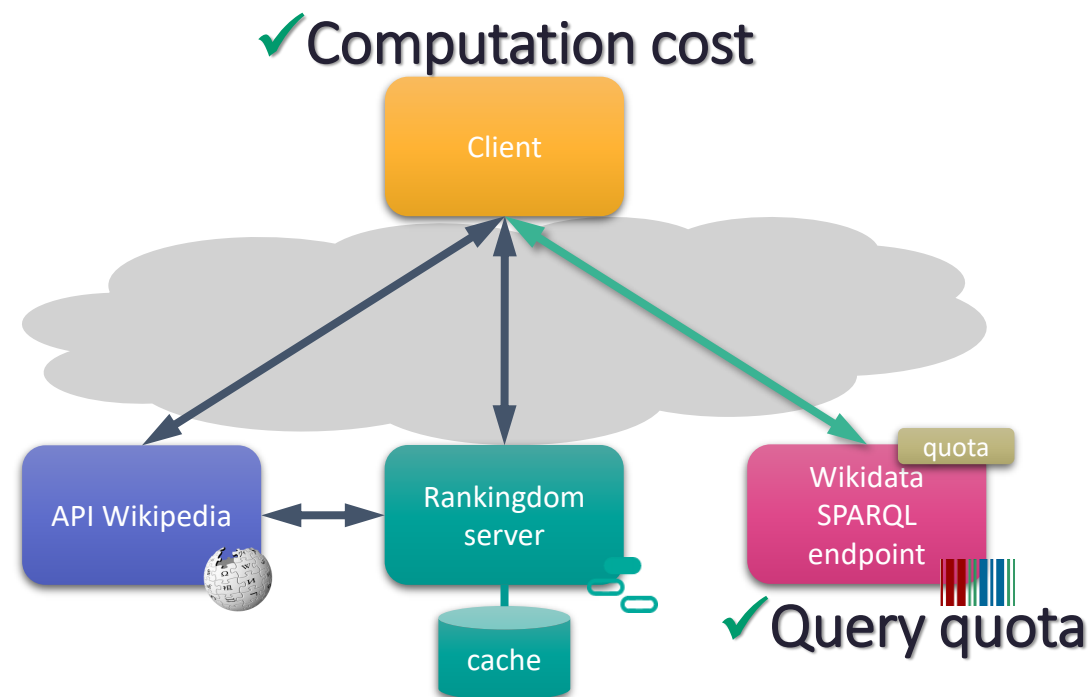
Soulet, A., & Suchanek, F. M. Anytime large-scale analytics of linked open data. In ISWC 2019.

Cooperative architecture for scalability

Centralized computation

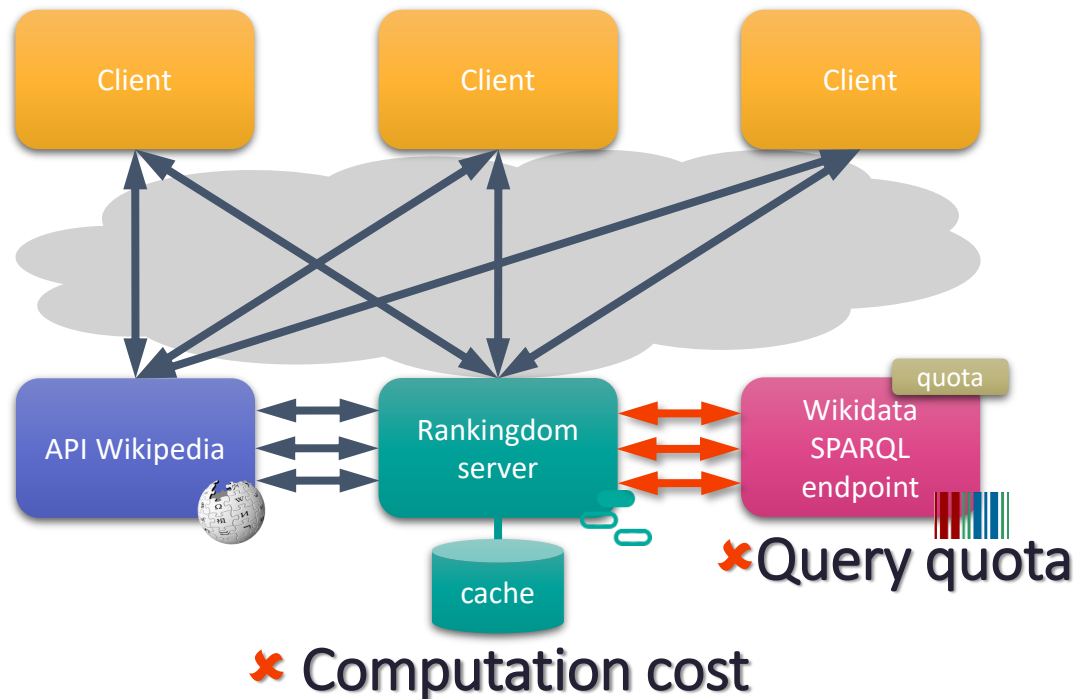


Decentralized computation

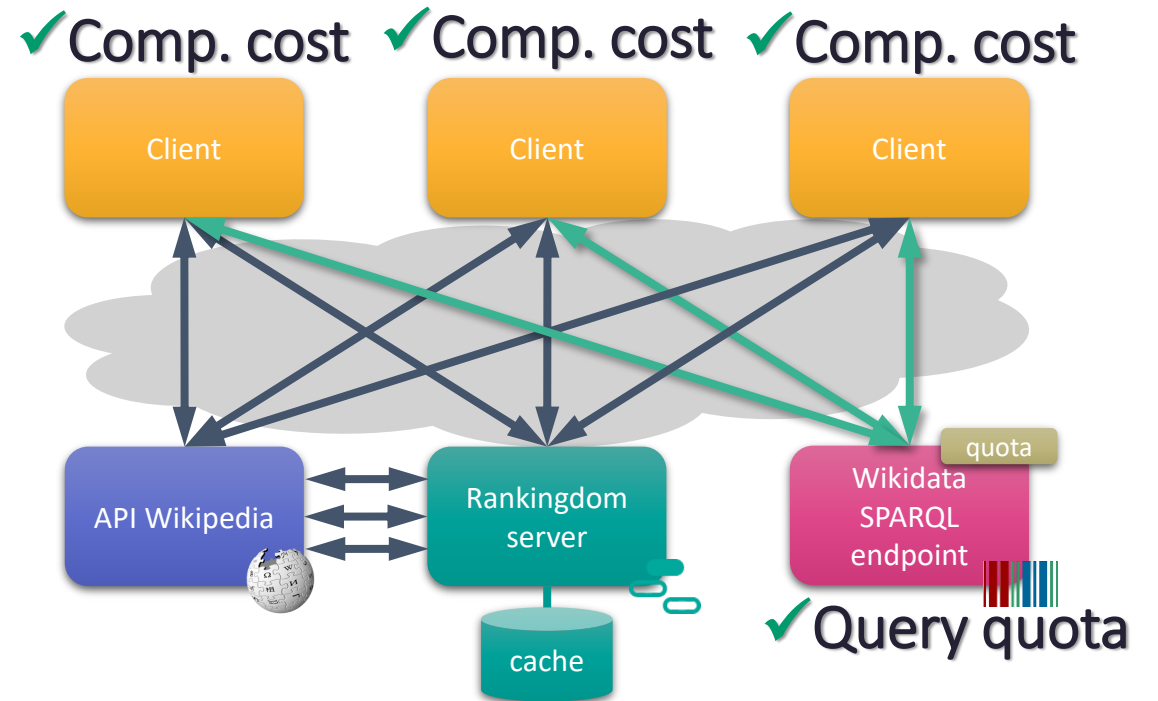


Cooperative architecture for scalability

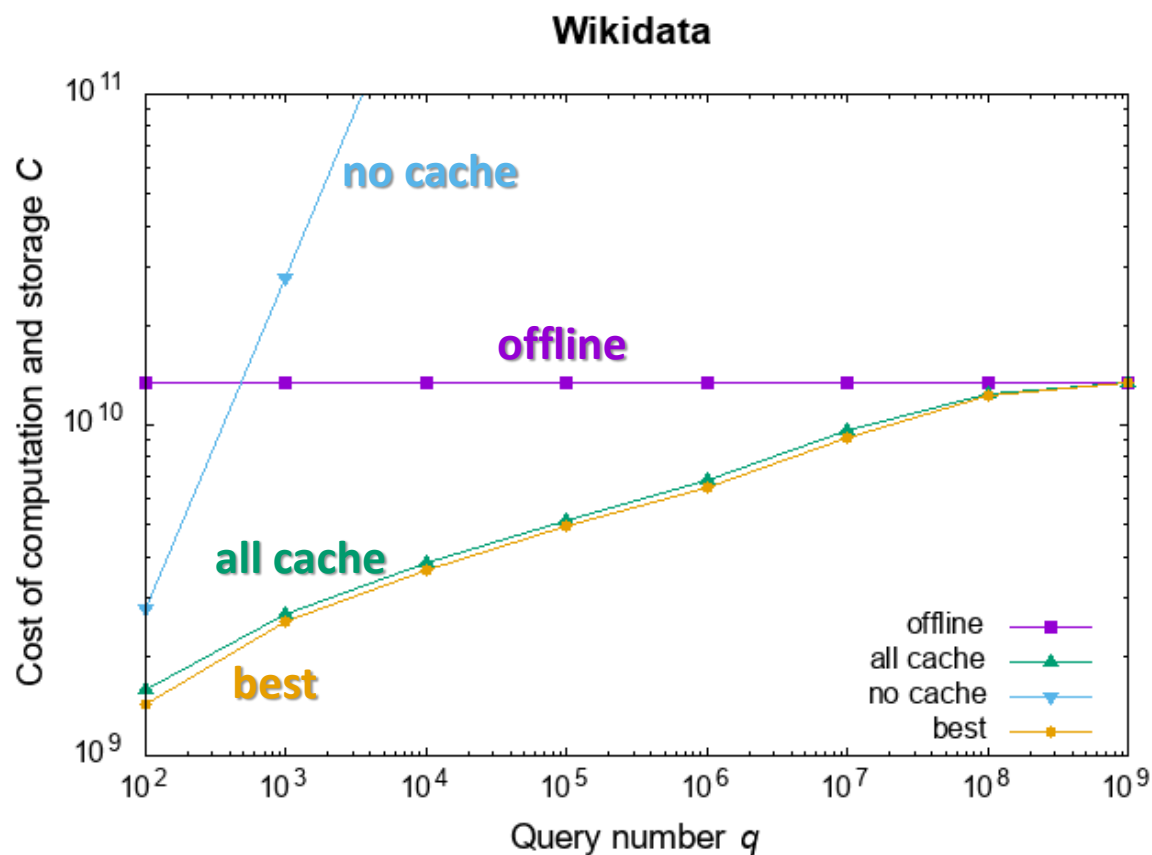
Centralized computation



Decentralized computation



Cooperative architecture for frugality



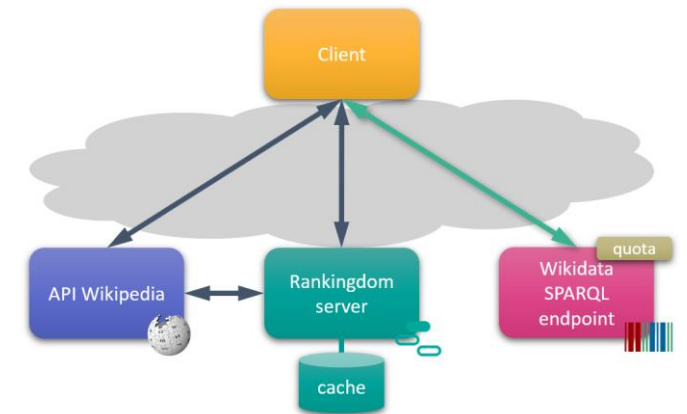
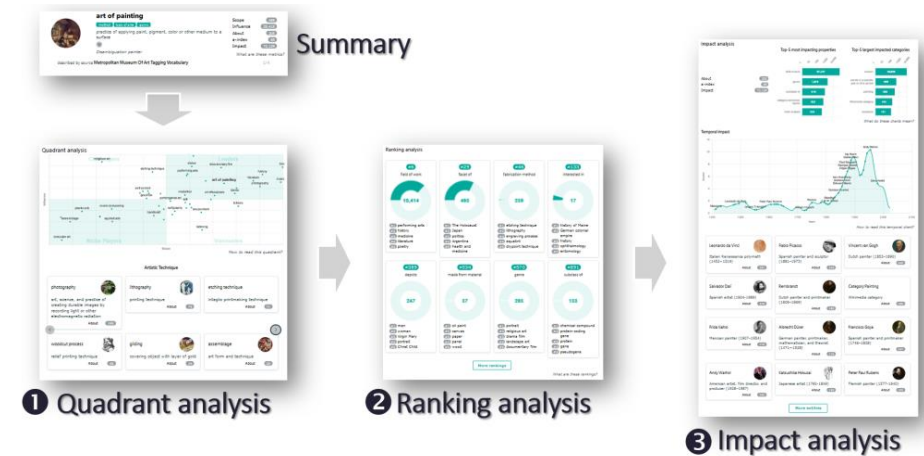
- **offline**: everything is computed offline
- **no cache**: everything is computed online at each query
- **all cache**: everything is computed once and cached
- **best**: optimal solution (everything is not cached)

On-demand computation

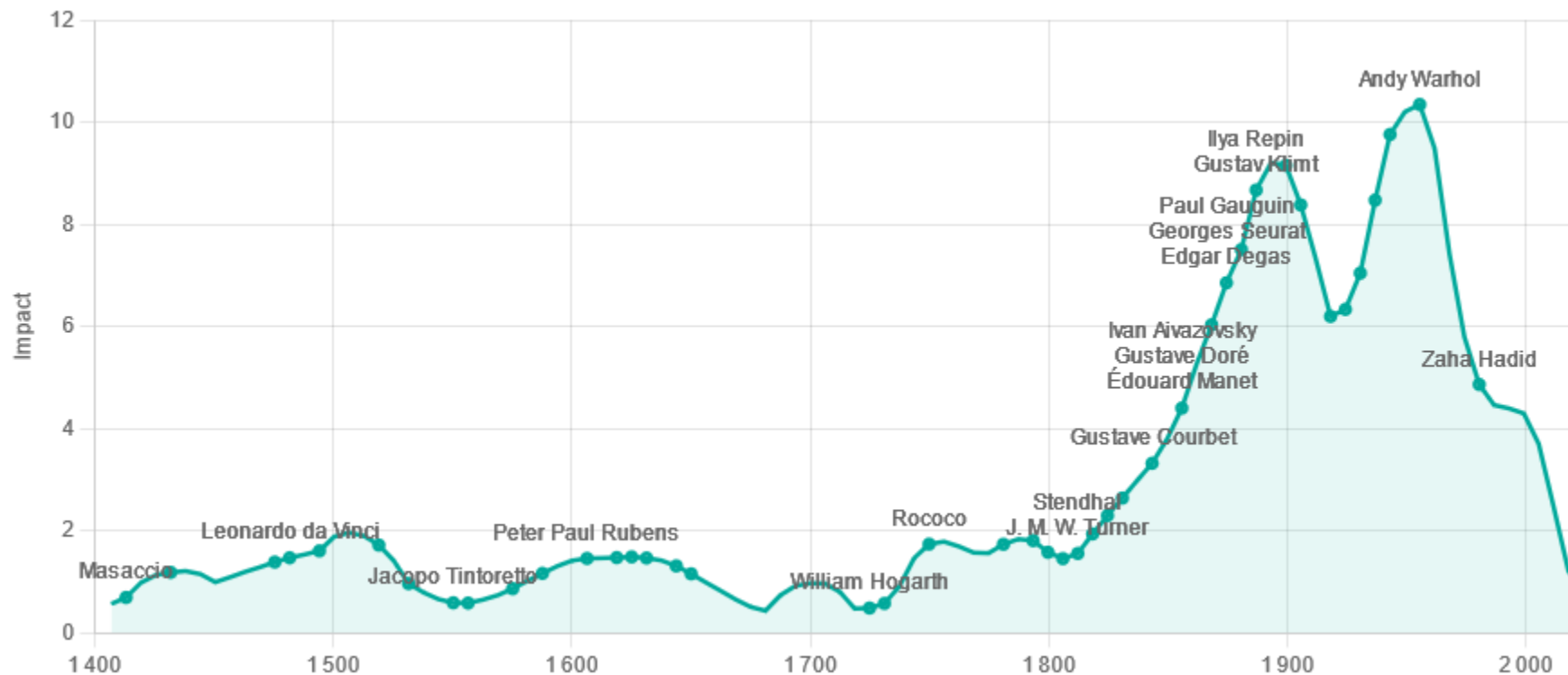
Soulet, A. Should We Consider On-Demand Analysis in Scale-Free Networks?. In IDA 2023.

Summary

- ❑ Data narrative from general to specific for homogeneous analysis
- ❑ Generic and semantic scores for multidimensional analysis
- ❑ On-demand computation for freshness and diversity
- ❑ Cooperative architecture for scalability and frugality



Challenges: availability and selection biases



How to measure inequality and diversity in knowledge graphs?



knowledge analytics on demand from 15,501,579,370 facts

🔍 Analyze...

Thank you for your attention and visit rankingdom.org!