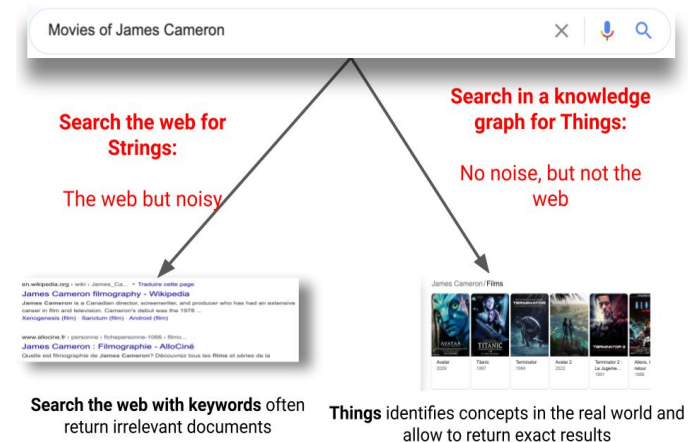


MiKroloG: The Microdata Knowledge Graph

Equipes
GDD Nantes
LACODAM Rennes

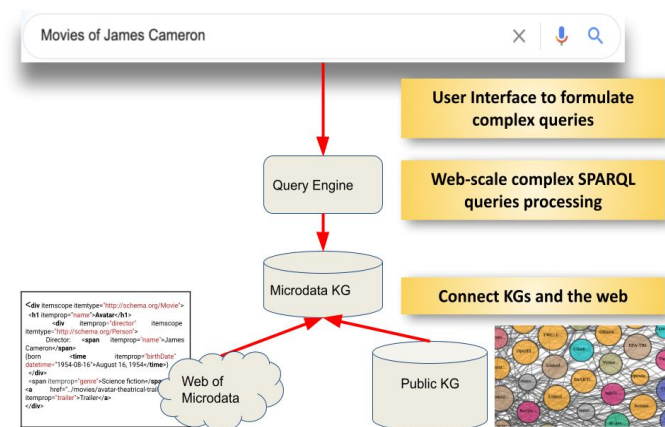
MiKroloG Context



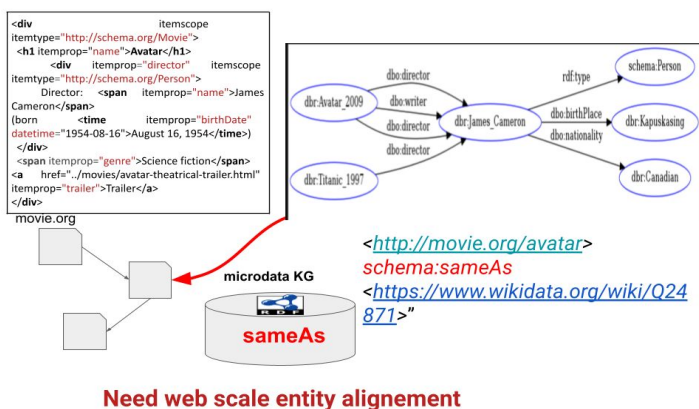
MiKroloG Objective

- Search the web with “Things” in order to get accurate results over a wide diversity of sources.
- MiKroloG makes virtually search for every “Things” on the Web:
 - MiKroloG University search
 - MiKroloG Company Search
 - MiKroloG Employee Search
 - MiKroloG Health Search
 - MiKroloG * Search

MiKroloG Approach



Machine learning to connect web entities with KG entities: sameAs KG



Web preemption for web scale complex query processing

- **Need preemptive ranking operators (Top-k) that can be suspended and resumed in constant time and bounded memory.**

T. Minier, Hala Skaf-Molli, Pascal Molli (2019) SaGe: Web Preemption for Public SPARQL Query services. The WWW Conference, The Web Conference, 2019.

A regular user uses Sparklis to formulate a complex query

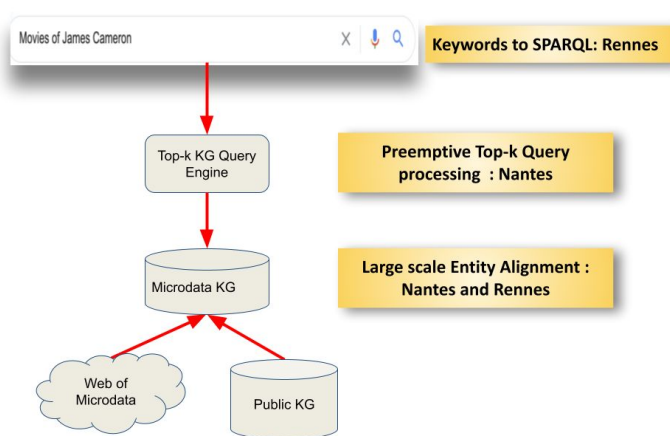
- **Need to connect Sparklis to a preemptive query engine**
- **Need to handle unnatural query formulation when traversing sameAs links in the microdata KG**

“give me every film whose director is (James Cameron or something that is the same as James Cameron) or that is the same as a film whose director is James Cameron”, where one would expect “give me every film whose director is James Cameron”.

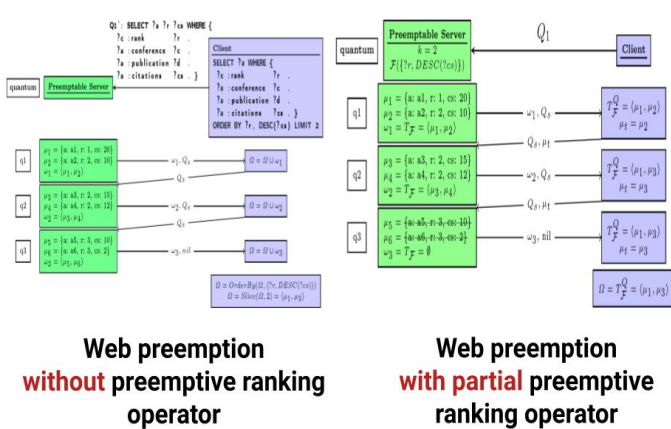
SELECT DISTINCT ?Person_1 (COUNT(DISTINCT ?Book_1 number_of_165) WHERE { ?Person_1 a dbo:Person . ?Book_165 a dbo:Book . ?Book_165 dbo:author ?Person_1 . } GROUP BY ?Person_1 ORDER BY DESC(?number_of_165)

Sébastien Ferré. Sparklis: an expressive query builder for SPARQL endpoints with guidance in natural language. Semantic Web 8.3 (2017).

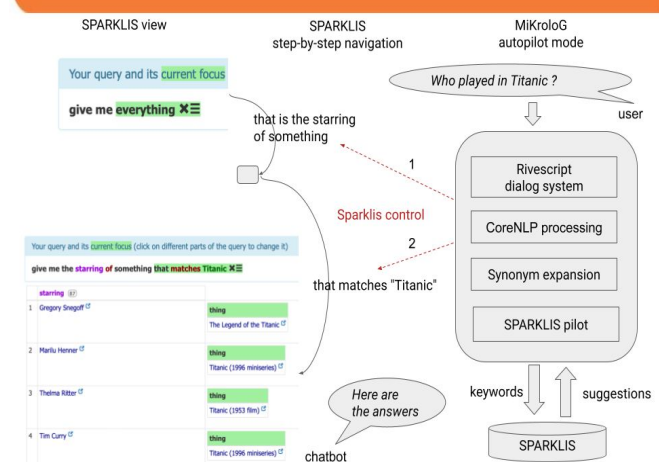
MiKroloG Organization



MiKroloG Results: Preemptive Top-k



MiKroloG Results: Autopilot Sparklis



MiKroloG use-cases: Bioschemas

UC1: Pandemic crisis.

- Explore recent research papers, related experimental data as well as lab protocols, data analysis tools, related training material

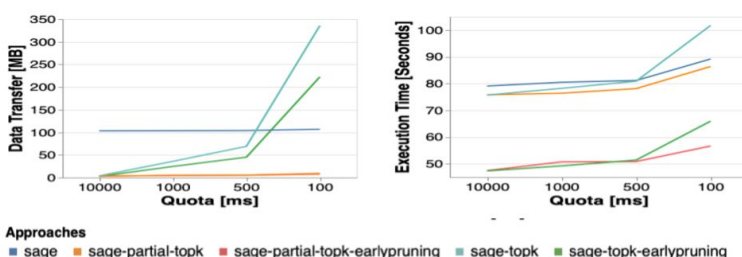
UC2: Data scientist.

- Queries on tool-oriented resources such as Bio.Tools to find appropriate data processing tools and design bioinformatic workflows

What is Bioschemas?

Bioschemas aims to improve the Findability on the Web of life sciences resources such as datasets, software, and training materials. It does this by encouraging people in the life sciences to use Schema.org markup in their websites so that they are indexable by search engines and other services. Bioschemas encourages the consistent use of markup to ease the consumption of the contained markup across many sites. This structured information then makes it easier to discover, collate, and analyse distributed resources.

<https://bioschemas.org/>



Average Performance on Wikidata Queries: A reduction in data transfer by a factor of 100 and a reduction of up to 39% in Wikidata query execution time

Processing SPARQL TOP-k Queries Online with Web Preemption, submitted to QuWeDa 2022: 6th Workshop on Storing, Querying, and Benchmarking the Web of Data at ISWC 2022

Conclusion

- **MiKroloG:**
 - Query the web with expressive queries and return complete results
 - Provide an alternative to search the web, query with “Things”
 - As microdata covers many domains including life science, education, people, events, etc...
 - MiKroloG, virtually search for every “Things”
 - MiKroloG University search
 - MiKroloG Company Search
 - MiKroloG Employee Search
 - MiKroloG * Search
- **ANR Mekano for querying the web of “Things” has just been accepted**