Advanced Methods of Information Retrieval
- Dataset Profiling -

Dr. Elena Demidova
SS 2018
Recap
A knowledge graph (KG)

- The term “knowledge graph” is currently used to refer to semantic Web knowledge bases
Overview of popular knowledge graphs

<table>
<thead>
<tr>
<th>Name</th>
<th>Instances</th>
<th>Facts</th>
<th>Types</th>
<th>Relations</th>
</tr>
</thead>
<tbody>
<tr>
<td>DBpedia (English)</td>
<td>4,806,150</td>
<td>176,043,129</td>
<td>735</td>
<td>2,813</td>
</tr>
<tr>
<td>YAGO</td>
<td>4,595,906</td>
<td>25,946,870</td>
<td>488,469</td>
<td>77</td>
</tr>
<tr>
<td>Freebase</td>
<td>49,947,845</td>
<td>3,041,722,635</td>
<td>26,507</td>
<td>37,781</td>
</tr>
<tr>
<td>Wikidata</td>
<td>15,602,060</td>
<td>65,993,797</td>
<td>23,157</td>
<td>1,673</td>
</tr>
<tr>
<td>NELL</td>
<td>2,006,896</td>
<td>432,845</td>
<td>285</td>
<td>425</td>
</tr>
<tr>
<td>OpenCyc</td>
<td>118,499</td>
<td>2,413,894</td>
<td>45,153</td>
<td>18,526</td>
</tr>
<tr>
<td>Google’s Knowledge Graph</td>
<td>570,000,000</td>
<td>18,000,000,000</td>
<td>1,500</td>
<td>35,000</td>
</tr>
<tr>
<td>Google’s Knowledge Vault</td>
<td>45,000,000</td>
<td>271,000,000</td>
<td>1,100</td>
<td>4,469</td>
</tr>
<tr>
<td>Yahoo! Knowledge Graph</td>
<td>3,443,743</td>
<td>1,391,054,990</td>
<td>250</td>
<td>800</td>
</tr>
</tbody>
</table>

(Paulheim, 2016)
Sources of knowledge graphs

- Knowledge graphs manually curated by experts
  - By organizations
  - By small closed communities
- Crowd-edited knowledge graphs
  - Large open community
- Knowledge graphs harvested from the Web
  - Created by heuristics
  - Automatically or semi-automatically
Linked Data principles

- P1: Use URIs as names for things
- P2: Use HTTP URIs so that people can look up those names
- P3: When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
- P4: Include links to other URIs, so that more things can be discovered

https://www.w3.org/DesignIssues/LinkedData.html
Linked Open Data Cloud

- Over 1000 datasets
- Majority of links connect equivalent entities in the different datasets

http://lod-cloud.net/
Dataset Profiling
Aims of the session “Dataset Profiling”

- **Lecture:**
  - Understand the concepts of:
    - Dataset
    - Dataset profile and dataset profile features
  - Familiarize with:
    - Selected vocabularies to describe dataset
    - Selected methods to infer missing information
    - Challenges in dataset search

- **Hands-on:**
  - Get practical experience with:
    - Dataset catalogues
    - Vocabularies for dataset profiling
    - Computation and inference of dataset profile features
Motivation

- Number of available datasets grows constantly
- Dataset characteristics vary with respect to
  - quality, provenance, interlinking, licenses, statistics and dynamics
- Reliable information on dataset characteristics is needed to empower applications using these datasets

2007

2009

2017

http://lod-cloud.net/
Terminology

*What is a dataset?*
Terminology

**An RDF Dataset** is a set of RDF triples that are published, maintained or aggregated by a single provider. (VoID Vocabulary)[1]
Terminology

An RDF Dataset is a set of RDF triples that are published, maintained or aggregated by a single provider. (VoID Vocabulary)[1]

Is a knowledge graph an RDF dataset?
## DBpedia datasets (an excerpt)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>en</th>
<th>de</th>
<th>es</th>
<th>fr</th>
<th>ja</th>
<th>nl</th>
<th>pt</th>
<th>ru</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citation Data</td>
<td>tql ?</td>
<td>tql ?</td>
<td>tql ?</td>
<td>tql ?</td>
<td>tql ?</td>
<td>tql ?</td>
<td>tql ?</td>
<td>tql ?</td>
</tr>
<tr>
<td>Citation Links</td>
<td>tql ?</td>
<td>tql ?</td>
<td>tql ?</td>
<td>tql ?</td>
<td>tql ?</td>
<td>tql ?</td>
<td>tql ?</td>
<td>tql ?</td>
</tr>
</tbody>
</table>
Where do we find relevant datasets?

Example: building an application that presents movie information

• Where do we find relevant datasets?
• How do we search for them?

In the LOD cloud?
Beyond?

Any ideas?

http://lod-cloud.net/
Web Search Engines

Google

movie dataset download

About 3.520.000 results (0.40 seconds)

The Movies Dataset | Kaggle
https://www.kaggle.com/ounsabanki/the-movies-dataset
Nov 10, 2017 - Context: These files contain metadata for all 45,000 movies listed in the Full MovieLens Dataset. The dataset consists of movies released on or before July 2017. Data points include cast, crew, plot keywords, budget, revenue, posters, release dates, languages, production companies, countries, TMDB vote...

TMDB 5000 Movie Dataset | Kaggle
https://www.kaggle.com/tmdb/tmdb-movie-metadata
Sep 29, 2017 - Background: What can we say about the success of a movie before it is released? Are there certain companies (Pixar?) that have found a consistent formula? Given that major films costing over $100 million to produce can still flop, this question is more important than ever to the industry. Film aficionados...

UCI Machine Learning Repository: Movie Data Set
https://archive.ics.uci.edu/ml/datasets/Movie
Data Set Download: Data Folder, Data Set Description: Abstract: This data set contains a list of over 10,000 films including many older, odd, and cult films. There is information on actors, casts, directors, producers, studios, etc.

IMDB 5000 Movie Dataset - dataset by popculture | data.world
https://data.world/popculture/imdb-5000-movie-dataset
Chuan Sun (@sundeepblue on Github) scraped tons of metadata using a combination of www-the-numbers.com, IMDB.com, and a Python library called "scrapy". He was able to obtain 29 variables for 5043 movies and 4906 posters (988MB), spanning across 100 years in 66 countries. There are 2399 unique director...

MovieLens | GroupLens
https://grouplens.org/datasets/movielens/
These datasets will change over time, and we are not responsible for providing correct publicly. We will...

Dataset Catalogues

31 datasets found for "Movie"

Order by: Relevance

Movies from argentina
This dataset has no description

Movies Directed By Kubrick
Information about movies directed by Kubrick. Various datasources.

Registered demonstrators movies
List of registered demonstrators films on the Russian Federation territory, indicating their location and accessories for cinema networks with an indication of the last date of...

Internet Movie Database
Large film/movie database containing 425,000+ titles, 1,700,000+ filmographies of cast and crew members Films from 1891 to Present Foreign and independent movies, television...

MovieLens Data Sets
This data set contains 10000054 ratings and 95560 tags applied to 10081 movies by 71507 users of the online movie recommender service MovieLens. Users were selected at random...

datahub.io
What is a relevant dataset?

Example: building an application that presents movie information

- How do we identify which dataset to use?
- Which dataset characteristics are relevant to take the decision?
Terminology

A Dataset Profile is a formal representation of a set of dataset characteristics (features).

A dataset profile describes the dataset and aids dataset discovery, recommendation and comparison with regard to the represented features.

A Dataset Profile Feature is a characteristic describing a certain dataset attribute.
Example of a basic dataset profile

The following example declares the resource :DBpedia as a void:Dataset and provides general dataset metadata using the Dublin Core Metadata Terms vocabulary:  

(VoID Vocabulary)[1]

:DBpedia a void:Dataset;
  dcterms:title "DBPedia";
  dcterms:description "RDF data extracted from Wikipedia";
  dcterms:contributor :FU_Berlin;
  dcterms:contributor :University_Leipzig;
  dcterms:contributor :OpenLink_Software;
  dcterms:contributor :DBpedia_community;
  dcterms:source <http://dbpedia.org/resource/Wikipedia>;
  dcterms:modified "2008-11-17"^^xsd:date;

https://www.w3.org/TR/void/
Dataset profile features categories

Categories of dataset profile features

- General
- Qualitative
- Provenance
- Links
- Licensing
- Statistical
- Dynamics
General features of a dataset profile

General dataset profile features are features carrying high-level semantic information.

Examples of general features:
- Domain / topic
- Contributors
- Language
- ...


dcterms:subject <http://dbpedia.org/resource/Computer_science>;
dcterms:contributor :OpenLink_Software;
Qualitative features of a dataset profile

Data quality is generally conceived as fitness for use, i.e. the capability of data to respond to the demands of a specific user given a specific use case.

Quality dimensions:
- Accessibility dimensions
- Intrinsic dimensions
- Trust dimensions
- Contextual dimensions
- Representational dimensions
- …

(Zaveri et al., 2016)
Qualitative features: accessibility

- Accessibility dimensions include:
  - availability, licensing, interlinking, security and performance.
- Availability metrics
  - Accessibility of SPARQL endpoints
  - Accessibility of RDF dumps
  - Dereferencable URIs

Recap: What is a dereferencable URI?
Qualitative features: intrinsic

- Intrinsic dimensions:
  - focus on whether information correctly and compactly represents the real world data and whether information is logically consistent in itself
  - accuracy
  - consistency
  - conciseness
Qualitative features: intrinsic / accuracy

• **Accuracy** is the extent to which data is correct.
• **Syntactic accuracy**: the degree to which data values are close to its corresponding definition, i.e. free of syntax errors.
• **Semantic accuracy**: the degree to which data values correctly represent the real world facts.

**Metrics:**
- Outlier detection
- Inaccurate values, facts
- Malformed literals / literals incompatible with datatype range
- Erroneous annotations
- Inaccurate annotations, labeling, classification

Which of these metrics refer to semantic / and which to syntactic accuracy?
Qualitative features: intrinsic / consistency

Consistency: a knowledge base is free of (logical/formal) contradictions with respect to particular knowledge representation and inference mechanisms.

Examples:

- Invalid usage of classes and properties
- Classes and properties used without any formal definition
- Redefinition of external classes and properties

Assessing consistency: an inference engine or a reasoner that supports the respective expressivity of the knowledge representation.
Qualitative features: intrinsic / conciseness

Conciseness:
redundancy of entities, at the schema or the data level.

• Schema level: the data does not contain redundant attributes.
  • E.g. dbo:starring, http://schema.org/actors)
  • Metrics:
    • Proportion of unique attributes in a dataset.

• Data level: the data does not contain redundant instances.
  • Metrics:
    • Proportion of duplicate instances in a dataset.

How to identify redundant attributes?
How to find duplicate instances in a dataset?
Qualitative features: trust

• Reputation
  • A user judgment to determine data source integrity.

• Believability
  • The extent to which information is regarded as true and credible.

• Verifiability
  • The degree by which a data consumer can assess the dataset correctness.

• Objectivity
  • The degree to which the data is unbiased, unprejudiced and impartial.
  • Metrics: Fact confirmation by independent sources
Example: fact confirmation by independent sources

Winston Churchill’s date of birth in Wikidata

<table>
<thead>
<tr>
<th>date of birth</th>
<th>30 November 1874</th>
</tr>
</thead>
<tbody>
<tr>
<td>stated in</td>
<td>Integrated Authority File</td>
</tr>
<tr>
<td>retrieved</td>
<td>9 April 2014</td>
</tr>
<tr>
<td>stated in</td>
<td>data.bnf.fr</td>
</tr>
<tr>
<td>retrieved</td>
<td>10 October 2015</td>
</tr>
<tr>
<td>reference URL</td>
<td><a href="http://data.bnf.fr/ark:/12148/c2119b36090">http://data.bnf.fr/ark:/12148/c2119b36090</a></td>
</tr>
<tr>
<td>stated in</td>
<td>RKDartists</td>
</tr>
<tr>
<td>reference URL</td>
<td><a href="http://explore.rkd.nl/nl/explore/artists/16810">http://explore.rkd.nl/nl/explore/artists/16810</a></td>
</tr>
<tr>
<td>RKDartists ID</td>
<td>16810</td>
</tr>
<tr>
<td>title</td>
<td>Winston Leonard Spencer (Lord) Churchill (English)</td>
</tr>
<tr>
<td>stated in</td>
<td>SNAC</td>
</tr>
<tr>
<td>SNAC Ak ID</td>
<td>wsprt5g0</td>
</tr>
<tr>
<td>named as</td>
<td>Winston Churchill</td>
</tr>
<tr>
<td>retrieved</td>
<td>9 October 2017</td>
</tr>
<tr>
<td>stated in</td>
<td>Nationalencyclopedin</td>
</tr>
<tr>
<td>NE.se ID</td>
<td>winston-s-churchill</td>
</tr>
<tr>
<td>named as</td>
<td>Winston Churchill</td>
</tr>
<tr>
<td>retrieved</td>
<td>9 October 2017</td>
</tr>
<tr>
<td>stated in</td>
<td>Find a Grave</td>
</tr>
<tr>
<td>Find a Grave memorial ID</td>
<td>2194</td>
</tr>
<tr>
<td>named as</td>
<td>Winston Leonard Spencer Churchill</td>
</tr>
<tr>
<td>retrieved</td>
<td>9 October 2017</td>
</tr>
<tr>
<td>stated in</td>
<td>Babelio</td>
</tr>
<tr>
<td>Babelio author ID</td>
<td>42789</td>
</tr>
<tr>
<td>named as</td>
<td>Winston Churchill</td>
</tr>
<tr>
<td>retrieved</td>
<td>9 October 2017</td>
</tr>
</tbody>
</table>
Qualitative features: contextual

- **Completeness**: the degree to which all required information is present in a particular dataset.
  - (a) Schema completeness: the degree to which the classes and properties of an ontology are represented.
  - (b) Property completeness / missing values for a specific property.
  - (c) Population completeness: the percentage of all real world objects of a particular type represented in the dataset.
  - (d) Interlinking completeness

- **Amount-of-data**: the quantity and volume of data that is appropriate for a particular task.

- **Relevancy**: provision of information which is in accordance with the task at hand and important to the users’ query.
Qualitative features: representational

- Representational conciseness
  - Compact, well formatted, clear and complete data representation

- Representational consistency
  - The format and structure of the information conforms to data from other sources (i.e. vocabulary reuse)

- Understandability
  - The ease of data comprehension for a human user.

- Interpretability
  - Use of an appropriate notation, conformance to the technical specification of the consumer.

- Versatility
  - Internationalized representation, alternative representations, alternative access methods for a dataset.
Provenance

Search result for “What is the population of Germany?”

Wikidata entry for Germany
https://www.wikidata.org/wiki/Q183
Provenance dataset profile features

“Provenance is defined as a record that describes the people, institutions, entities, and activities involved in producing, influencing, or delivering a piece of data or a thing.”

PROV-N: The Provenance Notation. https://www.w3.org/TR/prov-n/
Describing provenance with the PROV ontology

https://www.w3.org/TR/prov-o/
Links dataset profile features

Links features reflect the interlinking of the dataset with other datasets.

Interlinking is also a quality dimension of the dataset with corresponding metrics:

- Interlinking degree
- Links to external data providers

Recap:
What are the links to external data providers in RDF triples?
Examples: external links

**Identity Links** point at URI aliases used by other data sources to identify the same real-world object or abstract concept.

dbr:Brunswick_stew owl:sameAs
https://www.wikidata.org/wiki/Q994169

**Relationship Links** point at related things in other data sources, for instance, people, places or ingredients.

https://www.wikidata.org/wiki/Q994169 dbo:ingredient dbr:Maize

**Vocabulary Links** point from data to the definitions of the vocabulary terms that are used to represent the data, as well as from these definitions to the definitions of related terms in other vocabularies.

dbr:Brunswick_stew dbo:ingredient dbr:Maize
Licensing dataset profile features

Licensing is granting permission to re-use a dataset under certain conditions.

As a dataset profile feature, licensing is information regarding the license, under which the content can be used.

Licensing is also a quality dimension with the corresponding metrics:

- Machine-readable indication of a license
- Human-readable license
- Specific permissions (reproduction, distribution, modification)

`dcterms:license <http://www.opendatacommons.org/licenses/by/>;`
Dataset reuse statistics vs. license

- An increase in open license dataset reuse in scientific publications

(Endris et al., 2017)
Statistical dataset profile features

Statistical features are statistical characteristics of the dataset at the schema and instance level (e.g. size, coverage, etc.).

- Schema-level
  - Type usage
  - Properties usage

- Instance-level
  - URI usage per subject (/object)
  - number of triples with literals
  - number of internal and external links
  - number of used languages per literal

How to measure type usage with SPARQL?
Statistical dataset profile features

Statistical features are statistical characteristics of the dataset at the schema and instance level (e.g. size, coverage, etc.).

- **Schema-level**
  - Type usage
  - Properties usage
- **Instance-level**
  - URI usage per subject (/object)
  - number of triples with literals
  - number of internal and external links
  - number of used languages per literal

---

*How to measure type usage with SPARQL?*

```sparql
SELECT (COUNT(?building) AS ?count)
WHERE {
}
```
Dynamics dataset profile features

Dynamics include currency, volatility, and timeliness; dynamics is also a quality dimension of the dataset.

- Currency: how promptly the data is updated.
  - Metrics:
    - Difference between the observation and the last modified time
    - Proportion of outdated data
- Volatility: the frequency with which data varies in time.
- Timeliness: how up-to-date data is, relative to a task.
Vocabularies for dataset profiles
DCAT: Data Catalog Vocabulary

DCAT is an RDF vocabulary for representing dataset catalogs published on the Web.

Main classes:
• dcat:Catalog represents the catalog.
• dcat:Dataset represents a dataset in a catalog.
• dcat:Distribution represents an accessible form of a dataset as for example a downloadable file, an RSS feed or a web service that provides the data.

https://www.w3.org/TR/vocab-dcat/
The Vocabulary of Interlinked Datasets (VoID)

VoID Vocabulary
an RDF Schema vocabulary for expressing metadata about RDF datasets.

VoID also allows the description of **RDF links** between datasets.

A linkset is a collection of RDF links between two datasets, i.e. RDF triples whose subject and object are described in different datasets.

https://www.w3.org/TR/void/
Refinement of dataset profiles through knowledge graph completion
Knowledge graph completion

- Knowledge graph completion means increasing coverage of the knowledge graph
  - Prediction of missing entity types
  - Prediction of missing entities
  - Prediction of missing relations between entities

- Type information is:
  - Crucial for many applications (and thus particularly relevant for dataset profiles)
  - Often missing in knowledge graphs

- **How can we infer missing entity types?**
Methods for Type Inference

• Internal methods
  • Using information in the knowledge graph (e.g. links)

• External methods
  • Using information outside of the knowledge graph (links to web pages, Wikipedia link graph, text corpora)

(Paulheim et al., 2016)
Type Inference: SD-Type Algorithm

- Idea: type prediction as a link-based instance classification
- Each link to / from an instance is a type indicator
- Intuition: predicate usage can help to determine types of subject and object.
  - E.g. getting married is typical for instances of type person
  - When observing the triple:

  ![Diagram](image)

  - What can we say about the type of Eiffel Tower? Paris?

(Paulheim and Bizer, 2014)
Type Inference: SD-Type Algorithm

For each property, compute statistical distributions:

- The number of instances of certain types in the subject and object position. For example:

  Type distribution of the property `dbo:location`, i.e. statistics for the triples of the form: (subject `dbo:location` object)

<table>
<thead>
<tr>
<th>rdf:type</th>
<th>subject, %</th>
<th>object, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>dbo:Place</td>
<td>69.8</td>
<td>87.6</td>
</tr>
<tr>
<td>dbo:Building</td>
<td>34.0</td>
<td>0.0</td>
</tr>
<tr>
<td>dbo:City</td>
<td>0.0</td>
<td>24.2</td>
</tr>
</tbody>
</table>

Table: type distribution of the property `dbo:location` (Paulheim and Bizer, 2014)
Type Inference: SD-Type Algorithm

Given: (dbr:Eiffel_Tower dbo:location dbr:Paris)
When observing ?x dbo:location ?y, assign probabilities:

• $P(\text{?x rdf:type dbo:Place}) = 0.698$
• $P(\text{?x rdf:type dbo:Building}) = 0.34$
  • i.e. $P(\text{dbr:Eiffel_Tower rdf:type dbo:Building}) = 0.34$

<table>
<thead>
<tr>
<th>rdf:type</th>
<th>subject, %</th>
<th>object, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>dbo:Place</td>
<td>69.8</td>
<td>87.6</td>
</tr>
<tr>
<td>dbo:Building</td>
<td>34.0</td>
<td>0.0</td>
</tr>
<tr>
<td>dbo:City</td>
<td>0.0</td>
<td>24.2</td>
</tr>
</tbody>
</table>

Table: type distribution of the property dbo:location  
(Paulheim and Bizer, 2014)
Type Inference: SD-Type Algorithm

Given: (dbr:Eiffel_Tower dbo:location dbr:Paris)

When observing ?x dbo:location ?y, assign probabilities:

- $P(?y \text{ rdf:type dbo:Place}) = 0.876$
- $P(?y \text{ rdf:type dbo:City}) = 0.242$
- i.e. $P(dbr:Paris \text{ rdf:type dbo:Place}) = 0.876$

<table>
<thead>
<tr>
<th>rdf:type</th>
<th>subject, %</th>
<th>object, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>dbo:Place</td>
<td>69.8</td>
<td>87.6</td>
</tr>
<tr>
<td>dbo:Building</td>
<td>34.0</td>
<td>0.0</td>
</tr>
<tr>
<td>dbo:City</td>
<td>0.0</td>
<td>24.2</td>
</tr>
</tbody>
</table>

Table: type distribution of the property dbo:location

(Paulheim and Bizer, 2014)
Type Inference: SD-Type Algorithm

- Given resource r with a set of properties Prop, how likely is that this resource is of type t?

\[
conf(t(r)) = \frac{1}{|Prop|} \sum_{prop \in Prop} P(t(r)|(\exists prop. t) (r))
\]

- prop may be incoming or outgoing property

- Example:
  - given r= **dbr:Eiffel_Tower** and a statement
    - (dbr:Eiffel_Tower dbo:location dbr:Paris)
  - Infer:
    - the type t of **dbr:Eiffel_Tower**

- Conditional probabilities are obtained directly from statistical distributions

(Paulheim and Bizer, 2014)
Type Inference: SD-Type Algorithm

- Problems:
  - Not all properties are equally informative for resource type prediction.
  - For example, `rdfs:label` and `owl:sameAs` are more frequent, and independent of the resource type.
  - Some types can dominate the knowledge base (skewed distribution)

- Property weighting:
  \[
  w_{prop} = \sum_{all\ types\ t} (P(t) - P(t|(\exists prop. t)))^2
  \]
  - \(P(t)\) is an apriori probability of the type \(t\).
  - The weight measures the deviation of the property from apriori type distribution.

\[
conf(t(r)) = \frac{1}{\sum_{prop \in Prop} w_{prop}} \sum_{prop \in Prop} w_{prop} * P(t(r)|(\exists prop. t) (r))
\]
(Paulheim and Bizer, 2014)
Outlook: Dataset Profiles & Dataset Search

Open data portals

- Centralized repositories with fixed categories and basic keyword and faceted search
- Scope is currently restricted to returning (RDF) datasets
- Queries issued on data portals differ from those issued to web search engines in terms of length and structure

Future directions

- Better support dataset profile features in dataset search
- Define dataset profiles that better support dataset search applications

Examples:

https://datahub.io/search
https://figshare.com/
https://zenodo.org/
Outlook: Research Questions in Dataset Search

- Understanding how users search for data
  - How do we know that query intent is to find data?
  - How do users search for data?
- Improving the quality of the data
  - How do we understand the provenance of the data?
  - When are two datasets the same? Similar?
  - How do we know which data is trustworthy?
- Learning metadata from the data
  - Can we learn what the semantics of the data is?
  - What does a particular dataset represent?
  - Learn from similar data deposited in a different repository?
- Linking data extracted from PDFs to structured datasets associated with the paper

Natasha Noy, Google, keynote at the SAVE-SD workshop, 2018
Leibniz Symposium “Machine Learning – Intelligent Digitization”

Date: 24.5.2018
Location: Main Building
Leibniz Universität Hannover

Presentations from research and industry, on Machine Learning Methods and ML for Intelligent Mobility, Production and Health

Market place of the future, more than 40 projects about many different ML applications

Expected attendance: 200 - 300 persons

https://machine-learning.ama-academy.eu/
References and Further Reading

[1] Describing Linked Datasets with the VoID Vocabulary https://www.w3.org/TR/void/


Thank you!

Questions, Comments?

Dr. Elena Demidova

L3S Research Center
Leibniz University of Hannover

email: demidova@L3S.de
www: https://demidova.wordpress.com