Deliverable from the COST Action CA19134 “Distributed Knowledge Graphs”

Teaching Curriculum

Due date: 30 March 2024

Edited by: Tobias Käfer (DE) and Eleni Ilkou (DE)

cost
EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY

With contributions from:
Amr Azzam (AT), Anastasia Dimou (BE), Andras Micsik (HU), Andreas Harth (DE), Andrei Olaru (RO), Antoine Zimmermann (FR), Chang Sun (NL), Chiara Ghidini (IT), Christian Esposito (IT), David Chaves-Fraga (ES), Edlira Vakaj (UK), Eleni Ilkou (DE), Giorgos Flouris (EL), Ioannis Andreadis (EL), Ioannis Chrysakis (BE), Johannes Theissen-Lipp (DE), John McCrae (IE), José Emilio Labra Gayo (ES), Kārlis Čerāns (LT), Katja Hose (DK), Manuel Paneque (ES), Marco Hauff (DE), Marcos Da Silveira (LU), María del Mar Roldán (ES), Michel Dumontier (NL), Neki Frasheri (AL), Nuno Cruz Garcia (PT), Olaf Hartig (SE), Özgür Gümüş (TR), Rob Brennan (IE), Roman Lewandowski (PL), Ross Horne (LU), Sadok Ben Yahia (EE), Samir Omanović (BA), Sijin Cheng (SE), Stevan Gostojić (RS), Tobias Käfer (DE), Uldis Bojārs (LT)

This deliverable is based upon work from COST Action "Distributed Knowledge Graphs", supported by COST (European Cooperation in Science and Technology).

COST (European Cooperation in Science and Technology) is a funding agency for research and innovation networks. Our Actions help connect research initiatives across Europe and enable scientists to grow their ideas by sharing them with their peers. This boosts their research, career and innovation.

www.cost.eu

Preface

This deliverable has been compiled by the network of the COST Action "Distributed Knowledge Graphs". This COST Action is a research and innovation network to connect research initiatives around the topic of Distributed Knowledge Graphs. As part of the work during the COST Action, the network compiled a teaching curriculum for its field of research.

The work on this curriculum has been mainly carried out during our plenary meeting of the third Grant Period in September 2023, followed by a virtual session in January 2024. Further work has been carried out during a short-term scientific mission (STSM) in May 2024. This has been followed by further editing by the Action Chair. Parts of the STSM output are going to be presented as a poster at the 23rd International Semantic Web Conference (ISWC).

We identified topics by soliciting from the network the topics they teach in the field of Knowledge Graphs. In further sessions, we refined and detailed out the topics, and reflected on the issue of decentralisation and distribution. In between and afterwards, we clustered the topics in order to generate a condensed overview.

Most of the topics are of technical nature and cover standards, methods, and algorithms to work with knowledge graphs.

In this deliverable, we present this condensed overview.

While not manifested in the headlines of this document, the notion of decentralisation and distribution is woven into the technologies and practices discussed: Technologies to embrace distribution and decentralisation can be found in the Web topic, federated SPARQL, methods for working with data from decentralised and distributed sources, e.g. using formal reasoning and machine learning. The practices underlying the data modelling and publishing also embrace decentralisation and distribution: the Linked Data principles, the FAIR principles, standardisation and community practices, all contain the social fabric for making Distributed Knowledge Graphs work.

We hope that this deliverable helps the larger community to inspire and to improve their teaching practices, and assists in the development of new and existing courses.

Karlsruhe (DE) and Hannover (DE)

Tobias Käfer
Eleni Ilkou

Introduction

We group the topics identified into two classes: Core topics are topics that are taught by almost the entire DKG network. Elective topics are topics that are taught by fewer people from the network. The elective topics are roughly ordered by how many people teach at least a subtopic of the given topic in descending order. For each topic, we list a set of sub-aspects that some people teach. The list is non-exhaustive.

Core Topics

Knowledge Representation using the Resource Description Framework (RDF)

- RDF syntax (Turtle, RDF/XML, JSON-LD) and abstract concepts

Querying RDF using SPARQL (the SPARQL Protocol and Query Language)

- Syntax of the SPARQL query language

Elective Topics

Advanced Knowledge Representation and Reasoning using vocabularies and ontologies

- Vocabularies and semantics
    - OWL (Web ontology language) vocabulary, profiles, and semantics. Description logics.
    - RDFS (RDF Schema) vocabulary and RDF(S) semantics
    - Domain-specific and general-purpose vocabularies (friend-of-a-friend - FOAF, Dublin Core - DC, Data Catalogue - DCAT, Fast Healthcare Interoperability Resources - FHIR, schema.org, …)
- Reasoning techniques, e.g.
    - Rules
    - Backward chaining, forward chaining
- Ontology Engineering and related topics
    - Ontology 101
    - Ontology design patterns
    - Ontology alignment and merging

Knowledge Graphs on the Web

- Linked Data and the Linked Data Principles
- Web architecture, Representational State Transfer (REST)
- Linked Data Platform
- Solid (authentication, authorisation)

Knowledge Graph Governance, Practices, and Quality

- FAIR principles (findable, accessible, interoperable, reusable)
- Governance models and policies for maintaining KGs
- Knowledge Graph publishing practices (see also Knowledge Graphs on the Web)
- Techniques for assessing and improving data quality in KGs.
- Open standardisation practices in by the communities at IETF and W3C
- Social semantics
- Data provenance
- Compliance with regulations (e.g., the General Data Protection Regulation - GDPR).

Advanced Knowledge Graph Querying

- SPARQL Semantics
- SPARQL Protocol
- Federated SPARQL
- SPARQL + Entailment
- RDF and Property Graphs

Knowledge Graph Construction

- From structured data - Ontology-based Data Access (OBDA), RDF mapping language (RML), …
- From text - see Knowledge Graphs from Natural Language Processing and Machine Learning
- Refinement and interlinking - see Knowledge Graphs from Natural Language Processing and Machine Learning and Knowledge Graph Governance, Practices, and Quality

Knowledge Graphs from Natural Language Processing and Machine Learning

- Named entity recognition, disambiguation, extraction and linking
- Relation extraction, Knowledge Graph population, knowledge extraction
- Knowledge graph completion, link prediction
- Knowledge Graph embeddings

Knowledge Graph Applications, Systems, and Tools

- Concrete case studies (e.g. from industry)
- Generic use cases specific to verticals (e.g. healthcare, e-commerce).
- Practical examples of generic KG-based systems (e.g. fact checking, question answering)
- KG tools and libraries to build KG-based systems (e.g. triple stores, APIs in programming languages, ontology editors)
- Architectures of KG-based systems (ETL, OBDA, …)
- General-purpose tools applied to Knowledge Graphs (e.g. visualisation)

Graph Theory

- Introduction to graph theory concepts (nodes, edges, etc.)
- Types of graphs (property, directed, etc.)
- Graph algorithms for KGs (e.g., community detection, centrality measures)
- Graph traversal techniques