# Tutorials

## A Beginner’s Guide to Reasoning: How to reason your way to better data

#### Outline

Reasoning has become an increasingly valued tool in the world of connected data, and yet to many it’s still a black box solution. Perhaps more tragically, despite the explosion of its development in recent years, many still perceive it as a slow, cumbersome, and ultimately impractical technology, which could not be further from true today. Whether you’re looking to harness reasoning for your own goals, or to peek behind the curtains of someone else’s solution, now is your time to learn. Get hands on with a reasoning engine in this interactive walkthrough: A Beginner’s Guide to Reasoning.

You’ll come away understanding the power of reasoning, what it can add to your data, and the fundamentals of how to apply it yourself. With technology in this space running away, there’s never been a better time to learn!

This tutorial will touch on the basics of SPARQL, OWL, and Datalog, before diving into reasoning at a technical level. Each participant will come away having built a reasoning solution for themselves, guided along the way by knowledge engineers and subject experts. No prior knowledge is required.

#### Website

[www.oxfordsemantic.tech/events/semantics-2023-a-beginners-guide-to-reasoning](http://www.oxfordsemantic.tech/events/semantics-2023-a-beginners-guide-to-reasoning)

#### Organized by

* **Peter Crocker** (CEO, Oxford Semantic Technologies)
* **Valerio Cocchi** (Senior Knowledge Engineer, Oxford Semantic Technologies)
* **Thomas Vout** (Marketing Executing, Oxford Semantic Technologies)

## The Key to Sustainable Enterprises: ESG, Knowledge Graphs, and Digitalization

#### Outline

Environmental - Social - Corporate Governance (ESG) is a framework that helps corporations measure their behavior in relation to global sustainability issues. In the next coming years, initiatives and regulations will keep organizations busy with ensuring their business remains not only compliant with ESG directives and laws but also competitive and socially responsible. In simple terms, ESG compliance will be a necessary task for all organizations.

Considering the complexity of the topic and the vast amounts of data consumed within this domain, ESG requires the use of a dedicated knowledge management strategy and systems.

This tutorial/workshop will explain why Semantic AI and knowledge graphs should have a fundamental role in ESG strategies and will:

* Introduce ESG and cover the challenges associated with it
* Survey a collection of practical use cases
* Preview a core ESG knowledge model
* Explain how the knowledge model can be extended into a knowledge hub, among other solutions.

The tutorial/workshop will provide a deeper look into selected case studies related to ESG that have been successfully rolled out with semantic technologies. Themes that will be addressed include the practical application of knowledge graphs, the need for digitization and collaboration, and ESG as an opportunity for innovation.

#### Website

https://www.poolparty.biz/blogposts/semantics-2023-esg-knowledge-graphs

#### Organized by

* **Heather Hedden** ([Linkedin profile](https://www.linkedin.com/in/hedden/))
* **Andreas Blumauer** ([Linkedin profile](https://www.linkedin.com/in/andreasblumauer/))
* **Dr. Lutz Krüger** ([Linkedin profile](https://de.linkedin.com/in/dr-lutz-krueger))

## Knowledge Engineering of Taxonomies and Ontologies

#### Outline

Semantic technologies support linking and connecting data and enable the targeted retrieval of information. This requires knowledge organization systems, especially taxonomies and ontologies, whereby concepts are tagged to content or linked to data and linked to each other. While published knowledge organization systems exist, it's best to have a knowledge organization system that is custom-created to serve the needs of its users and reflect specific data and content. The goal of this tutorial is to teach participants how to design and build such knowledge organization systems.

This tutorial teaches the fundamentals and best practices for creating quality taxonomies and simple domain ontologies, based on Semantic Web standards, whether for the enterprise or for specific knowledge bases in any industry. Emphasis will be on serving users and use cases, rather than on theory.

Topics to be covered:

* Kinds of knowledge organization systems (taxonomies, thesauri, ontologies, etc.)
* Kinds of knowledge organization differences and their benefits and suitability for different purposes
* Semantic Web standards and other standards
* Taxonomy/thesaurus concept creation
* Taxonomy/thesaurus relationship creation
* Ontology modeling basics for classes, relations and attributes
* Linking knowledge organization systems
* Enriching a taxonomy or thesaurus to become an ontology and a knowledge graph

#### Website

<http://www.hedden-information.com/courses-workshops/conference-taxonomy-workshops/>

#### Organized by

* **Heather Hedden** ([Linkedin profile](https://www.linkedin.com/in/hedden/))

## Modular Approach to Solve Problems at Scale with Healthcare NLP

#### Outline

Despite ongoing efforts towards using natural language processing (NLP) in information extraction from electronic health records (EHR’s), current solutions require healthcare AI practitioners to make unacceptable trade-offs between delivering state-of-the-art accuracy, generalizing over unseen data points, and preventing the sharing of personal data or intellectual property.

Spark NLP for Healthcare aims to bridge this gap by providing an accurate, scalable, private, and tunable software library that helps healthcare & pharma organizations build longitudinal patient records and knowledge graphs on real-world EHR data.

Spark NLP is a Natural Language Processing (NLP) library built on top of Apache Spark ML. It provides simple, performant & accurate NLP annotations for machine learning pipelines that can scale easily in a distributed environment. Spark NLP comes with 20.000+ pretrained pipelines and models in more than 200+ languages. It supports nearly all the NLP tasks and modules that can be used seamlessly in a cluster. Being downloaded more than 3 million every month and experiencing 20x growth for the last one year, Spark NLP is used by 54% of healthcare organizations as the world’s most widely used NLP library in the enterprise.

In this workshop, attendees will get familiar with the library’s healthcare components and learn how to solve any NLP problem in healthcare with the state-of-the-art methods and practices across the industry. Additionally, they will familiarize with best practices for building production-grade solutions around the latest research, especially the healthcare-specific large language models (LLMs) fine-tuned to solve certain tasks such as clinical notes summarization and knowledge retrievals.

#### Website

<https://www.johnsnowlabs.com/>

#### Tutorial Materials

<https://github.com/JohnSnowLabs/spark-nlp-workshop/tree/master/healthcare-nlp>

#### Organized by

* **Veysel Kocaman**, PhD, Head of Data Science at John Snow Labs ([Linkedin profile](https://www.linkedin.com/in/vkocaman/))

## The Open Research Knowledge Graph - A Lighthouse in the Publication Flood

#### Outline

Science is facing a challenge: In the ever growing flood of publications, maintaining an overview becomes increasingly hard. The underlying issue is that we rely on outdated methods of scholarly communication. Traditional articles in pdf format are written by humans for human consumption and do not allow for machine assistance in finding, comparing and linking relevant research contributions. This document-centered information flow contributes to numerous issues currently seen in science such as the deficiency of peer-review or the reproducibility crisis.

The Open Research Knowledge Graph (ORKG1) aims to provide a solution.

We structure the contents of scholarly publications in a knowledge graph, making them human- and machine-actionable. At its core, the ORKG serves as a central hub for organizing scientific information from scholarly publications, including papers, datasets and software. The contents are curated by researchers in a crowd-based approach.

In this tutorial we will give a brief introduction to ORKG’s vision, followed by instructions on using the platform and a hands-on-session on curating content. We will also discuss methods to enhance the curation process with NLP techniques.

#### Website

<https://orkg.org/page/orkg-tutorial-at-semantics>

#### Organized by

* **Anna-Lena Lorenz**
* **Jennifer D'Souza**
* **Marina Wurzbacher**

## Streamlining License Customization and Clearance with Open DALICC Framework

#### Outline

**What is it about?**

In this tutorial we give an in-depth introduction into the Open DALICC framework (<https://dalicc.net>), current implementation projects and the open developer program accompanying the DALICC community initiative.

In the modern IT landscape, applications increasingly retrieve, store, and process data from a variety of sources to generate derivative works. This often leads to questions about the compatibility of components offered under different licenses and the application's legal compliance. Ensuring this legal compatibility in ICT- enabled, interconnected systems can be time-consuming, expensive, and complex, and is a challenge for both non-legal and legal professionals. To address these issues, we have developed DALICC, a software framework in line with semantic web technologies. This tool enables well-structured and machine-readable representation of software and data licenses. Consequently, this information can be used for a range of tasks, including the creation of custom licenses, license search with specific criteria, and compatibility checks between licenses.

DALICC serves as a comprehensive solution for creating, managing, and clearing licenses in digital asset management. It supports legal experts, innovation managers, and application developers in legally reusing third-party digital sources. The DALICC framework significantly simplifies the clearance of rights, thereby substantially reducing the costs associated with rights clearance during the creation of derivative works. It opens up opportunities for legally secure utilization of third-party digital assets. At its core, DALICC helps to identify which assets can be shared, with whom, and under what conditions, thereby reducing rights clearance costs and stimulating the data economy.

The DALICC framework, its technical components and its datasets are available via GitHub (<https://github.com/dalicc>) and provided under open licenses (MIT / CC-BY / CC-BY-ND) for sharing and reuse.

**Who should participate?**

This tutorial is designed for anyone involved in LegalTech and compliance management from both scientific and industrial fields. License issues are relevant to nearly everyone involved in creating or managing software or data-related products and services. This includes developers, data engineers, library personnel, legal staff, sales personnel, and many others. Participants of this tutorial will learn about the specific problems DALICC can solve, explore use cases, and find a platform to exchange ideas and foster collaboration for further development of the framework, extending beyond just license clearance.

**When & where?**

This will be a half day tutorial taking place in a hybrid form at the Semantics conference.

#### Organized by

* **Tassilo Pellegrini**
* **Giray Havur**