yoast

What is schema.org and why you should care

Paolo L. Scala – WordCamp Italia 2022 – 11, 12 November

Outline

- Context
- The problem
- Semantic Web: the story so far
- Enter schema.org
- An example of implementation
- Key takeaways

Context

Context

- Broad context: SEO
- I have *something* I want to *give* to **others**
- *Something* might be anything
- *Give* in exchange of something else

Context

Trade

- the transfer of goods and services from one person or entity to another (often in exchange for money)
- originated from human communication in prehistoric times

The problem

The problem

Who are the **others**?

- Before
  - Others -> humans
  - Marketplace -> physical
- Now
  - *Others -> machines*
  - *Marketplace -> digital*

The problem

Semantic gap

- The difference in meaning between constructs formed within different representation systems

- How do we bridge the Semantic gap?

Semantic Web

Semantic Web

- A field of research rather than a concrete artifact
- Just like AI
- Can be analysed under three possible perspectives

Semantic Web – Perspective 1

Semantic Web as an enhancement of the current World Wide Web

- Machine-understandable information + services (intelligent agents) consuming it
- Use of metadata
- Fairly old perspective (2001)

Semantic Web – Perspective 2

Semantic Web as a mean to establish efficient tools for data sharing, discovery, integration and reuse

- The WWW may or may not be used as medium
- Methods and tool developed by the field can provide value in absence of intelligent agents consuming the data
- Fairly recent perspective

Semantic Web – Perspective 3

Semantic Web as a mean to investigate applications of ontologies, linked data, and knowledge graphs with the W3C standards RDF, OWL, and SPARQL at its core.

- Way more restrictive than the previous perspectives
- More practical

Semantic Web – The story so far

Conventionally born in 2001:

- Berners-Lee et al. The Semantic Web (May 2001)

Previous efforts:

- 2000-2006 DARPA Agent Markup Language (DAML) program
- 1997 First draft of the Resource Description Framework (RDF)

Semantic Web – The story so far

Three overlapping phases

- 2000-mid 2005s Ontologies
- mid 2005s-2010 Linked data
- 2010s Knowledge graphs

Semantic Web – Ontologies

- A formal, explicit specification of a shared conceptualization
- A knowledge of concepts (types, classes) and their relationships specified in a knowledge representation language based on a formal logic
- A main vehicle for data integration, sharing, and discovery
- Driving idea: reusability

Semantic Web – Ontologies

- 2004 the Web Ontology Language (OWL) became a W3C standard
- based on a description logic designed in such a way that logical deductive reasoning over the language is decidable
- 2004 RDF became a W3C standard
- a syntax for expressing directed, labeled, and typed graphs

Semantic Web – Ontologies

Examples:

- The **Gene Ontology** had its beginnings in 1998 and is now a very prominent resource
- **SNOMED CT**, which can be traced back to the 1960s, now fully formalized in OWL and widely used for electronic health records

Semantic Web – Ontologies

- Difficult to maintain (often based on ad-hoc modeling)
- Hard to reuse reuse (too specific)
- Considerable up-front cost to develop good ontologies (at that time)

Semantic Web – Linked data

- a set of RDF graphs linked together
- many IRI in the graphs appear also in other, sometimes multiple, graphs
- Datasets cover a variety of subjects

Semantic Web – Linked data

- The number of publicly available linked RDF graphs grew significantly during the first decade
- more than 37 billion triples from over 650,000 data documents in 2015
- Information in RDF graphs was kept simple (compared to ontologies)
- Allows for simpler data integration, management and applications

Semantic Web – Linked data

Examples

- DBpedia
- a linked dataset extracted from Wikipedia
- The April 2016 releases covers about six million entities and about 9.5 billion RDF triples

Semantic Web – Linked data

- Wikidata

- based on a similar idea as Wikipedia (crowdsourcing)
- is about creating structured data that can be used by programs or in other projects
- over 66 million data items, over one billion edits since project launch, over 20,000 active users

Semantic Web – Linked data

- Integrating and utilizing linked data took more effort than some initially expected
- Shallow non-expressive schemas appeared to be a major obstacle to reusability
- Interlinks between datasets would not account for this weakness

Semantic Web – Knowledge graph

- Term coined by Google, identifying a proprietary technology
- Accessible by public API
- Mainly a “re-brand” of well-known ideas from the ontologies era, with some differences

Semantic Web – Knowledge graph

- Openness: less open, as many related initiatives and results are industry-led
- Centralized control vs community-driven organisation
- Shift from academic research to use in industry

Enter schema.org

Enter schema.org

A **collaborative**, **open** effort in curating a **shared vocabulary** that **webmasters could use** to **annotate** their websites

- Born in 2011
- Between Linked Data and Knowledge Graph

schema.org – The organization

- Collaborative
- Google, Bing, Yahoo, Yandex
- Open
- Changes, additions and extensions to schema.org are discussed in a dedicate public forum
- Open contribution through schema.org GitHub repository
- Communities of practice can contribute specific extensions to the schema

schema.org – The structure

Shared vocabulary

- 797 Types
- 1453 Properties
- 14 Datatypes
- 86 Enumerations
- 462 Enumeration members

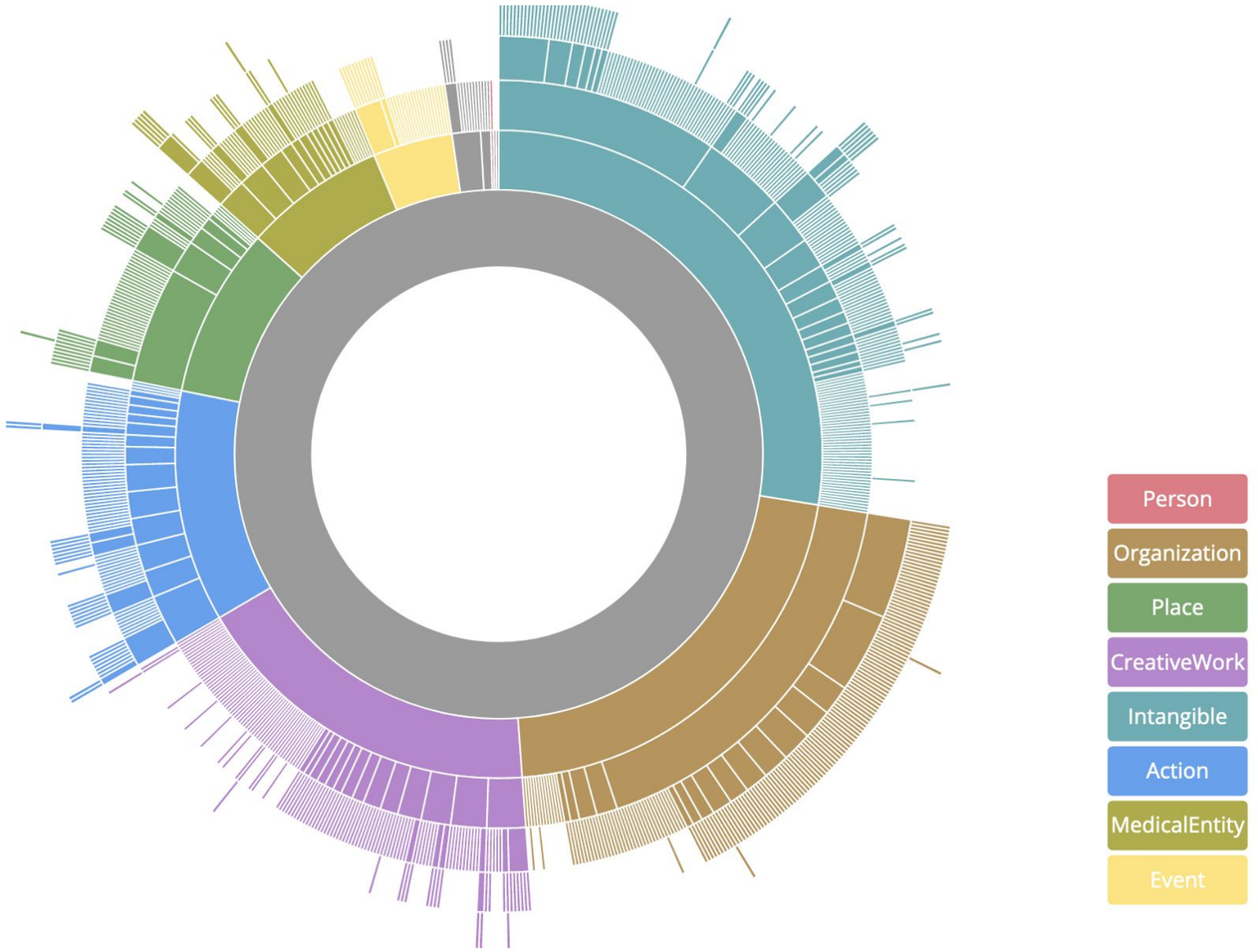
schema.org – The structure

- Types
- Entities such as *Person*, *WebPage*, *Action*…
- Properties
- Attributes attached to actions and describe them through their value

schema.org – The structure

- Datatypes
- Used to express values of properties
- Enumerations
- Collect accepted values for a specific property

schema.org – Entities hierarchy



http://bl.ocks.org/danbri/1c121ea8bd2189cf411c

schema.org – Target audience

Producer point of view:

- Webmasters

- Easy to use to annotate webpages
- Supports multiple syntaxes (RDFa, MicroData, JSON-LD)
- Both machine-friendly and human-friendly

schema.org – Target audience

- Entity reconciliation is a shared effort between humans and machines
- Machines use all the properties available to identify the descriubed entity
- Humans can explicitly set a *SameAs* property
- Add complexity through subsequent iterations

schema.org – Target audience

Consumer point of view:

- Google

- Not only for rich snippets and Knowledge Graph
- More and more advanced tools rely on schema
- (quality) Structured data = access to a wider market
- Schema as Google’s source of truth

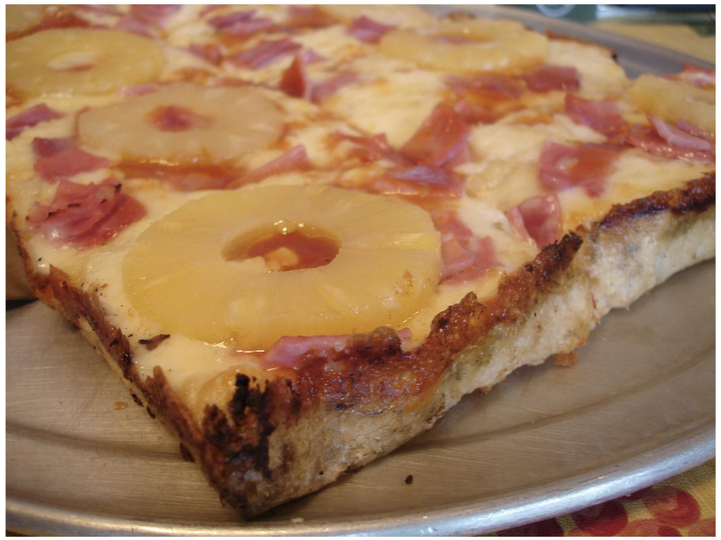
schema.org – Target audience

- Facebook, Pinterest, Instagram
- Influeces how, when, where and **if** your content is offered

A schema example

schema.org – An example

Simple post with image



(Picture: @joefoodie from USA, CC BY 2.0 https://creativecommons.org/licenses/by/2.0, via Wikimedia Commons)

October 15, 2022 Paolo Scala Uncategorized

schema.org – An example


```
{
   "@context": "https://schema.org",
   "@graph": [
      {
         "@type": "ImageObject",
         "inLanguage": "en-US",
         "@id": "https://basic.wordpress.test/simple-post-with-image/#primaryimage",
         "url": "https://basic.wordpress.test/wp-content/uploads/2022/10/Hawaiian_pizza_1.jpg",
         "contentUrl": "https://basic.wordpress.test/wp-content/uploads/2022/10/Hawaiian_pizza_1.jpg",
         "width": 1600,
         "height": 1200
      }
   ]
}
```

schema.org – An example


```
{
    "@context": "https://schema.org",
    "@graph": [
        {
            "@type": "WebPage",
            "@id": "https://basic.wordpress.test/simple-post-with-image/",
            "url": "https://basic.wordpress.test/simple-post-with-image/",
            "name": "Simple post with image - Basic"
        },
        {
            "@type": "ImageObject",
            "inLanguage": "en-US",
            "@id": "https://basic.wordpress.test/simple-post-with-image/#primaryimage",
            "url": "https://basic.wordpress.test/wp-content/uploads/2022/10/Hawaiian_pizza_1.jpg",
            "contentUrl": "https://basic.wordpress.test/wp-content/uploads/2022/10/Hawaiian_pizza_1.jpg",
            "width": 1600,
            "height": 1200
        },
    ]
}
```

schema.org – An example


```
{
    "@context": "https://schema.org",
    "@graph": [
        {
            "@type": "WebPage",
            "@id": "https://basic.wordpress.test/simple-post-with-image/",
            "url": "https://basic.wordpress.test/simple-post-with-image/",
            "name": "Simple post with image - Basic",
            "primaryImageOfPage": {
                "@id": "https://basic.wordpress.test/simple-post-with-image/#primaryimage"
            },
            "image": {
                "@id": "https://basic.wordpress.test/simple-post-with-image/#primaryimage"
            },
        },
        {
            "@type": "ImageObject",
            "inLanguage": "en-US",
            "@id": "https://basic.wordpress.test/simple-post-with-image/#primaryimage",
            "url": "https://basic.wordpress.test/wp-content/uploads/2022/10/Hawaiian_pizza_1.jpg",
            "contentUrl": "https://basic.wordpress.test/wp-content/uploads/2022/10/Hawaiian_pizza_1.jpg",
            "width": 1600,
            "height": 1200
        }] }
```

schema.org – An example


```
{
    "@context": "https://schema.org",
    "@graph": [
        {
            "@type": "WebPage",
            "@id": "https://basic.wordpress.test/simple-post-with-image/",
            "url": "https://basic.wordpress.test/simple-post-with-image/",
            "name": "Simple post with image - Basic",
            "isPartOf": {
                "@id": "https://basic.wordpress.test/#website"
            },
            "primaryImageOfPage": {
                "@id": "https://basic.wordpress.test/simple-post-with-image/#primaryimage"
            },
            "image": {
                "@id": "https://basic.wordpress.test/simple-post-with-image/#primaryimage"
            },
        },
        {
            "@type": "WebSite",
            "@id": "https://basic.wordpress.test/#website",
            "url": "https://basic.wordpress.test/",
            "name": "Basic",
            "description": "Just another WordPress site",
        }
    ]
}
```

schema.org – An example


```
{
    "@context": "https://schema.org",
    "@graph": [
        {
            "@type": "WebPage",
            "@id": "https://basic.wordpress.test/simple-post-with-image/",
            "url": "https://basic.wordpress.test/simple-post-with-image/",
            "name": "Simple post with image - Basic",
            "isPartOf": {
                "@id": "https://basic.wordpress.test/#website"
            },
            "primaryImageOfPage": {
                "@id": "https://basic.wordpress.test/simple-post-with-image/#primaryimage"
            },
            "image": {
                "@id": "https://basic.wordpress.test/simple-post-with-image/#primaryimage"
            },
            "author": {
                "@id": "https://basic.wordpress.test/#/schema/person/948ea13d5"
            },
        },
```

schema.org – An example


```
{
    "@type": "Person",
    "@id": "https://basic.wordpress.test/#/schema/person/948ea13d5",
    "name": "Paolo Scala",
    "image": {
        "@type": "ImageObject",
        "inLanguage": "en-US",
        "@id": "https://basic.wordpress.test/#/schema/person/image/",
        "url": "https://secure.gravatar.com/avatar/6edc8cc7190",
        "contentUrl": "https://secure.gravatar.com/avatar/6edc8cc7190",
              "caption": "Paolo Scala"
    },
    "url": "https://basic.wordpress.test/author/pls78/"
}
```

schema.org – An example


```
{
    "@context": "https://schema.org",
    "@graph": [
        {
            "@type": "WebPage",
            "@id": "https://basic.wordpress.test/simple-post-with-image/",
            "url": "https://basic.wordpress.test/simple-post-with-image/",
            "name": "Simple post with image - Basic",
            "isPartOf": {
                "@id": "https://basic.wordpress.test/#website"
            },
            "primaryImageOfPage": {
                "@id": "https://basic.wordpress.test/simple-post-with-image/#primaryimage"
            },
            "image": {
                "@id": "https://basic.wordpress.test/simple-post-with-image/#primaryimage"
            },
            "author": {
                "@id": "https://basic.wordpress.test/#/schema/person/948ea13d5"
            },
```

schema.org – An example

```
"potentialAction": [
    {
        "@type": "ReadAction",
        "target": [
            https://basic.wordpress.test/simple-post-with-image/"
              ]
    }
  ]
},
{
    "@type": "WebSite",
    "@id": "https://basic.wordpress.test/#website",
    "url": "https://basic.wordpress.test/",
    "name": "Basic",
    "description": "Just another WordPress site",
    "potentialAction": [
    {
        "@type": "SearchAction",
        "target": {
            "@type": "EntryPoint",
            "urlTemplate": "https://basic.wordpress.test/?s={search_term_string}"
        },
        "query-input": "required name=search_term_string"
    }
  ] }
```

schema.org – …but how?

How to know which types/properties are needed?

- Google General structured data guidelines
- https://developers.google.com/search/docs/appearance/structured-data/sd-policies
- Technical/quality guidelines about how to implement structured data

schema.org – …but how?

- Google Search Gallery
- https://developers.google.com/search/docs/appearance/structured-data/search-gallery
- Required/recommended properties for each relevant type

schema.org – …but how?

- schema.org vocabulary reference
- https://schema.org/docs/schemas.html
- Useful reference for advanced users

schema.org – Useful Tools

- https://classyschema.org/Visualisation
- Great validation and schema visualization tool
- Very fast
- Interactive schema navigation

schema.org – Useful tools

- https://search.google.com/test/rich-results
- Google tool to test schema
- tests if page supports rich results
- Slow

schema.org – Useful Tools

- https://validator.schema.org/
- Official schema.org validator
- Simple visualization

schema.org – Implementation

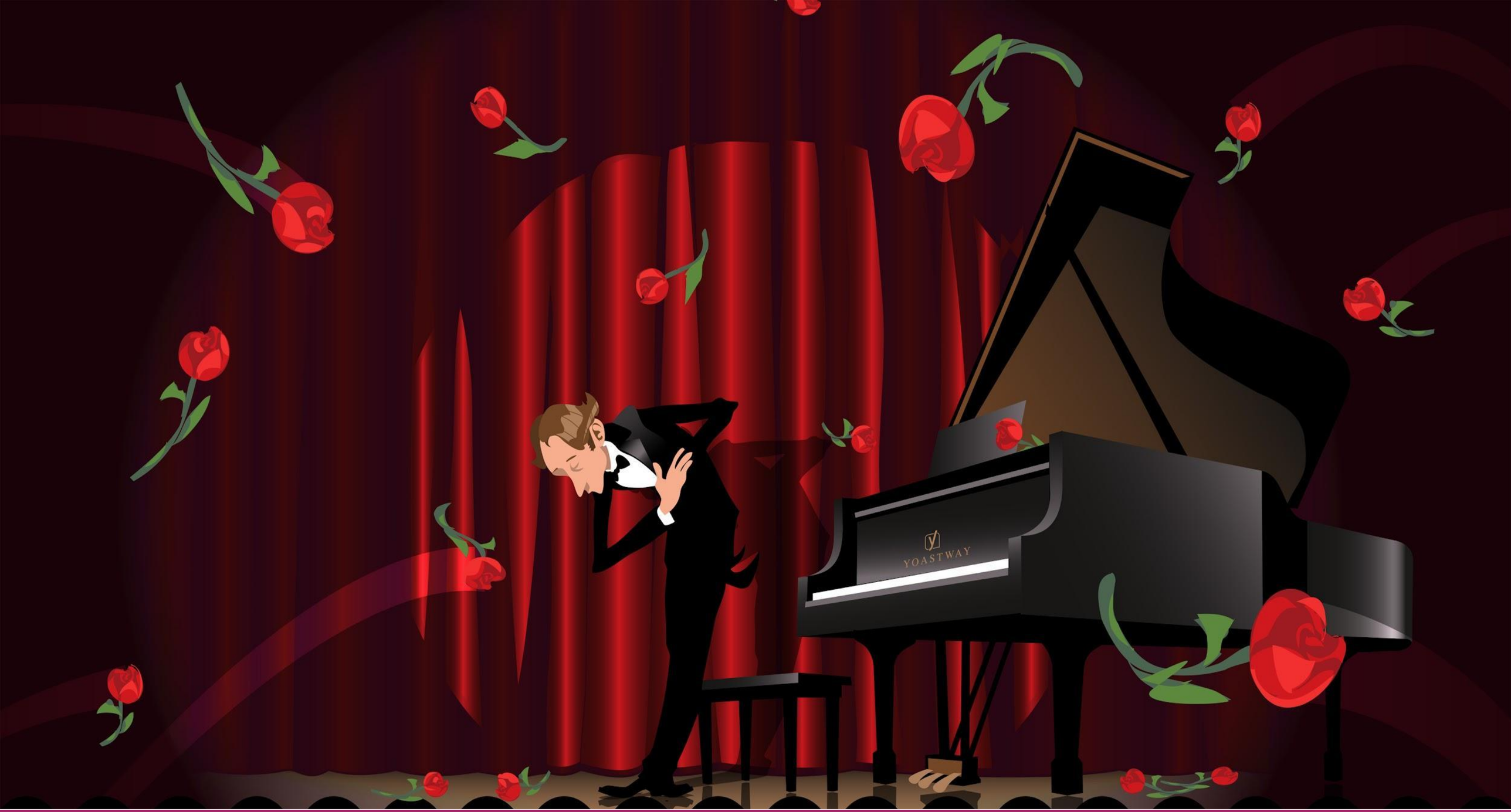
[Visual Studio Code Interactive Session]

Key take-aways

- Schema.org helps bridging the Semantic gap
- Is already important and will become more and more important
- Is much more than rich snippets

Key take-aways

- Is not only about Google
- Simple to understand
- Difficult to get it right


YOASTWAY

References

- The Semantic Web – T. Berners-Lee, J. Hendler, O. Lassila - Scientific american, 2001 – JSTOR
- Schema. org: evolution of structured data on the web – R.V. Guha, D. Brickley, S. Macbeth - Communications of the ACM, 2016 – https://dl.acm.org
- A review of the semantic web field – P. Hitzler - Communications of the ACM, 2021 – https://dl.acm.org
- A deep dive into schema.org and structured data for SEO – J. Alderson - https://wordpress.tv/2020/11/27/jono-alderson-a-deep-dive-into-schema-org-and-structured-data-for-seo/
- Schema.org - What, How, Why? - C. McCathieNevile – Open Web Camp VI, 2014 –
- https://youtu.be/hcahQfN5u9Y