Schema App

Guide to Entities & Knowledge Graphs for SEO

Introduction

THE WAY WE SEARCH HAS EVOLVED SIGNIFICANTLY SINCE THE ADVENT OF THE INTERNET.

From typing keywords into a desktop computer to using voice search on a mobile or smart home device, we can access information at any moment, from anywhere.

Over the past decade, search engines have evolved from lexical to semantic search. Instead of using keyword matching and rudimentary algorithms to rank results, search engines now use complex AI-powered algorithms to understand the concepts and meaning in a query and on a page to provide relevant search results.

Generative Artificial Intelligence (AI) is also disrupting the world of search by changing the way information is presented on the search engine results page (SERP). These AI search engines provide summarized answers, follow-up questions, and relevant information that the user might want. When this happens, the user's query may be answered resulting in a zero-click search, or they may want to explore more on the presented relevant sites.

What we do know, is that in order to stay relevant in this new AI search experience, you must clearly communicate the meaning and intent behind your content. To effectively communicate in this new search language, you must clearly define your entities and showcase the connections between your entities through a knowledge graph using Schema Markup. That way, search engines can derive meaningful context from your content and provide more accurate search results.

In this eBook, we delve into the impact of entities and knowledge graphs on your SEO strategy and guide you through the process of defining and linking the entities on your site to construct a robust reusable knowledge graph using Schema Markup.

Let's dive in.

What are Entities?

Definition of an Entity

‘An entity is a thing or a concept that is singular, unique, well-defined and distinguishable.’

- Google

An entity can be a person, place, organization, item, abstract concept, concrete element, or any combination thereof.

They *take up space* (be it physical, digital or conceptual), have *attributes* (like size, weight or prices) and most importantly, are *understood in relation to other things*.

Let's consider the example of "faralang."

This is a string of characters that means nothing to the human brain, and therefore won’t mean anything to a machine either. But if we were to name a new band “Faralang”, this jumble of letters now becomes an identifiable entity. Entities need to be described to have any meaning.

As humans, we use context clues to make sense of new things and search engines do the same thing. That said, machines and search engines are not as good at understanding things in the same way that a human brain can.

To improve the search engine’s understanding of language, Google uses natural language processing models like BERT to analyze the proximity and frequency of certain terms, phrases and entities in search queries and the content on your pages. But the technology is far from perfect.

FOR ONE, WORDS CAN BE AMBIGUOUS.

In the English language, certain words have multiple meanings. For example, the word ‘Apple’ is both a fruit and a brand. Without additional context, it can be hard for the natural language processing models to disambiguate which “Apple” the searcher refers to.

Apple (Fruit) **or** **Apple (Brand)**

Language differences also pose a challenge to natural language processing models. Different languages tend to phrase things differently. For example, the term ‘rebord de fenêtre’ in French translates directly to ‘edge of window’ in English. But it is actually referring to a windowsill.

Luckily, there are ways to make statements about entities more explicit for search engines.

WHY ARE ENTITIES IMPORTANT IN SEO?

To provide users with more accurate search results, search engines are now more semantic. They look at the meaning and intent in a search query and on a page and try to provide searchers with results that are most relevant to their query or research journey.

Entities play an important role in this semantic understanding. They can be accurately defined and distinguished, with details of their relationship to other things on the web or within a company offering layers of context to search engines. Organizing data and establishing clear connections between entities is vital for search engines and machines to understand your content.

By incorporating and connecting entities, websites can establish clear connections and relationships between different pieces of information. This interconnected web of entities forms the foundation of knowledge graphs, which not only help search engines understand the content within a website but also establish connections with external databases and authoritative sources of information.

When search engines comprehend content in this detailed manner, they can provide more accurate and relevant search results to users. This leads to improved user experience, as searchers receive results that closely match their query intent.

In essence, entities and knowledge graphs enable search engines to deliver precise, contextually rich, and highly relevant results. Let’s dive deeper into knowledge graphs and their role in SEO.

What are Knowledge Graphs?

Definition of a Knowledge Graph

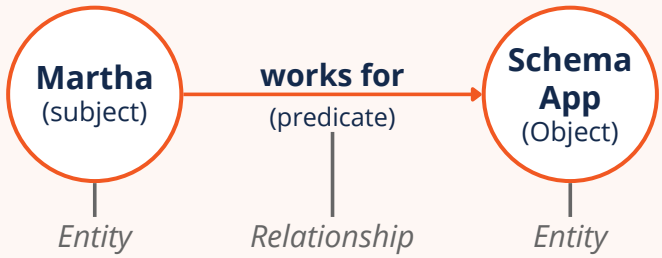
A knowledge graph is a collection of relationships between entities defined using a controlled vocabulary, from which new knowledge may be gained by way of inferencing.

In technical terms, a knowledge graph is a structured knowledge representation that leverages an ontology to express relationships between entities in the form of Resource Description Framework (RDF) triples.

WHAT IS AN RDF TRIPLE?

An RDF Triple is a framework used to express data as a directed graph using **subject-predicate-object** statements known as triples. They help machines understand the relationship between entities. By combining these triples, vast interconnected graphs of resources can be created.

For example, Martha van Berkel and Schema App are both entities. Both entities are related to each other because Martha van Berkel works for Schema App.



These triples are typically represented in a machine-readable format like JSON-LD to ensure information is organized and accessible for machine processing. In SEO, we often use JSON-LD to express the entities and connections in your content using the Schema.org vocabulary.

Schema App

ABOUT THE SCHEMA.ORG VOCABULARY

Schema.org is an initiative that emerged in 2011 as a collaborative effort between Google, Bing, Yahoo, and Yandex.

At its core, Schema.org bridges the gap between human language and machine understanding through the creation of a standardized vocabulary, aptly named the Schema.org vocabulary.

When websites express their content using the Schema.org vocabulary in formats like JSON-LD, it transforms into Schema Markup. The Schema.org vocabulary currently consists of over 800 Types and 1500 properties.

Schema.org Types

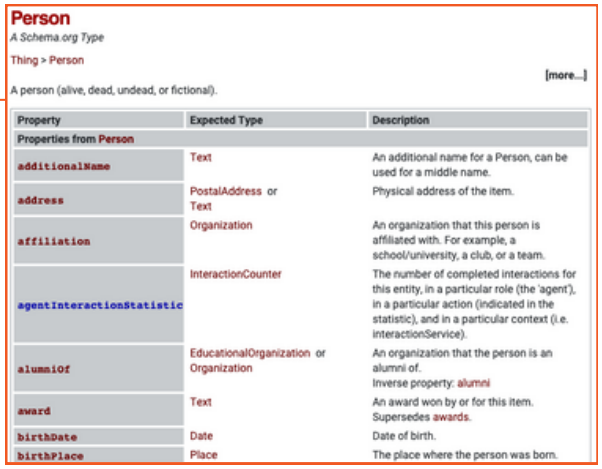
A Schema.org Type categorizes entities as specific kinds of things.

For example, the 'Person' type is used to represent an individual that could be alive, dead, undead or fictional, while the 'Product' type is used to present an entity that is available for purchase.

Schema.org Properties

A Schema.org property is an attribute or characteristic that provides additional details and contextual information about a given entity.

For example, the 'Person' type can use properties like 'name', 'date of birth', and 'address'.



By using these properties, you are providing machines with specific information about an individual. These properties describe not only the qualities of the entity but also its relationships with other entities.

LET'S ILLUSTRATE THIS USING AN EXAMPLE.

1

The content on this web page tells people that the Schema App Highlighter (an entity) is from the brand Schema App (an entity).

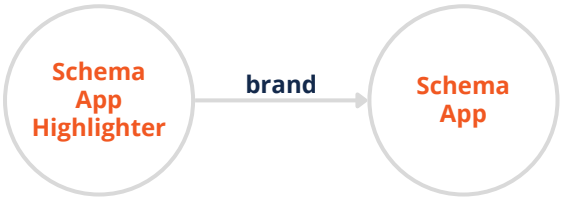
2

That content is translated into Schema Markup and expressed in JSON-LD telling search engines that the Schema App Highlighter is a product from the brand Schema App.

```
<script type="application/ld+json">
{
   "@context": "http://schema.org",
   "@type": "Product",
   "name": "Schema App Highlighter",
   "@id":
"https://schemaapp.com/highlighter/#Product",
   "brand":{
       "@type": "Brand",
       "@id": "https://schemaapp.com/#Brand",
       "name": "Schema App"
    }
}
</script>
```

3

Schema Markup, expressed in the form of JSON-LD helps search engines understand this content as a connected graph of RDF triples.



When you connect the different entities on your site, you are effectively developing an internal knowledge graph about your organization.

OTHER KNOWLEDGE GRAPHS IN SEO

The concept of knowledge graphs has been around for a long time. Social media platforms like Facebook and Twitter have leveraged graph technology to gain insights into users' interests, hobbies, preferences and social connections. Cultural heritage institutions like libraries and museums have also turned to graph technology to structure their information in a meaningful way.

Some external knowledge graphs like Google's knowledge graph, Wikipedia and Wikidata are also useful for search engine optimization (SEO) purposes. SEO teams can connect their internal knowledge graphs to external knowledge graphs to help search engines disambiguate the entities on their website.

IMPORTANCE OF KNOWLEDGE GRAPHS IN SEO

Improve Search Engine's Semantic Understanding

Knowledge graphs are a collection of relationships between the entities on your site and beyond. Therefore, your knowledge graph can provide search engines with more contextual information about your content. With a better semantic understanding of your content, search engines can better match your page to relevant user queries.

Create a Control Point for AI Search Engines

Over this past year, we've also seen generative AI-powered search features like Google's Search Generative Experience, ChatGPT and the New Bing adding complexity to the search landscape. These new search features use Large Language Models (LLMs) to summarize complex search results into contextual responses. Despite the efficiencies and benefits it offers, LLMs also have their challenges.

Much of the content on the internet is formatted as unstructured data (i.e. plain text, images, videos, etc.), and search engines use LLMs to try to understand, summarize, translate, predict and generate new content from this data.

However, LLMs are black-box models often falling short of capturing and accessing factual information. Without the ability to perform a fact check, they are prone to hallucinations and end up providing inaccurate answers. You don't want Google to provide your users with inaccurate information about your organization, do you?

Fortunately, knowledge graphs can overcome these issues because they are structured knowledge models that explicitly store domain-specific, timely, rich, factual information about your organization. This makes it easier for AI search engines to consume and infer factual information about your organization while also giving you a control point for how you want your content understood.

Entities are fundamental to the development of your knowledge graph. Your knowledge graph can ground AI search engines with factual information about your organization. **In the next section of this guide, we'll explore how you can define and connect the entities on your site to develop your knowledge graph.**

Interested in diving deeper into knowledge graphs?

Enroll in our free Content Knowledge Graph Fundamentals course to learn more about the anatomy of a knowledge graph and more!

Explore the course

How to Define and Connect the Entities On Your Site to Develop Your Knowledge Graph

The process of developing a knowledge graph using your website content involves 3 main steps:

1. Design your website content to talk about specific topics in depth

2. Use Schema Markup to describe the entities in your content and their relationships to other things

3. Link known entities on your page to external knowledge graphs

Schema App

1 Design your content on your website to talk about specific topics in depth

Before you can go about defining your entities and developing your knowledge graph, you need to create the content for it.

Identify the topics you want to be known for in search or topics your audience might be interested in

Instead of looking at keywords that you want to rank for, look at the topics you want to be known for or things you want search engines to know about your organization.

For example, if you run an SEO software company that provides on-page and off-page SEO services, you'd want search engines to know about your organization, the features offered by your software, your employees, and your areas of expertise.

On top of information about your organization and your product, your audience might also be searching for complementary topics like on-page SEO tips or SEO metrics.

But how do you figure out what relevant information a searcher might want to know about a topic?

You can use tools like **Ask the Public** to see relevant questions that people might be searching for around a topic.

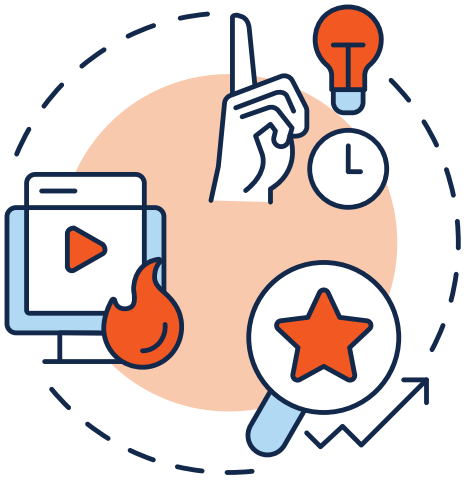
You also use the **Schema.org** vocabulary to see the available properties to make your content more descriptive. This is especially useful for entities like services, physicians, local businesses, products and events.

Once you've identified the topics you want to be known for, you should create high-quality content for those topics and focus on one topic per page.

Focus on one topic per page

The purpose of creating high-quality content is to answer all questions that a searcher might need regarding a specific topic. However, to prevent your pages from cannibalizing each other on the search engine results page, you should focus on one topic per page and include information relevant to your topic to make the content more comprehensive and holistic.

For example, if your company offers multiple types of software, you'd want to create a page for each software.



You might also want to include information about the software such as the:

- price,
- functionalities,
- ratings and reviews, and
- frequently asked questions.

This is also important because we're seeing more natural and conversational search queries with the advent of generative AI search features. People are querying the search engine the same way they'd query other people in a conversation, and they can even ask follow-up questions.

For example, a searcher can query "Show me a female OBGYN near me who accepts Aetna Medicare Advantage insurance".

If your page only includes content stating that the doctor is an OBGYN operating in a location near the patient, without including any information about the insurance they accept, search engines might not deem your page relevant in a search result.

They might retrieve an answer based on multiple sources. However, adding more relevant information about a specific topic on the page can only do your site more good than harm.

2 Use Schema Markup to describe the entities in your content and their relationships to other things

Once you've created and published the content on your site, it's time to translate your unstructured content into machine-readable code like Schema Markup.

WHAT IS SCHEMA MARKUP?

Schema Markup, also known as structured data, is a standardized vocabulary used by search engines to understand the content on your web pages.

By implementing Schema Markup on your website, the data in the form of JSON-LD, search engines can contextualize and fully understand your content, giving them high confidence in presenting it to users searching for relevant and related topics.

Implement Schema Markup on all relevant pages on your site

Your website has content about your organization, products and services, authors, case studies, blogs, events and more. These are all entities in your brand's knowledge graph that you want search engines to recognize and comprehend.

Implementing Schema Markup allows you to categorize your entities and explicitly relate them to each other to provide search engines with helpful contextual information about your content.

To implement Schema Markup on a page, simply use the Schema.org Types and properties to identify and describe the entity on a page.

Let's illustrate this using an example.

EXAMPLE

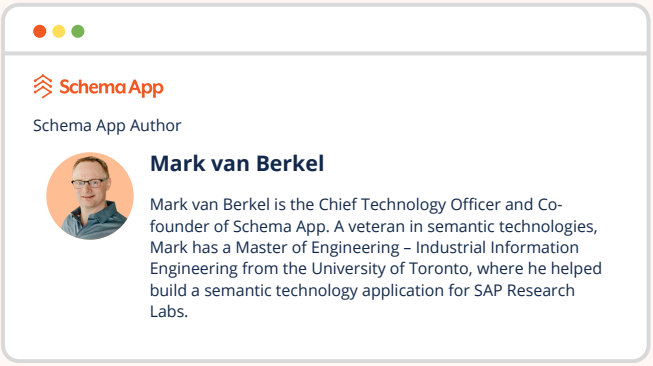
At Schema App, we have author pages on our website and we want search engines to understand who a person is and what their areas of expertise are.

According to Mark van Berkel's author page, he is the Co-founder and Chief Technology Officer at Schema App. He is an alumnus of the University of Toronto and he knows about semantic technologies.

Based on the content on his author page, we can use the Schema.org vocabulary to make a series of statements that describe Mark.

- This entity is a Person.
- This Person is named Mark van Berkel.
- This Person works for Schema App.
- This Person has the job title Chief Technology Officer.
- This Person knows about Semantic technologies.

When you use Schema Markup to describe the entities on your site, you are building RDF triples. We can then express this data using JSON-LD.




```
<script type="application/ld+json">
{
   "@context": "https://schema.org",
   "@id": "https://www.schemaapp.com/author/vberkel/#Person",
   "url": "https://www.schemaapp.com/author/vberkel/",
   "name": "Mark van Berkel",
   "worksFor": {
     "@type": "Organization",
     "name": "Schema App",
      },
   "jobTitle": "Chief Technology Officer",
   "knowsAbout": {
     "@type": "Thing",
     "name": "Semantic Technology",
      },
}
</script>
```

JSON-LD that describes the content on Mark's author page

Implementing and managing your Schema Markup at scale can be tedious, especially if done manually.

Most Schema Markup plugins in the market will automatically add default Schema Markup to certain pages. However, this method restricts the amount of control you have over your Schema Markup and won't give your content the richly descriptive Schema Markup that best supports search engines.

We recommend using tools like our Schema App Highlighter to generate customized Schema Markup at scale.

Schema App

Connect your entities using Schema Markup

You can also connect different entities on your site to each other. In our previous example, we mentioned that Mark van Berkel works for the organization, Schema App.

This organization is an entity, and we've identified and described this entity on our home page using Schema Markup.

We can express that the person, Mark van Berkel (an entity), works for the organization, Schema App (an entity), by nesting the organization markup under the worksFor property.

By connecting both entities, we are providing search engines with more contextual information about Mark and his relationship to Schema App.

```
<script type="application/ld+json">
{
   "@context": "https://schema.org",
   "@type": "Person",
   "@id": "https://www.schemaapp.com/author/vberkel/#Person",
   "url": "https://www.schemaapp.com/author/vberkel/",
   "name": "Mark van Berkel",
   "worksFor": {
     "@type": "Organization",
     "@id": "https://www.schemaapp.com/#Organization",
     "url": "https://www.schemaapp.com/",
     "name": "Schema App",
     "description": "Schema App is a schema markup solution...",
     "telephone": "+18554448624",
     "email": "support@schemaapp.com",
     "sameAs": "https://www.linkedin.com/company/2480720/",
     "sameAs": "https://twitter.com/schemaapptool",
     "sameAs": "https://www.youtube.com/channel/@SchemaApp",
      },
   "jobTitle": "Chief Technology Officer",
   "knowsAbout": {
     "@type": "Thing",
     "name": "Semantic Technology",
     },
}
</script>
```

It is also vital to describe the relationships between marked-up entities in detail to ensure accurate representation. You want to make sure you're using the property that best describes the relationship between both entities. The Schema.org vocabulary often has forty or more properties for each type of thing, giving you many options to accurately define the relationship.

In the example above, the worksFor property most accurately describes the relationship between Mark and Schema App. Other properties like affiliation or memberOf would not be an accurate description of their relationship in this situation.

Pro Tip: Looking to identify properties to connect your entities? Use our free Schema Paths Tool to identify available properties that you can use to connect your entities effectively.

Ensure your Entities have URIs

For the entities on your knowledge graph to be retrievable and identifiable, they must each have a distinct Uniform Resource Identifier (URI).

In JSON-LD, this is expressed with the '@id' attribute. The purpose of URIs is to uniquely identify resources (like entities) and enable the interlinking of data.

Think of it like a social security number in the US. There may be 125 different people named "Mark van Berkel" in the US, but each of them will have a unique social security number that differentiates them.

In our previous example, the organization entity created for Schema App's homepage had the @id: https://www.schemaapp.com/#Organization. If you navigate to this URI, it will take you to the page about the organization.

By connecting both entities, we are providing search engines with more contextual information about Mark and his relationship to Schema App.

When you publish your Schema Markup using the Schema App Editor or Highlighter, our tool automatically generates HTTPS URIs for the entities you define in your Schema Markup. That way, you can easily reference and link the entities across your website in your Schema Markup.

3 Use Linked Entities from External Knowledge Graphs

Using Schema.org to define and relate the entities across your site will help search engines understand your content. However, you can help search engines and other machines further disambiguate the entities on your website by connecting them to linked entities in external knowledge graphs.

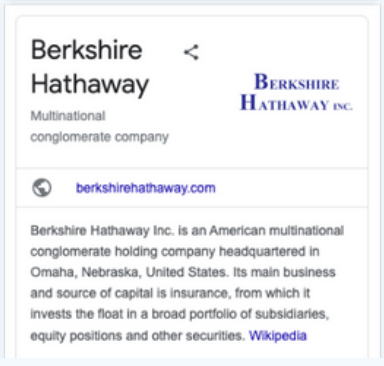
As mentioned earlier, words can be ambiguous. In the English language, the word 'Apple' is both a fruit and a brand. If you were referring to the brand Apple in your content, linking the entity on your site to the Apple brand entity in an external knowledge graph can help machines disambiguate which Apple you are referring to.

EXTERNAL KNOWLEDGE GRAPHS IN SEO

External knowledge graphs like Google's Knowledge Graph, Wikipedia and Wikidata are authoritative knowledge bases that contain records of millions of entities on the web and their relationships to other entities.

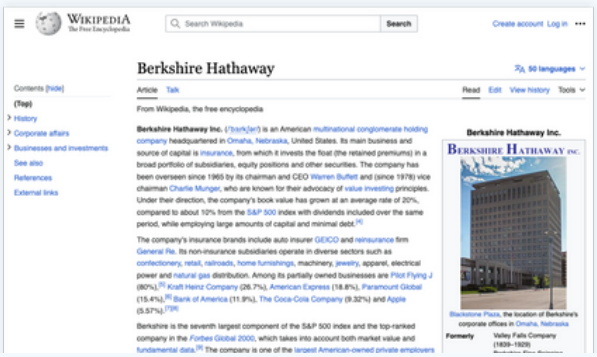
These entities link to other entities across the web which is why they're referred to as linked entities.

The linked entities identified in these external knowledge graphs also have unique identifiers (URIs) that you can use to link to your own entities.

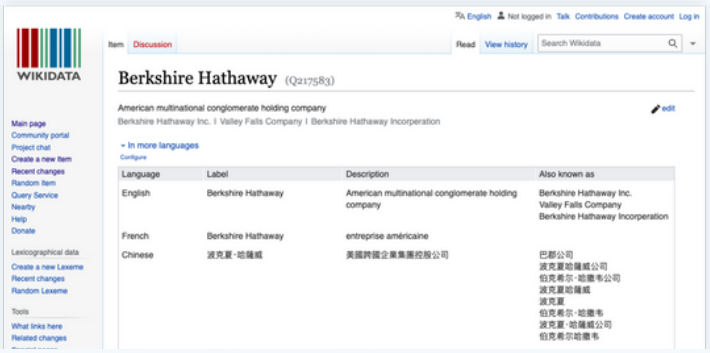


The entity 'Berkshire Hathaway' on Google's Knowledge Graph displayed as a Knowledge Panel on the SERP

URI: /m/01tmng

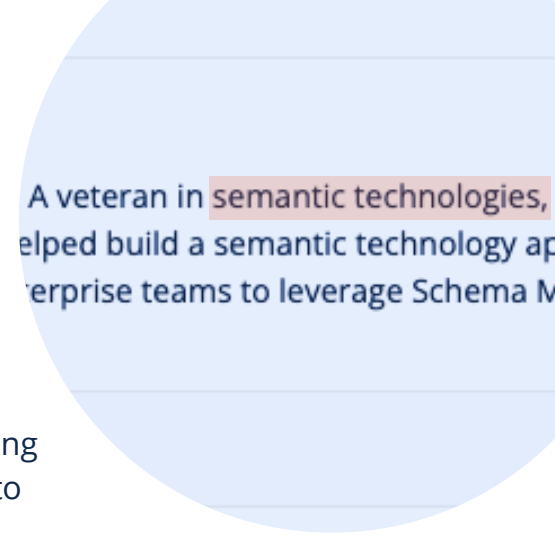


The entity 'Berkshire Hathaway' on Wikipedia
URI: https://en.wikipedia.org/wiki/Berkshire_Hathaway



The entity 'Berkshire Hathaway' on Wikidata
URI: https://www.wikidata.org/entity/Q217583

EXAMPLE



On Mark's author page, it says that he knows about Semantic Technology. However, 'Semantic Technology' isn't available in the Schema.org vocabulary.

But we can be more descriptive about this entity by categorizing it as a generic Thing and using the sameAs property to link it to entities in external knowledge graphs.

```
<script type="application/ld+json">
{
   "@context": "https://schema.org",
   "@type": "Person",
   "@id": "https://www.schemaapp.com/author/vberkel/#Person",
   "url": "https://www.schemaapp.com/author/vberkel/",
   "name": "Mark van Berkel",
   "worksFor": {
     "@type": "Organization",
     "@id": "https://www.schemaapp.com/#Organization",
     "url": "https://www.schemaapp.com/",
     "name": "Schema App",
     "description": "Schema App is a full-service schema markup solution.",
     "telephone": "+18554448624",
     "email": "support@schemaapp.com",
     "sameAs": "https://www.linkedin.com/company/2480720/",
     "sameAs": "https://twitter.com/schemaapptool",
     "sameAs": "https://www.youtube.com/channel/@SchemaApp",
      },
   "jobTitle": "Chief Technology Officer",
   "knowsAbout": {
     "@type": "Thing",
     "@id": "https://www.schemaapp.com/author/vberkel/#Thing",
     "name": "Semantic Technology",
     "sameAs": "https://www.wikidata.org/wiki/Q7449091",
     "sameAs": "/m/0c13x_",
     "sameAs": "https://en.wikipedia.org/wiki/Semantic_technology",
     },

}
</script>
```

This helps search engines to understand that:

- Mark van Berkel knows about a Thing called "Semantic Technology".
- This Thing is the same as the Semantic Technology entity defined on Wikidata, Wikipedia, and Google's knowledge graph.

Linking to entities from different sources can provide machines and search engines with more contextual information, enabling them to infer new knowledge from existing facts.

Entity Linking Methods

There are several ways you can link the entities on your site to external knowledge graphs. You can:

Manually link the entities using properties like sameAs or areaServed

> The sameAs property was created to state that an entity is exactly the same as an entity from another source. Using the sameAs property means the entity you're markup up inherits all of the same information, attributes and relationships of the external source. Read this article for more details on how to use the sameAs property for entity linking.
>
> Entities are also commonly linked to properties that expect a place. For example areaServed, addressCountry or location.
>
> This method of linking is static but not scalable. You are required to manually update the links if there are any changes to the entities on your pages.

Automatically link the entities using the Omni Linked Entity Recognition (LER) feature in the Schema App Highlighter tool

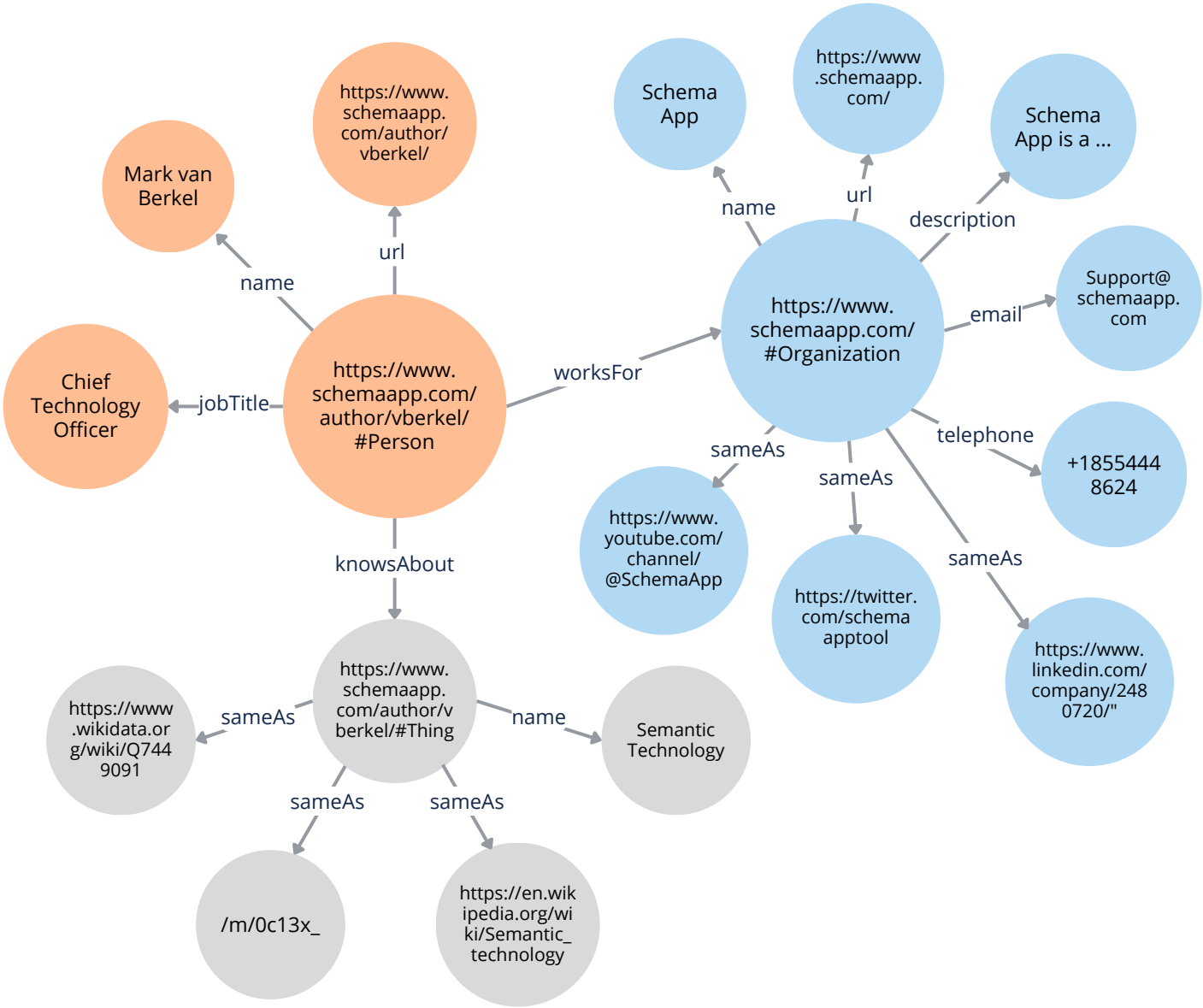
> If you are a Schema App Highlighter user, you can apply Omni LER tags to your Highlighter templates. Once applied, Omni LER runs text through an API to identify linked entities. If entities are recognized, the API returns URIs from the following sources, linking to your content using the sameAs property.
>
> This method is dynamic and scalable. When you make any updates to your content, Omni LER will automatically rerun your text through an API to identify any linked entities and add the URIs from those external knowledge graphs.

Automatically link the categories and tags on a WordPress site to existing linked entities using the Schema App Advanced WordPress plugin.

> If you are a Schema App Advanced WordPress plugin user, the plugin will automatically map the Category of the blog post to existing Wiki entities using the about property to your BlogPosting Schema Markup and map the Tags of the blog post using the mentions property. For more details on this method, read this article.
>
> This method is dynamic and scalable. However, it is only available to WordPress websites.

When you define each entity – what it's about and how it relates to other things on the web – using the Schema.org vocabulary, you are building a reusable knowledge graph.

This is what Mark's knowledge graph would look like once we've linked all the relevant entities.



Having a knowledge graph can improve the search engine's semantic understanding of your content and help them match your pages to relevant search queries with greater accuracy. However, there are benefits beyond search. The versatility of knowledge graphs provides limitless flexibility in their applications.

Value of Knowledge Graphs Beyond SEO

The knowledge graph produced using your website content is rich with factual information about your organization and can be reused for purposes beyond SEO. You can reuse your knowledge graph to train and ground your LLM, analyze your content, and more.

GROUNDING YOUR ORGANIZATION'S LLMS

Many organizations are turning to LLM-powered AI chatbots to support customer enquiries on their site. However, the success of your AI chatbot hinges on its ability to provide accurate information at all times and LLMs have a tendency to hallucinate due to their inability to fact-check information.

To improve the accuracy of your AI chatbot, you need to train your LLM to understand specific things about your business. This information is often available in your website content in the form of unstructured data.

By creating a knowledge graph using your website content with Schema Markup, you have a reliable marketing data layer that can be used to train and ground your LLM. This makes it easier for your LLM to consume and infer factual, domain-specific knowledge about your organization and gives you a control point for how your content is understood. The result is accurate, fact-based answers at a lower cost.

Unlike other LLM training data, your knowledge graph is also readily available, which means that you can quickly deploy it to train your LLM. If you are a Schema App customer, we can export your knowledge graph for you to train your LLM.

Schema App

ANALYZE YOUR CONTENT

Formatting and structuring your content in the form of a knowledge graph allows you to categorize and quantify your content library and identify gaps.

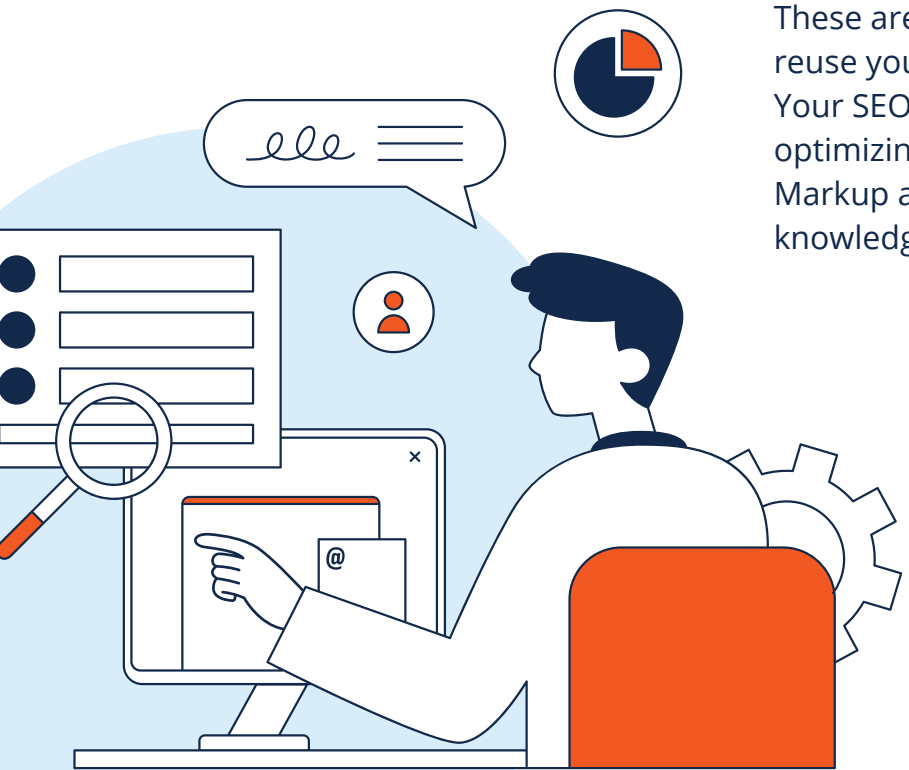
It also enables a deeper level of analysis to answer questions like:

- Which schema.org Types are used most often?
- How many entities are identified on each page?
- How many properties are being used?
- How much of your content is being marked up?

Your knowledge graph can also be used to assess your content qualitatively, by answering questions such as:

- How do your entities compare to the search terms you're targeting with your content?
- Are you linking to entities from external authoritative knowledge bases? If so, which ones?
- Are there any missing entities and topics that should be present considering your area of expertise?

These are just two of many ways you can reuse your knowledge graph beyond SEO. Your SEO team can lead the way by optimizing your website with Schema Markup and building a rich reusable knowledge graph as a by-product.



Leverage Entities & Knowledge Graph for SEO Success

As the SEO landscape becomes more semantically driven, it is pertinent for SEO teams to stay agile and implement semantic Schema Markup to maintain their visibility on the SERP.

Your website content is filled with information about your organization. By defining and connecting the entities on your site using the Schema.org vocabulary, you are developing a reusable knowledge graph that enhances search engines' understanding of your content and prepares your organization for AI.

This deeper comprehension ensures highly relevant search results, increasing the likelihood of user clicks and qualified traffic, and ultimately boosting CTR for relevant pages. It ultimately prepares your data for the new ways consumers will find and engage with your data in this ever-evolving AI and Search landscape.

If you are looking to leverage semantic Schema Markup or develop a robust reusable knowledge graph to futureproof your organization for AI search, we can help.

At Schema App, we help enterprise SEO teams leverage semantic Schema Markup to develop their knowledge graph and improve search performance. Visit our website to learn more about our end-to-end Schema Markup solution.

Get in Touch

Hello@SchemaApp.com

www.SchemaApp.com

Schema App