

# Minimal representations of information for web spiders

Dr. Alexej Schelle<sup>1</sup>

*<sup>1</sup>Freelancer (Senior) Lecturer @ IU Internationale Hochschule,  
email : alexej.schelle@gmail.com*

## Abstract

A short introduction on how to present the most important information of an enterprise or personal webpage to standard search engines like Google, MSN, or Yahoo, is presented. The information concept is built on JSON datasets and makes use of standard data schemes from Schema.org. Using the standard declarations of the datasets as presented, this provides a data feed for search engines from integrating the data with a link into the header of standard HTML websites, mobile-compatible websites, or standalone mobile-compatible websites.

## I. I. INTRODUCTION

Informal Dataset Representations for Web Spiders can be standardized as a technique for representing informal data in a compact form, which includes an individual JSON dataset containing the information of interest. It can be implemented equally well for websites, mobile-compatible websites, and standalone mobile-compatible websites and uses public website content, canonical tags, meta viewports, or HTTP headers, simply by placing a link, like

```
<script src="http://www.domainname/idr_plugin.js"></script>
```

into the JavaScript-sensitive section of the HTML (header) file, and further writing the required public information into the referred file in JSON format. Using such IDR for web spiders, one may profit from

1. A realistic and unique informal data model for search engines about the content of websites and SaaS applications
2. A generator for nice search engine results through individual data mapping and understanding of website content
3. A data map of company networks and relations in a compact data format for search engines to idealize the search engine results' network environment.

To fill the header of the IDR-JSON file containing all the necessary information for a web spider in a condensed and formally ideal form to find an optimal representation of abstracts, logos, and search engine ranking, a large range of meta tag attributes can be used to declare the main information framework of a website, such as:

```
<meta name = "title" content = "Your Company Name">  
<meta name = "keywords" content = "..."/>  
<meta name = "description" content = "..."/>  
<meta name = "subject" content = "...">
```

```

<meta name = "copyright" content = "...">
<meta name = "language" content = "...">
<meta name = "robots" content = "..."/>
<meta name = "revised" content = "..."/>
<meta name = "abstract" content = "...">
<meta name = "topic" content = "...">
<meta name = "summary" content = "...">
<meta name = "classification" content = "...">
<meta name = "author" content = "...">
<meta name = "designer" content = "...">
<meta name = "copyright" content = "...">
<meta name = "reply-to" content = "...">
<meta name = "owner" content = "...">
<meta name = "url" content = "...">
<meta name = "identifier-URL" content = "...">
<meta name = "category" content = "...">
<meta name = "coverage" content = "...">
<meta name = "distribution" content = "...">
<meta name = "rating" content = "...">
<meta name = "revisit-after" content = "...">
<meta http-equiv = "expires" content = "...">
<meta http-equiv = "pragma" content = "...">
<meta name = "viewport" content = "width = ..." >

```

Encoding more specific information about the content of a company or personal website requires data written into JSON-LD data columns, like datasets as presented on [schema.org](http://schema.org):

```
// Geschaeftsangaben :
```

```

<script type="application/ld+json"> {
  "@context": "http://schema.org",

```

```

"@type": "Organization",
"address": {
  "@type": "PostalAddress",
  "addressLocality": "...",
  "addressRegion": "...",
  "postalCode": "...",
  "streetAddress": "..."
}
,
"name": "Your Company Name",
"URL": "yourdomain",
"email": "...",
"telephone": "...",
"potentialAction": {
  "@type": "ViewAction",
  "target": ["yourdomain",
    {
      "@type": "EntryPoint",
      "urlTemplate": "yourdomain",
      "actionApplication": {
        "@type": "...",
        "name": "...",
        "operatingSystem": "..."
      }
    }
  ]
},
,
"yourdomain/",
{
  "@type": "EntryPoint",
  "urlTemplate": "yourdomain",
  "actionApplication": {

```

```

        "@type": "SoftwareApplication",
        "name": "...",
        "operatingSystem": "..."
    }
}
,
"yourdomain/",
{
    "@type": "EntryPoint",
    "urlTemplate": "yourdomain",
    "actionApplication": {
        "@type": "SoftwareApplication",
        "name": "...",
        "operatingSystem": "..."
    }
}
]
}
,
"exifData": [ {
    "@type": "PropertyValue",
    "name": "...",
    "value": "..."
}
,
"exifData": [ {
    "@type": "PropertyValue",
    "name": "...",
    "value": "..."
}
,

```

```

    "exifData":[ {
        "@type": "PropertyValue",
        "name": "...",
        "value": "...
    }
    ,
    "exifData":[ {
        "@type": "PropertyValue",
        "name": "...",
        "value": "...
    }
    ,
    "exifData":[ {
        "@type": "PropertyValue",
        "name": "...",
        "value": "...
    }
}
</script>

```

To improve the design of search engine results in search engine queries in standard applications, such as Google, MSN or Yahoo, at least the company logo and social profile should be added, with the specific data columns, indicated as

```

// Logo :

{
    "@context": "http://schema.org",
    "@type": "Organization",
    "URL": "...",

```

```

    "logo": "...
}

// Social Profile :

<script type="application/ld+json">

{
  "@context": "http://schema.org",
  "@type": "Person",
  "name": "your name",
  "URL": "",
  "relatedlink": [
    "http://www.facebook.com/your-profile",
    "http://instagram.com/yourProfile",
    "http://www.linkedin.com/in/yourprofile",
    "http://plus.google.com/your-profile"
  ]
}

</script>

```

Finally, to declare feeds for references to other partner companies or related websites, the following lines of code will be perfectly well suited to represent information to a public audience on the World Wide Web, in the format experienced by standard search engines:

```

// Forward Linking

<script type="application/ld+json">

```

```
{
  "@context" : "http://schema.org",
  "@type" : "WebSite",
  "url" : "Partner URL",
  "name" : "Partner",
  "author" : {
    "@type" : "...",
    "name" : "..."
  },
  "description" : "...",
  "publisher" : "Partner Website",
  "potentialAction" : {
    "@type" : "SearchAction",
    "target" : "http://.../?s={keyword}",
    "query-input" : "required name=keyword",
    "relatedlink" : "Related Domain"
  }
}
```

```
</script>
```

```
// Backward Linking
```

```
<script type="application/ld+json">
```

```
{
  "@context": "http://schema.org",
  "@type": "WebSite",
  "URL": "...",
  "name": "Your Company Name",
  "author": {
```



```

    "atype": "Person",
    "name": "...",
  },
  "description": "...",
  "publisher": "publisher name",
  "potentialAction": {
    "atype": "SearchAction",
    "target": "yourdomain/?s={keyword}",
    "query-input": "required name=keyword" },
  "relatedlink" : "http://..."
}

</script>

```

From the Data Science perspective, the above (minimal) informal dataset representation defines a minimal feed for standard search engines, such as Google, MSN or Yahoo, with all the necessary information to represent a company or personal website at a suitable ranking without the requirement for using further specific and partly costly SEO optimization campaigns.