BE DISCOVERED!

Indexation guide by Josée Plamondon

Document your content to be understood by the search engines

FIP
Fonds indépendant de production | Independent Production Fund

A publication of the Independent Production Fund
in collaboration with

SODEC Québec
TV5 QUÉBEC CANADA

Contents

Section 1 3

Foreword and credits 3

1.1. Who is this guide for? 4

1.2. Definitions: Content; Data; Metadata 5

1.3. This is not about web page optimization 6

1.4. Search engines: transformation in progress 7

1.5. Documented content vs non-documented content 8

Section 2 12

2.1. How to document content for the web? 12

A. Create an IMDb page 12

B. Create an article in Wikipedia 12

C. Index your webseries to create metadata for search engines 13

2.2 How to create a Wikipedia page: What you need to know 14

Where to start 15

Going one step further 17

Useful references 17

Section 3 19

3.1 How to create metadata for search engines 19

Where are the metadata that describe your content? 19

Describing web series with structured data 20

Structured data markup tool for television series 20

Do you have authorization to modify your web site? 20

Describe an episode of the webseries 21

3.2. This is just the beginning 23

Section 1

Foreword and credits

In a constantly evolving media world every word, number and image you produce attracts the attention of search engines – along with the other billions of words, numbers and images in the universe. How can you stand out? We thought it would be useful to create this Guide in order to demonstrate how, using simple tools, it is possible to maximize the visibility of your content in the semantic web. The Board of Directors of the Independent Production Fund hopes to help demystify "discoverability" in the context of linear content production. Ideally, by applying the suggestions in this Guide, you will be ensuring that your content becomes recommendations of choice by search engines. Your potential audiences will find you even if they don't know that you exist.

This Guide was created with the help of partners whom we warmly thank: Harold Gendron, Project Manager, Strategic Development Directorate and President's Office at Sodec; Benoît Beaudoin, Director, Innovation and Digital Lab at TV5, and his colleagues, Samuel St-Pierre and Jérôme Lapointe; Turbulent has undertaken its web integration.

We had the great opportunity to interest Josée Plamondon, consultant in digital content exploitation, in the drafting of this project. We appreciate her intellectual rigor but above all, her passion for words, data and metadata. We hope to transmit this enthusiasm to you because it is about the future of your content!

Claire Dion, Deputy Director General, Independent Production Fund

Andra Sheffer, CEO, Independent Production Fund

Credits:
Research and writing: Josée Plamondon (consultant)
Editor: Claire Dion (Fip)
Working Committee: Harold Gendron (Sodec), Benoît Beaudoin (TV5), Claire Dion
Research and programming: Samuel St-Pierre and Jérôme Lapointe (TV5)
Cover page: Bruno Provencher (TV5) and Claire Dion
English translation: Josée Plamondon and Andra Sheffer (IPF)
Web Integrator: Turbulent

Published in Canada in English and French: Independent Production Fund www.ipf.ca and www.ipf.ca/Fip
Published in Canada November 2017
A French version is also available at www. http://ipf.ca/FIP/

Date of Update: June 12th, 2019
Sections concerned: 2.1, 3.2, 3.3.
Reason: Google now gives preference to its media partners for content types *Movies, TV Series* and *Music*. However, using these types for content descriptions contraindicated, although the treatment by the search engine is unclear. The type *VideoObject* is a very interesting alternative for content that is made available on the web.

ISBN 978-0-9876748-3-8

creative commons
This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 2.5 Canada License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/2.5/ca/ or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

1.1. Who is this guide for?

Although this guide has been written to make concepts that are complex understandable to as many readers as possible, some sections may be of particular interest to certain groups of specialists.

Sections	
1 and 2	Producers, decision makers, managers, project managers, developers. Communication professionals, creators of content
3	Programmers, webmasters and anyone with the necessary knowledge to intervene in the programming of web pages.

1.2. Definitions

To avoid confusion, here are some important definitions for understanding the concepts and instructions that are provided in this guide.

Content

In this document, the term "content" refers to a video, sound recording, image, or any creation on a digital medium. A web page can bring together different contents: text, image and sound.

Data

Regardless of whether it is a sequence of alphanumeric characters, geographic coordinates, a person's name or the title of a video, data is the smallest unit of information.

- Considered individually, it provides partial, ambiguous or totally incomprehensible information.
- Linked to other data, it produces meaningful information.

Data can be found in formats that are standardized by international or industry organizations, such as the ISO format for a date or the MARC format for library records. But in general, apart from libraries or archives, where a common and international vocabulary has been defined to facilitate the exchange of information, data are not created according to recognized standards.

Data is therefore the basic element that can describe content.

Metadata

Metadata is used to describe a given data, that is, to provide its meaning. For example, the data "Venus" may have as its metadata: "goddess" or "planet" or "first name" (tennis player) etc. We find metadata in various forms and everywhere, both in our physical world and in the digital world. Here are some examples of metadata:

- The word "Ingredients" preceding the list of ingredients that makes up a cooked dish.
- The words "Lead Actors" in a film's credits.
- Column headings in an Excel spreadsheet.
- The acronym "ISAN", which indicates that the following numerical code, in a video-on-demand site, is a standardized unique identifier of an audiovisual work.

In a database, metadata describe the attributes and properties of the data. This allows the database management system to follow the relationships established between the data to generate the information needed in response to a query.

1.3. This is not about web page optimization

This guide does not concern the optimization of web pages (also called SEO).

Web page optimization, which relies mainly on keyword tracking, aims to improve the positioning of web pages in search engine results. Applications locate keywords, compare pages between them, and order them according to different factors, depending on the words used in a research query. The optimization of pages does not require any understanding, by the search engines, of texts, images and other content that is present in the web pages. This is however an intervention that should be periodically updated in order to maintain a positioning in a constantly fluctuating digital environment.

Content documentation (or content indexing) for search engines uses new and different techniques that allow machines and systems to understand texts, images and other content. It is no longer a matter of classifying a web page in a list of results, but of locating a content, for example a video, to understand the data describing it (subject, creators, producers, language, place of production, free or paid, etc.) and to link all this information to its knowledge classification system. The documentation of a video allows the machines to make links between it and other content, based on relevant relationships: actors profile, production companies, sound tracks, literary works, other videos of the same genre, etc.

Content indexing is an operation that does not require subsequent actions: the data that describes the content remains within the scope of the search engines as long as the page that contains them is online.

While page optimization allows websites to rank favorably among search results, content documentation allows search engines and other machines to understand them and to link information within the Web. ***The more possible links between content and other content in the web, the better is its potential to be recommended or used as a response by search engines.***

Web page optimization is used to provide a list of pages corresponding to the keywords of the request.

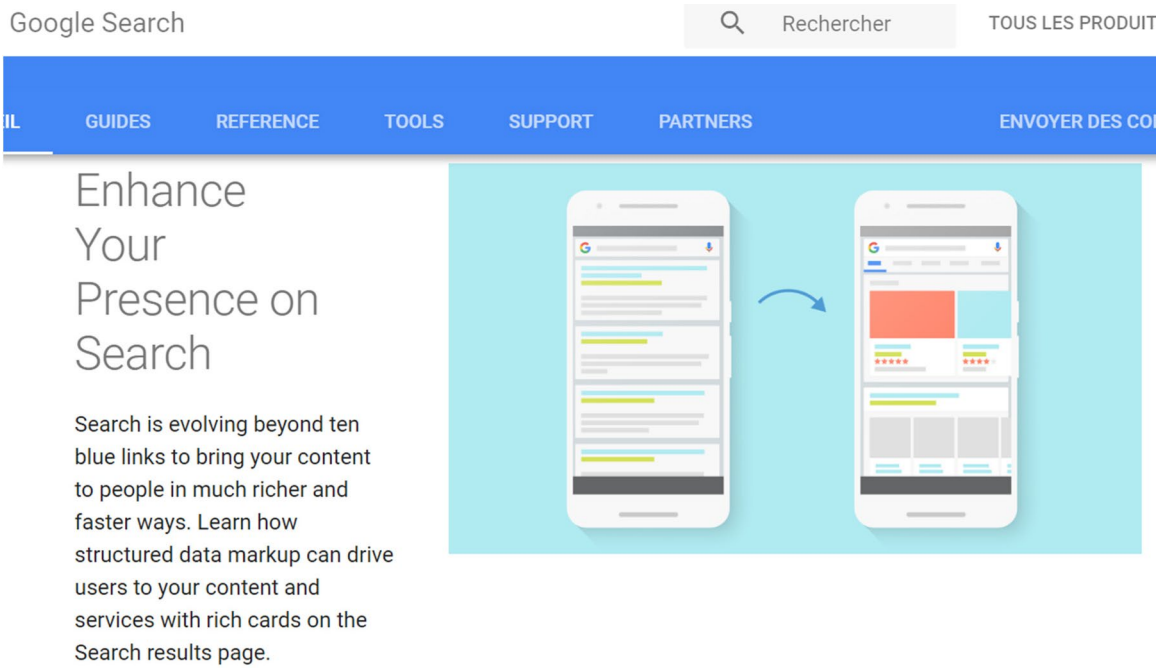
Content documentation is used to link information across the web, regardless of the pages from which it originates, to provide a response or suggestions through links that make sense. For example, a list of popular webseries in Canada – it figures out what you are really looking for!

These two practices are different and complementary and should be combined with marketing and promotion efforts. We will see in the following sections the tools that are available to us by the consortium of search engines (among others: Google, Bing and Yahoo!) so that we are able to adopt new practices to document our content according to established standards.

1.4. Search engines: transformation in progress

Whatever the platform, search engines are now raking the web to find content, preferably on pages that are described using data that their systems can understand and process.

On its *Developer Search Tips* home page, Google recommends documenting content using structured data (also known as semantic tags). As shown in the illustration below, on a smart phone, the list of search results (web pages) is replaced by a form containing information that is collected on the web and that is organized to provide a response and useful links for the user. All search engines, regardless of their platform, are now seeking to provide direct access to information by exploiting the data that describes the content.



https://developers.google.com/search/

Discovery offers more potential than search

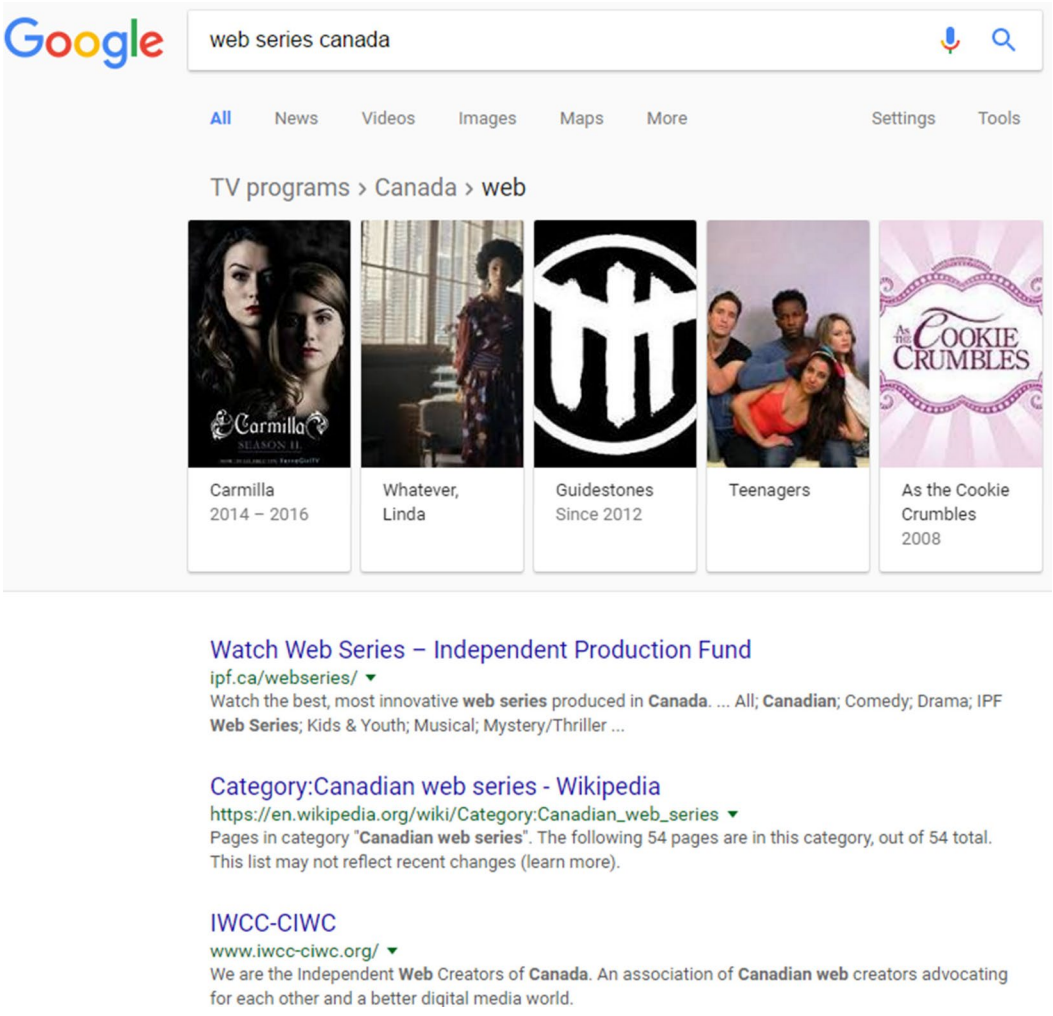
Documenting content allows search engines to create links between your content and any other relevant information on the web. The more persistent and explicit the links are that connect your content to other information in the web, the more likely it is to be found by internet users. Thanks to the work of machines that are searching and collecting information and understanding the meaning of the relationships between different sources of information related to your content, your content will become more discoverable.

Increasing the potential of content to be discovered by consumers makes it possible to reach out to new audiences: those who are unaware of its existence;: those who do not use platforms where the content is likely to be found, or those who simply are not

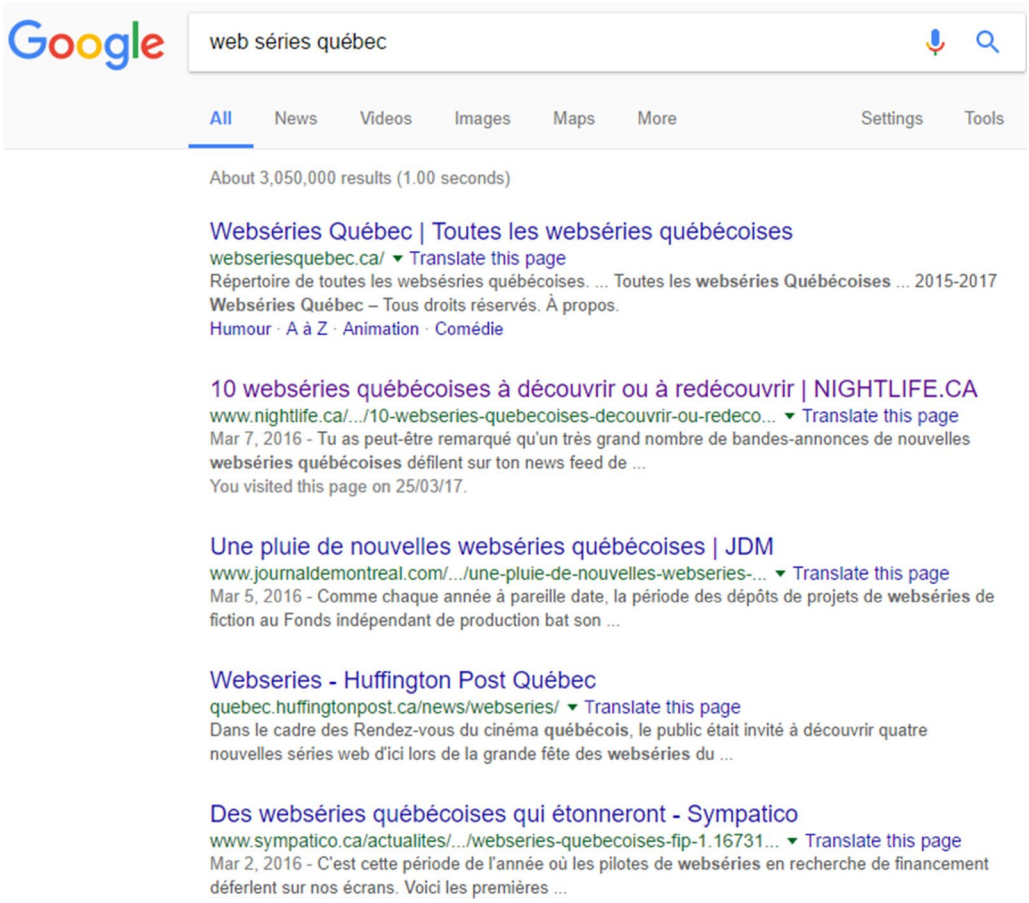
interested in your content, could change their minds when they discover the subject, or a favourite actress, or a particular recommendation that interests them.

1.5. Documented content vs non-documented content

Search engine optimization (SEO) strategies promote content by positioning their web pages among the first search results. However, when content is documented in a format that machines understand, search engines can recommend the right content directly as in the following example of a search carried out to find “webseries”. The carousel that appears above the list of results suggests content that is related to the query because these have been documented for search engines and are described as television series that are also available on the web. The presentation in the carousel provides more accurate and directly accessible information than the list of results located below the carousel.

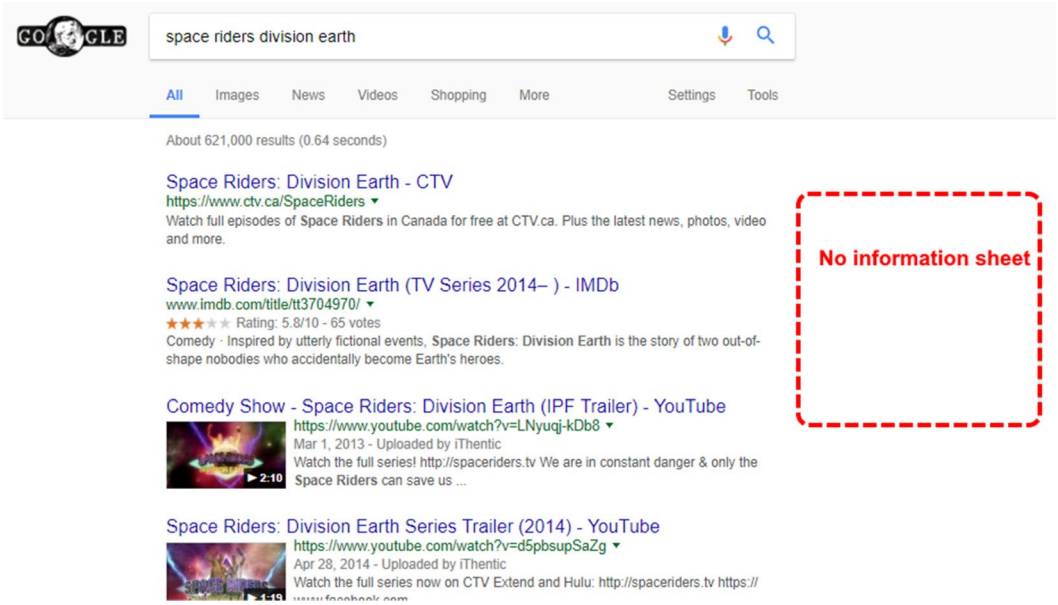


For comparison, here is the result of a search done in French to find webseries from Quebec. As there is not enough of this type of content that is documented for search engines, no recommendations can be displayed. Only the list of web pages that contain the keywords used is generated.

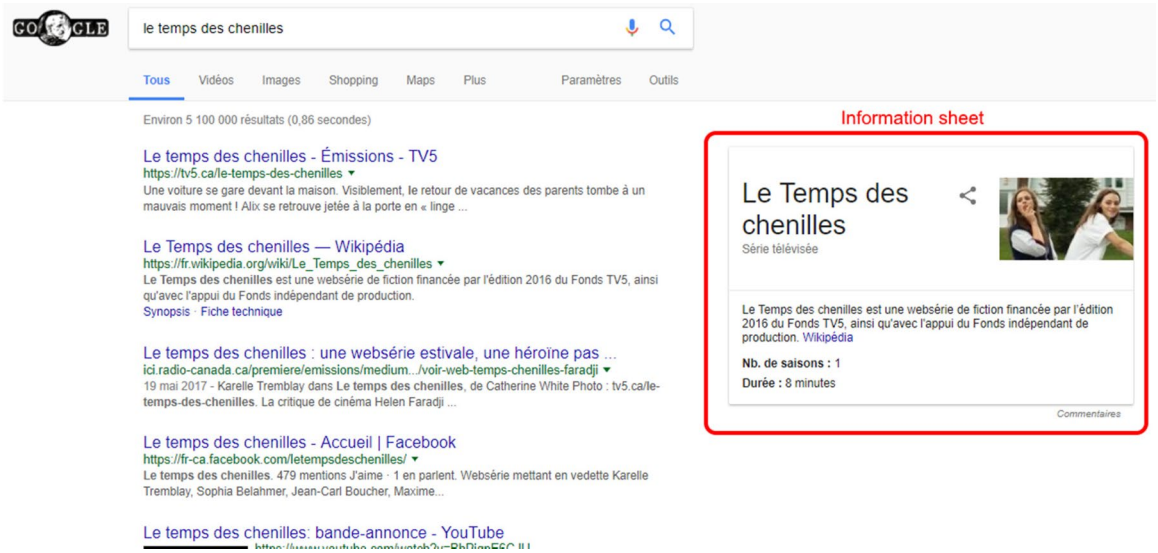


Suggesting content and presenting it visually in a carousel or in the form of an information sheet or knowledge card (shown to the right of the list of results) is also more attractive and useful because it provides information with the search results without the need to even open new pages or to rely on the editorial choice of sites to find content.

Even a search to find a specific web series will only produce a list of web pages if it is not documented. Here is an example:

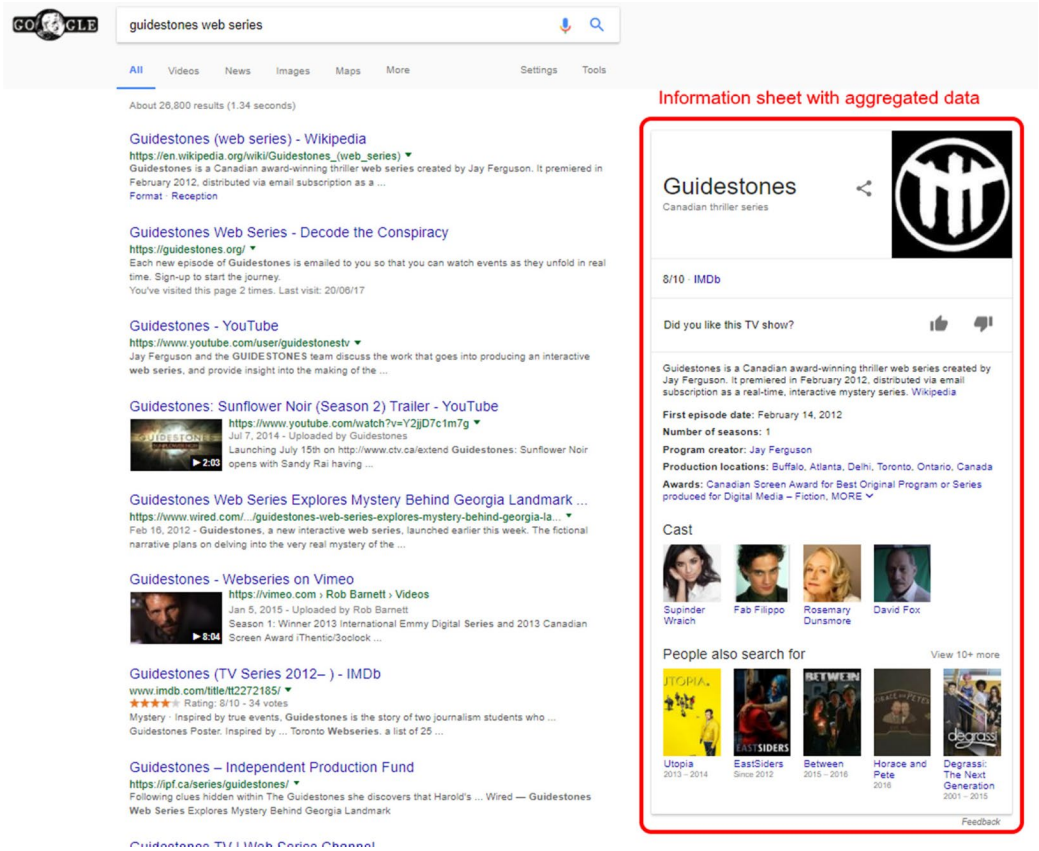


Below is the result of a search performed in French to find a webseries from Quebec, "Le temps des chenilles". The search engine can produce a partial fact sheet that includes the start of a Wikipedia article on the webseries. Each of the articles in the online encyclopedia is documented using metadata that can be understood and used by applications. Google can then link the request for information to the metadata of the specific article in Wikipedia.



Here is another example: it is a web series for which there is a dedicated web page, a Wikipedia page, an IMDb page, as well as links to related information in Wikipedia (actors, prices, shooting locations) and links to similar content proposals.

This web series is well documented on aggregators that are used by search engines to understand the nature of content, build links and generate a fact sheet.



In short: Why document your content for the web?

Content must be documented for the web because data production and reuse are strategic activities that are at the heart of a digital business model.

Mastering the production of descriptive data means multiplying the capacity of content propagation (discoverability) in order not to became dependent on the algorithms of social networks and platforms (YouTube, Vimeo) for the diffusion of content.

Section 2

2.1. How to document content for the web?

This guide does not cover promotion and distribution strategies on social media platforms. Content published on Facebook, for example, is not linked to the documentation of content and structured data on the web, but is being disseminated to a target audience – the users of this social network. There are no tools to produce metadata in Facebook because the platform does not open their content to indexing by external search engines.

Producing and exploiting data for the web requires the technical skills and resources that are specific to the semantic web (also known as web of data). There are, however, accessible and easy-to-use ways to produce metadata that allow applications to make the necessary links to process your content according to a semantic model. There are ways that are within the reach of all and require only the development of new habits and, of course, a little practice.

A. Create an IMDb page

Although IMDb (Internet Movie Database) is not the subject of this guide, it is important to note that it plays a key role for search engines by providing them with standardized information on film and television content (webseries are considered TV series by IMDb). As such, refer to the example in the previous section which represents the information sheet of the series *Guidestones*, with information provided by IMDb. Documenting a web series on IMDb is therefore to be considered as soon as it is released online.

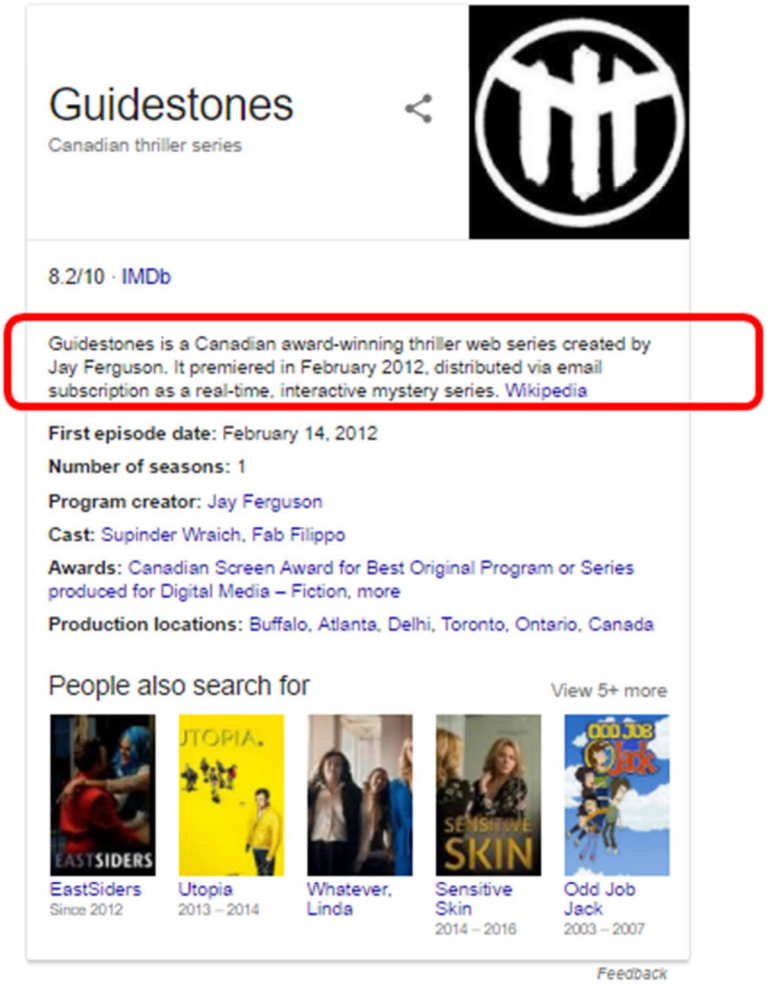
B. Create an article in Wikipedia

(See also: How to Create a Wikipedia page , Section 2.2)

*** Warning: Wikipedia is neither a commercial directory nor a promotional tool. It is important to respect the guiding principles of the encyclopedia, otherwise a contribution will be rejected.

Why create an article in Wikipedia? Because the encyclopedia is used as a reference by search engines and various systems for analyzing digital content in order to eliminate any ambiguity in the description of a specific piece of content and to search for information related to it.

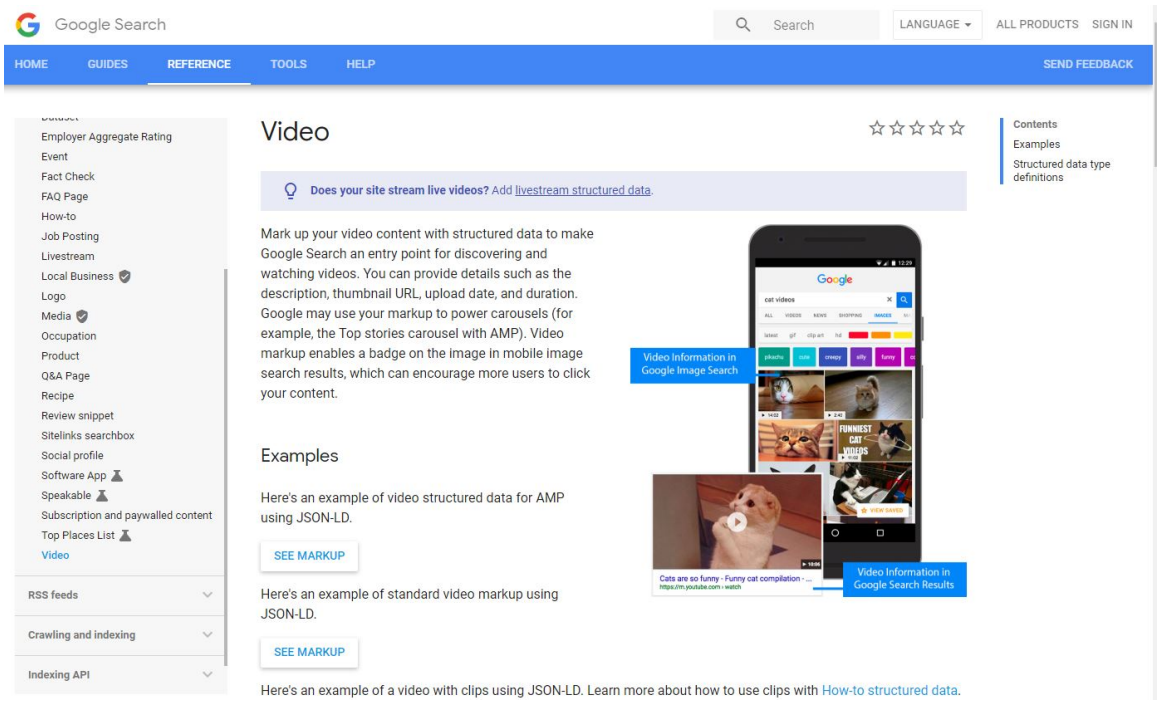
The presence of an article in Wikipedia (as well as the availability of other information on the web), allows the search engine to produce the information sheet that provides detailed information on a content (as shown in the example below).



C. Index your webseries to create metadata for search engines

Why index your webseries? Because by indexing content you are producing metadata to provide a description in the same way that we enter the information required to add content to an online directory or database. Search engines such as Google and Microsoft recommend the use of a metadata model, a kind of semantic markup, called **Schema**. It provides information about the type of content, subject, persons and entities, availability (watch, buy) and links to other metadata that are in a compatible format.

The metadata production guide, provided by Google, for video (https://developers.google.com/search/docs/data-types/video).



How do you do that? See Section 3 - How to create metadata for search engines

2.2 How to create a Wikipedia page: What you need to know

Wikipedia is one of the sources of information that is used by search engines to validate content because of the advantages it offers:

- Knowledge organization model that is open and universal, with descriptive metadata in standard format that can be understood by machines;
- Objective content (editorial independence, prohibited advertising), verifiable, multilingual and freely reusable;
- Content that can be perpetually validated and updated;
- Network of internal and external links connecting the content to the web of data.

Writing a page follows certain rules:

- A topic page is called an "article".
- The writing of an article must respect certain rules of editing and validation of the sources of information.
    - **Objectivity:** the article must have an objective and informative tone. It is about informing, not about promoting.

    - **Notoriety:** Content must have a certain scope and recognition that can be proven using external journalistic sources (your website does not count, nor that of the broadcaster).
    - Wikipedia generally **requires references** from national media to prove the interest and veracity of an article. You shouldn't create an article on a topic for which you don't have references. For web series, official regional media sources or recognized websites can be used, but expect some discussions with moderators.
    - **Do not copy text** from an external source, even if it's your own web site. Contributors and robots are responsible for enforcing intellectual property rights. If you are inspired by a source of information, write in your own words.
    - Do not use **images** unless you can prove you are the copyright holder.
- The encyclopedia is open to everyone: anyone can create, modify or enrich articles.
- Be prepared to discuss with other members of the community as it is common for some users to disagree with your material. You must be prepared to justify the legitimacy of your articles or information to the community.

Important: Keep the information up to date and make relevant links:

- Articles can be enriched and edited by everyone. It is therefore important to keep an eye on the articles or pieces of information that have been added to the encyclopedia (Watchlist).

Where to start

1. **Create an account** to create articles, access edit mode on the "Edit" and "Edit code" tabs, receive notifications and communicate with other contributors for help.

   Help: Getting started (Introduction and tutorials)
   https://en.wikipedia.org/wiki/Help:Getting_started

2. **Find a model:** Use an article on a similar topic that is complete and up-to-date, in order to inspire you with its structure. You can view the components of the structure on the "Edit" tab.

   For **web series**, inspire yourself from this article:
   https://en.wikipedia.org/wiki/Carmilla_(series).

3. **Prepare the content** of your article: Make a list of the information needed to write the article, search for reliable sources, identify relevant links to other articles in Wikipedia, assign at least one category (see "Categories" at the bottom of the page of the article on *Carmilla*).

   Notability (media) - Programming
   https://en.wikipedia.org/wiki/Wikipedia:Notability_(media)#Programming

Identifying reliable sources
https://en.wikipedia.org/wiki/Wikipedia:Identifying_reliable_sources#Context_matters

Linking articles:

Inbound links: If possible, link 3 other Wikipedia articles to yours so that it is not orphaned. These links can be created by editing the articles, as long as the relationships are appropriate. For example: articles about a financing fund, a production company or an actress who are associated with the web series.

In visual editing mode, select the text that contains the information to be linked and use the option represented by the "chain" icon, then "Search pages" and select the one that corresponds.

Outbound links (to other articles): When relevant, link information elements in your article to corresponding articles in Wikipedia. For example, if you have an episode that relates to tennis, you can link the word “tennis” to the Wikipedia tennis page.

External links: As far as possible, link only the official site(s) of the web series.

In visual editing mode, select the text that contains the information to link and use the option represented by the "chain" icon in the editor, then "External link" and paste the URL of the site.

Links to social networks (eg Facebook) and platforms (eg YouTube) are not allowed due to the difficulty of validating the information and ensuring the continued existence of the information.

Sources: Facts and affirmations must be confirmed by sources (article, website, book) in the form of notes that are inserted at the end of the sentences using the "Cite" option of the editor. The links will then automatically appear in a "References" section.

Ensure that the exact information that precedes the note is found in the source cited.
Do not put notes or references on a section title. If necessary, place a small sentence under the heading to insert the note.

4. **Create a draft page** to save your work, check your page layout and review your text without having to make it public.

How to create a draft page in your user sandbox:
https://en.wikipedia.org/wiki/Wikipedia:About_the_Sandbox

An article about a web series could contain the following elements:

Description (This is the introductory summary. It is not a section and has no title)

Short presentation of the subject: type of production and brief description.
Put the name of the series in bold at the beginning of the description.
Release date.

Organizations that have produced and funded the web series.

Plot

Episode list

Use bulleted lists to list episode titles.
Show episode titles in italics.

Cast and characters)

Bulleted list.

Production

Reception

Critical reviews, viewership.
Awards.

5. **Publish** the article after revising your content and validating your links. If this is a very brief description, it may be flagged as a stub. Do not worry if a banner appears later and is marked "stub": it is an encouragement to enrich the article addressed to you and possible contributors.

   Stub https://en.wikipedia.org/wiki/Wikipedia:Stub

6. **Develop links** by ensuring that your article finds itself in categories, portals or articles that would have been newly created.

   Wikipedia's contents: Categories:
   https://en.wikipedia.org/wiki/Portal:Contents/Categories
   Wikipedia's contents: Portals:
   https://en.wikipedia.org/wiki/Portal:Contents/Portals

Going one step further

The more links there are to allow search engines to get information in order to contextualize content such as a web page, the more they can link to other relevant information on the web.

The following information aggregators, in the field of audiovisual productions, are used by Google to produce the Knowledge card that results from a search on a web page, television content or a film:

- IMDb http://www.imdb.com/
- Rotten Tomatoes https://www.rottentomatoes.com/

Synopsis, authors, actors, producers, broadcasters: document your content. The more exhaustive the information, the more potential links there are between a web series and any other audiovisual content on these platforms.

Useful references

Wikipedia : Instructional material (tutorials, manuals, books, videos)
https://en.wikipedia.org/wiki/Wikipedia:Instructional_material

Article Wizard (With live help chat)
https://en.wikipedia.org/wiki/Wikipedia:Article_wizard

Wikipedia: Writing better articles
https://en.wikipedia.org/wiki/Wikipedia:Writing_better_articles

Illustrating Wikipedia (pdf)
https://en.wikipedia.org/wiki/File:Illustrating_Wikipedia_brochure.pdf

Section 3

3.1 How to create metadata for search engines

Metadata and data that are used by search engines are also referred to as "structured data" because each data is paired with metadata that help categorize and index your content. Structured data are presented in hierarchical and nested order to describe the links between data.

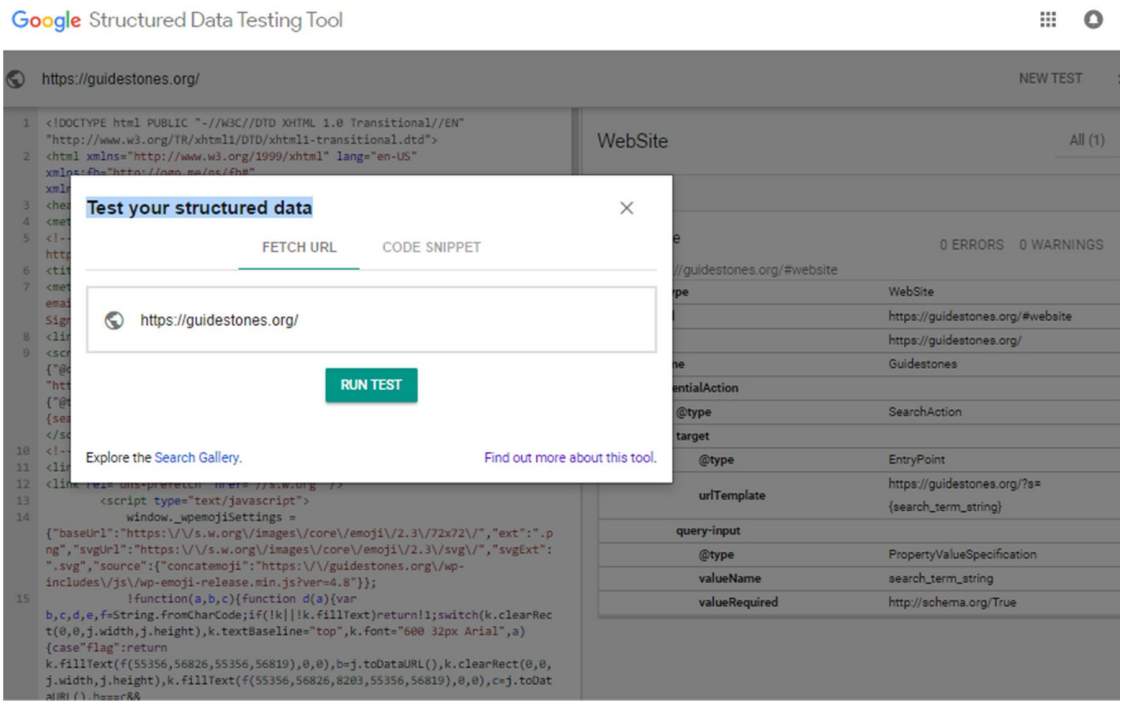
Where are the metadata that describe your content?
It is possible to check the presence on a web page of the type of metadata that Google is looking for. Metadata expressed using a specific vocabulary called Schema allow search engines to understand the description of the content and make links to other information in the web.

To verify the presence of metadata in a web page, simply use the test tool found in Google's search console:

https://search.google.com/structured-data/testing-tool

If there are metadata using the Schema model, they will be displayed on the right side of the screen. The left side will display the html code that is used to present the contents of the web page.

The application reports errors and missing metadata depending on the type of content described. It may also suggest corrections.



Describing web series with structured data

Google produces a guide to encourage the creation of metadata, for various types of content, according to the Schema model and in a semantic format (JSON-LD).

For example, here is the section on video:
https://developers.google.com/search/docs/data-types/video

The content description, or more precisely its indexing, requires a good grasp of the Schema model and technical resources or skills to encode and integrate the metadata into web pages.

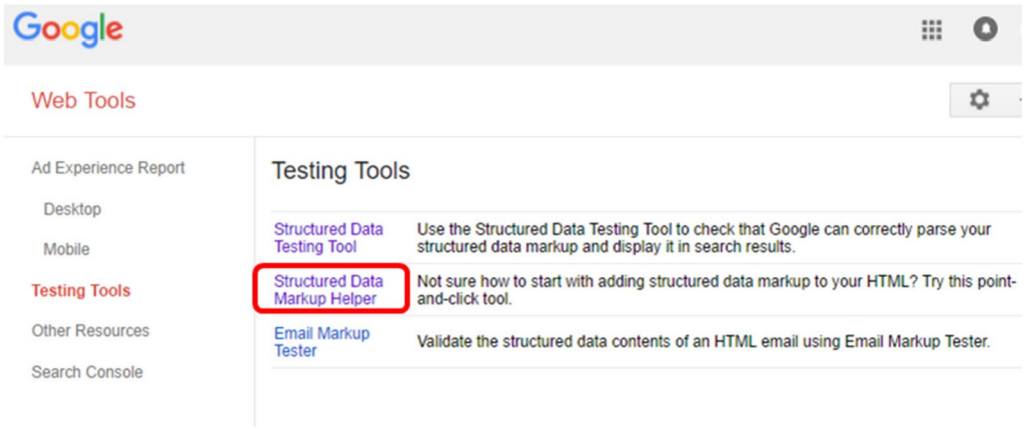
If you do not have the right skills, you can use a wizard developed by Google to help you create structured data and integrate it into the web pages of your web series. However, you must have access to the source code of your web site. The explanations below will guide you.

Structured data markup tool for television series

The Structured Data Markup Helper helps people who have access to the content management of their website to generate and integrate basic descriptive metadata, without programming, according to the vocabulary and format recommended by Google. This tool can be found in Google's Search Console:

https://www.google.com/webmasters/markup-helper/?hl=en

Warning: Google now gives preference to its media for content like Movies, TV Series and Music. However, using these types of content descriptions is not contraindicated, although the treatment by the search engine is unclear. The type *VideoObject* is a very interesting alternative for content that is made available on the web. Please note that this type of content is not available in the markup tool.



Do you have authorization to modify your web site?

To integrate metadata into the pages of your site, you must prove that you have the authorization to do so.

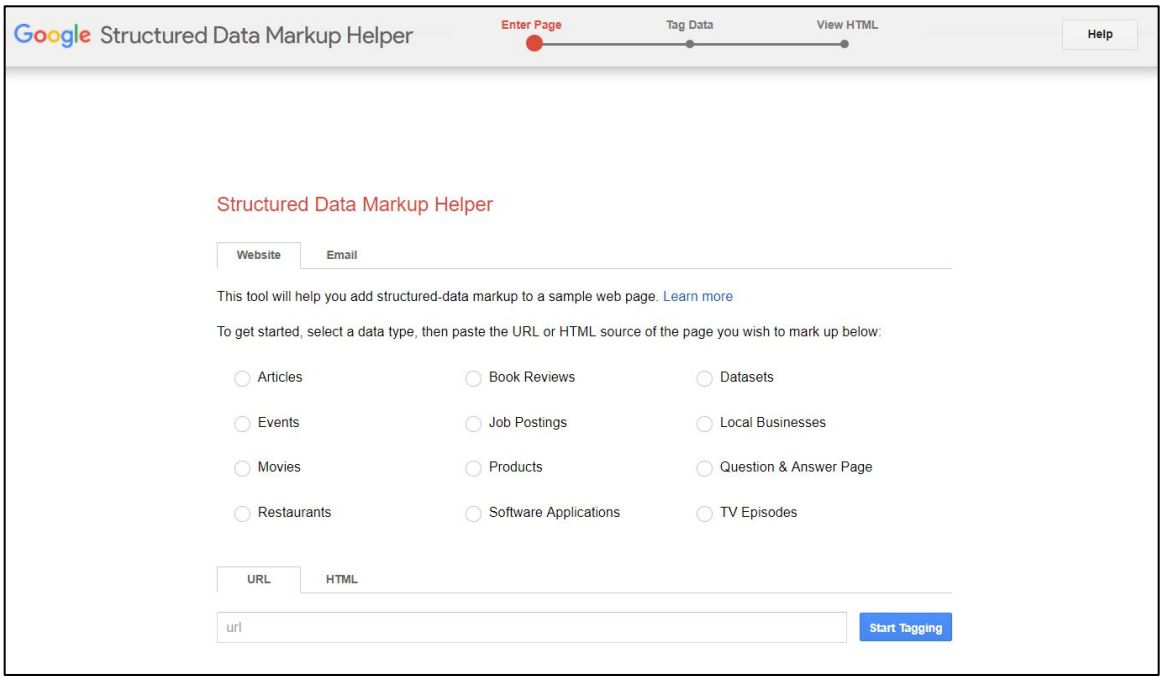
> You will have to comply with one of the property validation processes to prove that you are authorized to access the content and data of the site.
>
> You can save and manage all the pages (called properties) for which you want to produce semantic tags, in the search console.

You will find information about the validation of your access permission in the search console.

Describe an episode of the webseries

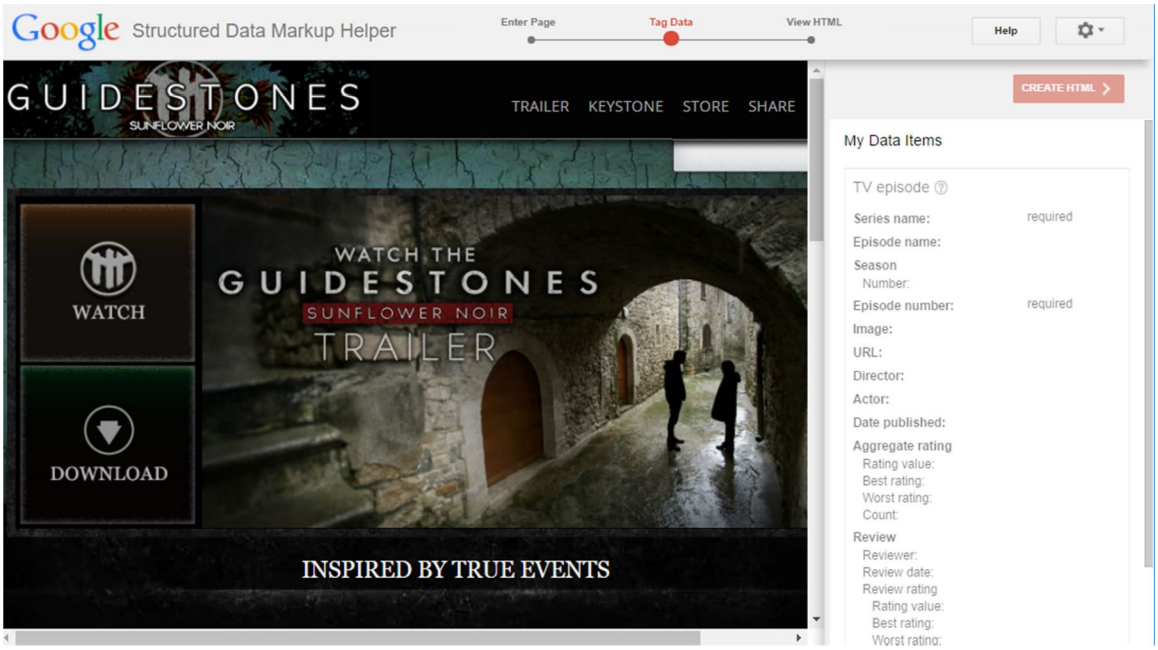
The markup tool allows you to create structured data for a specific content type. Simply select "TV Episodes" and copy-paste the url of the page corresponding to the content you want to describe in order to start documenting your web series and thus provide search engines with a basic description for your episodes of web series.

This point-and-click tool allows you to understand, by following the on-screen instructions, what data search engines are looking for. It is also a good exercise to improve the information offered on your web site or the structure of the presentation.



On the right side of the screen, you will find a list of metadata that provides a basic description for episodes of web series. On the left side of the screen, the page corresponding to the url of the episode is displayed.

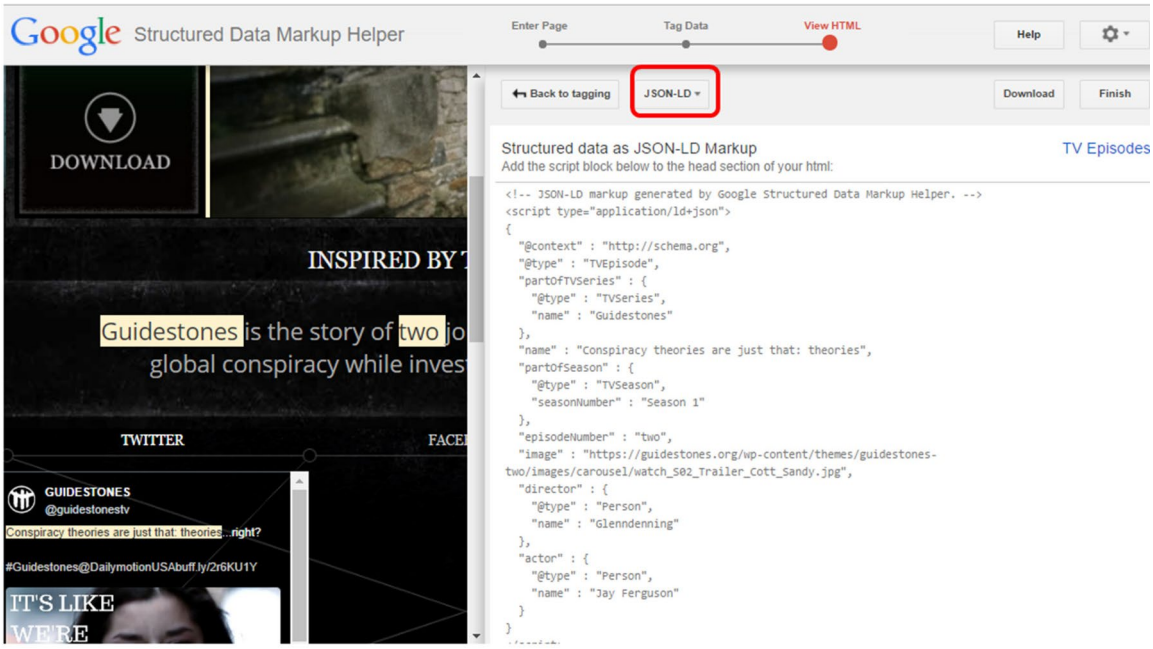
Simply select with your pointer information items that are on the web page to display a choice of metadata. Each of the choices will add information corresponding to the metadata displayed in the list on the right side of the screen. It is possible to use the "Actor" metadata several times.



When the description of the episode is finished, click on the red button in the upper right corner of the screen to create the html code. Then you will need to choose an encoding format at the top of the screen, on the left.

Select "JSON-LD", the encoding format recommended by Google.

Click the "Download" button in the upper right corner of the screen to get an html file that you can copy and paste into the source code of the web page of the web series episode.



The help page for the markup tool contains more information about the process:

https://support.google.com/webmasters/answer/3069489?hl=en

3.2. WordPress plugins for Schema metadata

Creating metadata for search engines must now be considered as a Web best practice. Content management systems for website publishing will very likely incorporate markup assistance functions that are similar to the one previously presented.

WordPress users can already rely on extensions to describe various kinds of content. They can thus produce and integrate metadata in the form of tags according to the Schema vocabulary, without having to program them.

You can find these extensions on the WordPress site, by searching in the Extension section with the term "schema".

https://wordpress.org/plugins/search/schema/

3.3. This is just the beginning…

This guide should raise awareness of the most important technological changes directly affecting content producers; and may be the beginning of concrete tests. It is essential knowledge for any creator or producer of content who wants to maximize their online discoverability potential.

This guide does not offer any recipes that will produce immediate results. It aims to contribute to reflections for informed decision-making in the context of business strategies and digital projects (for example, to provide structured data in the production of new content, whatever it may be).

The web, as most of us have seen it emerge and develop, is in transformation; and this is not about to stop.