SEMANTIC WEB TECHOLOGIES FOR TRAVEL & TOURISM

Sante J. ACHILLE
Search Marketing Consultant
LinkedIn: https://www.linkedin.com/in/sjachille/
Twitter: @sjachille

CATEGORY: Applied, Digital Tourism

INTRODUCTION

Millions of web pages are published every day by local authorities, companies and private citizens with the ambition to inform, share an experience, or tell a story. Are they findable? Is there a way to create a value-added infrastructure? Search engines face the challenge to understand content and make information available to the user in our real-time, mobile-first world. This contribution addresses the need and opportunity for the travel & tourism industry to embrace artificial intelligence (AI) in the shape and form of Structured Data to feed Machine Learning (ML) and Natural Language Programming (NLP) algorithms. It explains the underlying principles and the benefits of a renewed digital strategy.

From a Web of Documents to a Web of Things

The Internet is a collection of arbitrary typed objects (websites) based on the paradigm:

Anyone can say **A**nything about **A**ny topic (AAA).

Such an information web is an organic entity that grows from the interests and energy of the communities that support it[1]. Web pages, images and video have an underlying meaning we, as humans, can understand based on *context*. Search engines may fail to understand the context of a narration or reference to a place or person. They rely on probabilistic calculations to solve semantic problems of translation and interpretation in order to minimize *ambiguity* and reduce *uncertainty*.

- Ambiguity arises when things or concepts can be interpreted in more than one way.
- Uncertainty is a consequence of ambiguity that triggers doubt or confusion and may favour inaccurate conclusions or decisions.

Structured Data, organised in schemas[2], is a means to understand and organize information across the web. Vocabularies[3] define schemas concepts and relationships (Types and Properties) used to describe and represent an area of interest used in a particular application.
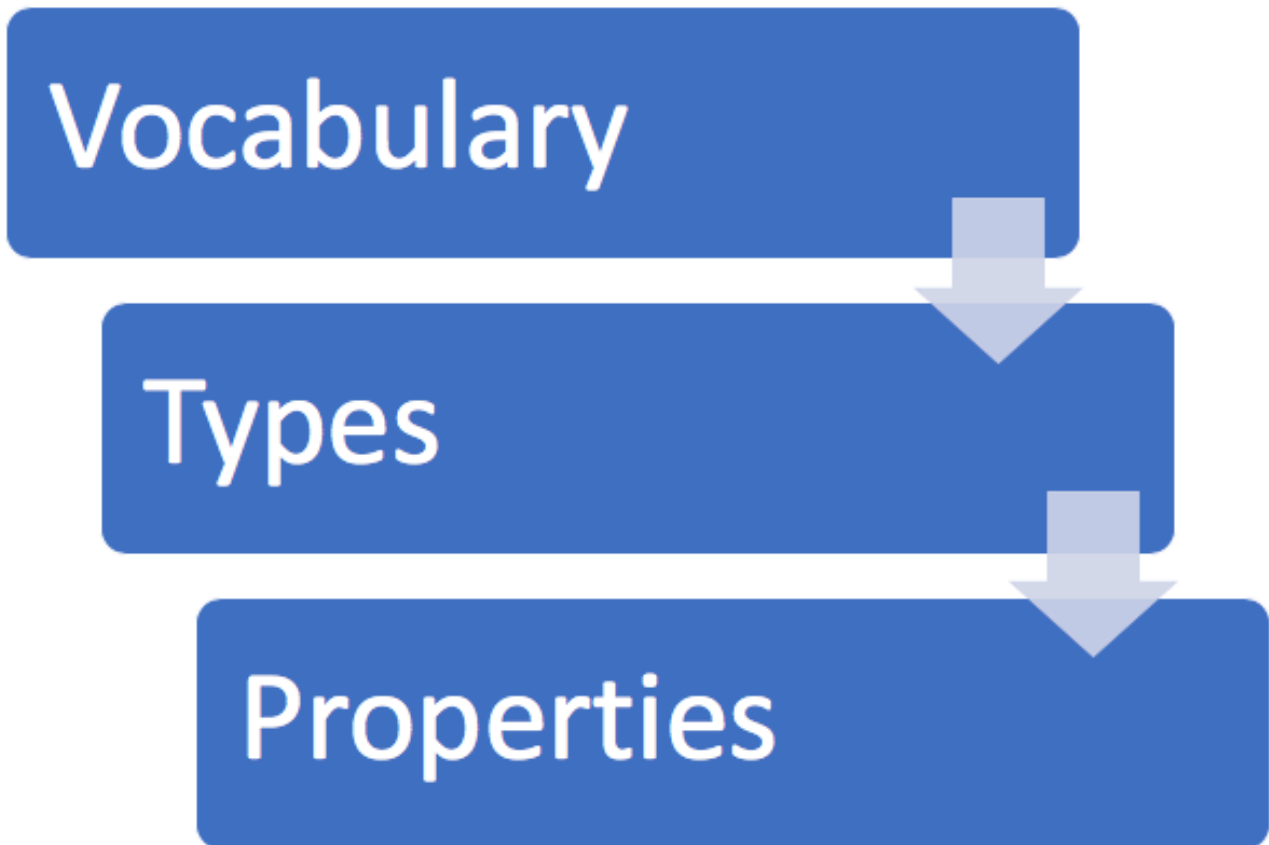
Figure 1: How schemas are organised

The benefits are an increased online visibility, attracting relevant users interested in consuming the content.

The advantage w.r.t. a traditional website is three-fold:

1. An unambiguous identification of *Entities* (Things, Places, People, Actions, Data, etc.) allowing easier more reliable machine learning.
2. The creation of a knowledge graph[4] which, from a practical point of view, is an edge/node model (RDF graph[5]).
3. The possibility to network and interconnect the knowledge graph with bigger more complex knowledge graph models – ideally local travel & tourism datasets.

The end result is the transformation of an HTML website into a data model which refers to attractions and services (Entities) available to tourists.

For example, the Hotel Schema[6] provides the opportunity to describe a hotel, identified as a local business, in a particular location, offering a wide range of services for tourists. The hotel would connect to a higher-level knowledge graph describing the city. In the same way, museums and other tourist attractions would connect to the city and establish an indirect connection between the city, hotel and museums[7].
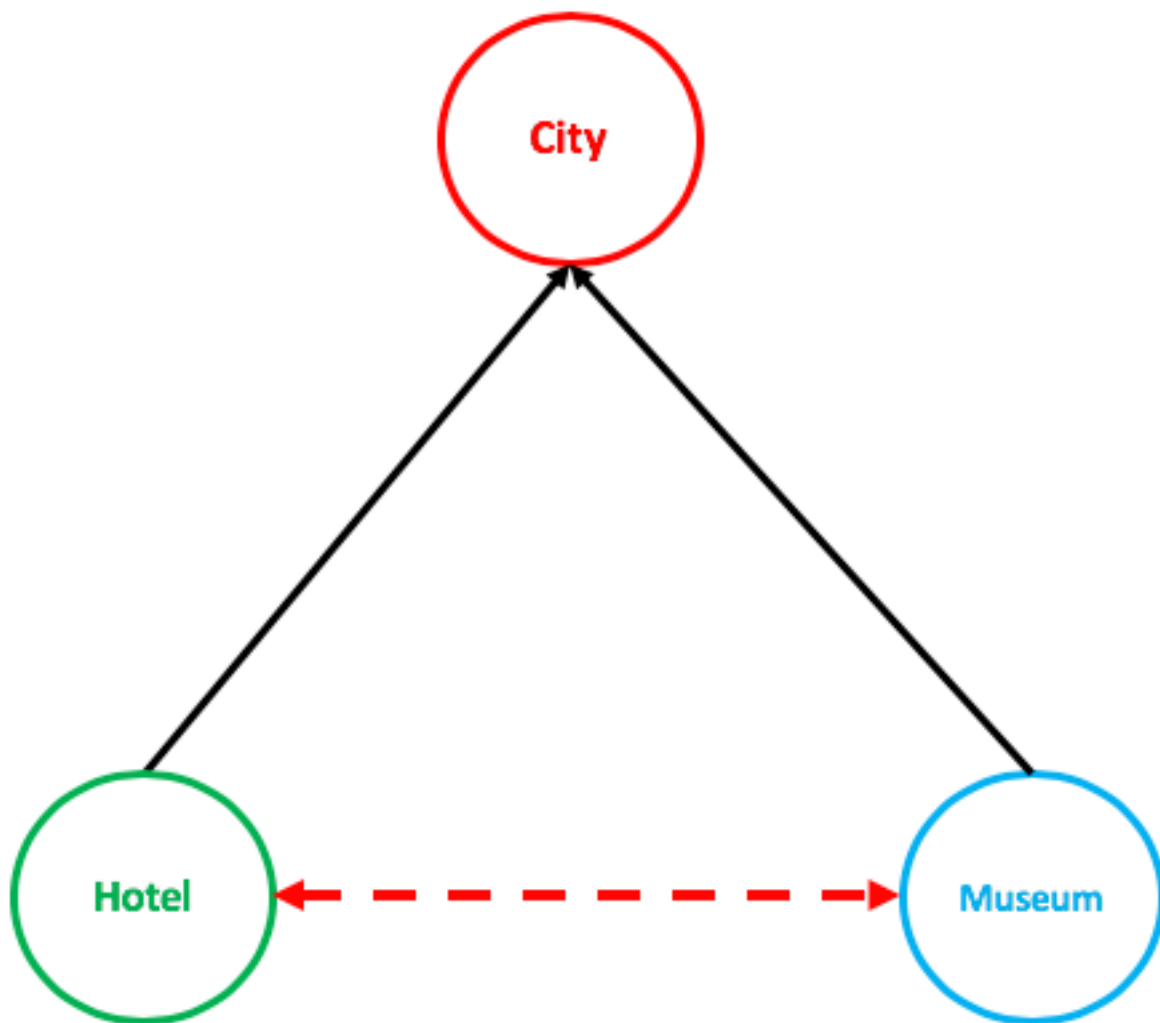
This approach is "fail safe". As more websites adopt the technology, the overall network gains strength: each knowledge graph contributes in the creation of a "Smart Grid" of information, with no need for centralised management (too expensive and ultimately impossible to achieve).

Implementation is achieved via JSON-LD (JSON for Linking Data[8]). JSON-LD is a means to create a network of standards-based, machine-readable data across Web sites. It allows an application to start at one piece of Linked Data, and follow embedded links to other pieces of Linked Data that are hosted on different sites across the Web. JSON-LD is non-intrusive and does not interfere with existing website layout or design and can be implemented anywhere at page level.
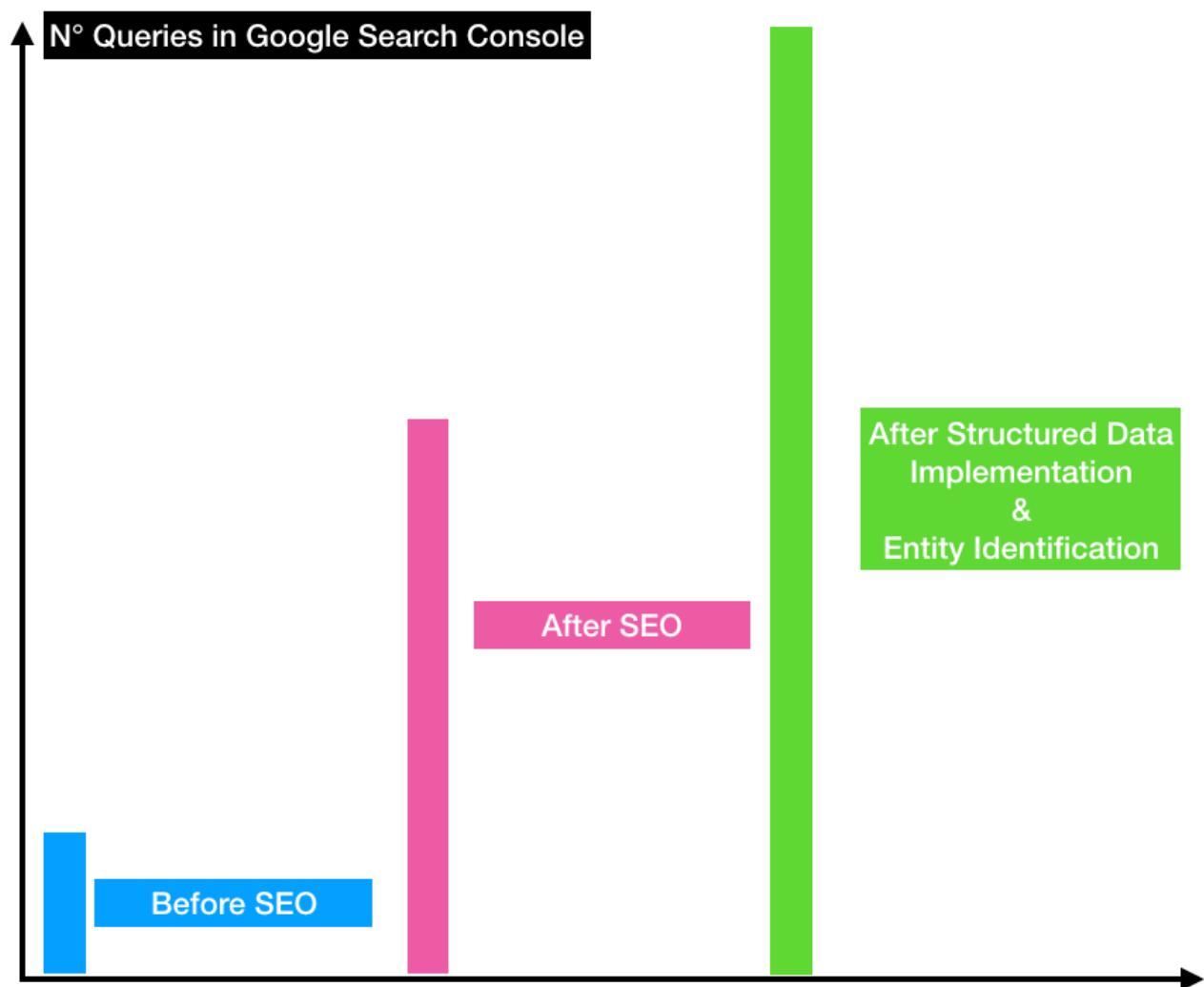
Figure 3 - Benefits of Structured Data and Disambiguation

Fig. 2 offers a qualitative representation of benefits after implementing knowledge graph and entity identification for a travel related website operating in Sardinia. The market segment is yacht services and logistics. The analysis reveals a transformation of quantity and quality of organic traffic[9] with a significant increase in commercial contacts and request for services.

An Integrated Approach to the Semantic Web

There is very limited awareness of the semantic web technology amongst most webmasters and digital marketers. Innovation is this direction is hindered by limited awareness of the benefits, but mostly by the complexity of the topic.

The marketplace suffers from both a skillset shortage and lack of specific Vocabularies dedicated to local attractions and services relevant to tourists.

A pan-European educational effort and consequent creation of standards inspired by The Web Content Accessibility Guidelines (ISO/IEC 40500[10]) and Ontology Development[11]

Universities and public institutions could foster this new approach to web design and development for travel & tourism. A higher-level initiative to fund local institutions would trigger the level of attention such an initiative would require.

References

[1] Semantic Web for the Working Ontologist, Dean Allemang, 2nd edition – Jim Hendler

[2] Schemas are a set of 'types', each associated with a set of properties. The types are arranged in a hierarchy - https://schema.org/docs/schemas.html

[3] What is a Vocabulary? - https://www.w3.org/standards/semanticweb/ontology

[4] Graph Theory, Victor Adamchik (2005) - https://www.cs.cmu.edu/~adamchik/21-127/lectures/graphs_1_print.pdf

[5] RDF 1.1 Concepts and Abstract Syntax - https://www.w3.org/TR/rdf11-concepts/#data-model

[6] Hotel schema in schema.org - https://schema.org/Hotel

[7] NATURAL LANGUAGE INFERENCE, Bill MacCartney (2009) - https://nlp.stanford.edu/~wcmac/papers/nli-diss.pdf

[8] JSON-LD is a lightweight Linked Data format. It is easy for humans to read and write. It is based on the already successful JSON format and provides a way to help JSON data interoperate at Web-scale. JSON-LD is an ideal data format for programming environments, REST Web services, and unstructured databases such as CouchDB and MongoDB – https://json-ld.org/

[9] The term defines users who visit a website by clicking through on an unpaid (or *organic*) search result in a search engine (Google, Bing, etc.)

[10] Web Content Accessibility Guidelines (WCAG) 2.0 - https://www.w3.org/TR/WCAG20/

[11] ISO 19150-2:2015: Geographic information -- Ontology Rules for developing ontologies in the Web Ontology Language (OWL)