



> Semantic Web Use Cases and Case Studies

Case Study: Contextual Search for Volkswagen and the Automotive Industry

William Greenly, Charles Sandeman-Craik, Yago Otero, and John Streit, [Tribal DDB UK](#)

October 2011



Introduction

The contextual search project marks a hallmark of achievement in Volkswagen content and data strategies. It represents technical innovation in one of its truest forms and presents a paradigm shift in the way Volkswagen uses and shares data across the web.

Objectives

A new site search and browse engine means we can get people to the information and content they want, quicker. Getting people the information and content they need quicker inevitably enables people to make decisions more quickly and ultimately increases the rate of conversion. Equally, the ability to promote and re-purpose content in different contexts along with the facility to suggest and find related things that weren't directly searched creates a means to surface seasonal, campaign related or tactically important information.

However as the solution to the problem became more apparent, long term and strategic opportunities were identified that would span the immediate need and many more.

Approach

Initial concepts were drawn up in Jan 2010 with the emphasis on finding an off-the-shelf solution. Two product candidates were identified, [Google GSA](#) and [Apache Solr](#). By May 2010 both products were prototyped and on trial. However, it soon became apparent that Google GSA didn't provide any benefit and came with a heavy price while Apache Solr on its own wouldn't provide much benefit, but could be extended using proprietary/closed schemas. It was therefore concluded that:

- Investing time in developing a proprietary/closed schema in Solr had little value
- It would be better to identify existing web vocabularies by which to describe content
- And develop and publish a new vocabularies to describe domain specific concepts not already covered
- Ultimately enabling unstructured site content to be combined and enriched with structured data

Solution Design

In summary the solution was rationalised as follows:

- Annotate HTML content with RDFa using relevant vocabularies, enriching existing unstructured content
- Transform existing product data into RDF using relevant vocabularies
- Extract RDF from RDFa using [Nutch](#) and merge with Product Data
- Reason over the merged dataset
- Post complete dataset to StructWSF (Solr/SPARQL wrapper)
- For syntactic search, use StructWSF search endpoint
- For SPARQL query, use SPARQL endpoint

This is illustrated in the diagram below:

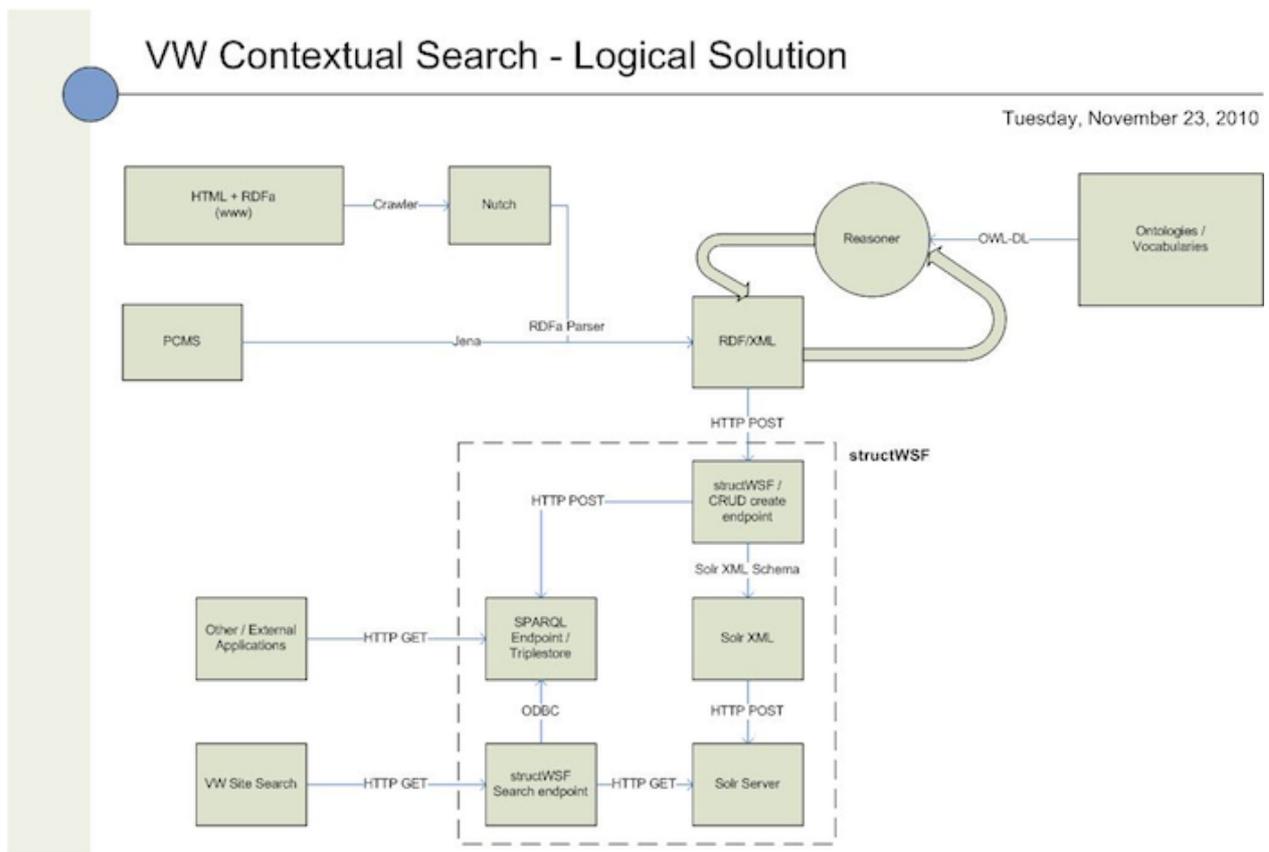


Figure 1. Solution design. (A [larger version of the figure](#) is also available.)

The core component was the [StructWSF](#) framework, an opensource project providing interfaces and wrappers to both SPARQL and Apache Solr acting as a binding between the two.

Existing Vocabularies

The following existing vocabularies were deemed relevant for site and product content:

- [Dublin Core](#)
- [Good Relations](#)
- [Vehicle Sales Ontology](#)
- [Ontology for Innovation](#)

New Vocabularies

Where terms did not exist, new vocabularies were created. These were procured from Prof. Dr Martin Hepp, author of the popular Good Relations vocabulary and Vehicle Sales Ontology. The vocabularies created consisted of a generic, industry wide [Car Options Ontology](#) to describe models, trims, derivatives, components and component compatibility along with the [Volkswagen Vehicles Ontology](#) to describe Volkswagen specific concepts. The former extends both the Good Relations vocabulary and the Vehicle Sales Ontology, whilst the latter extends the former.

Benefits

As well as the functional benefits for the site and site search, the system and its components also provide operational benefits for all Volkswagen 2011/2012 projects. New search facets can be added easily without costly additional development, enabling content to be reused and repurposed across the site within different contexts in an "Elastic search", as visualised below.

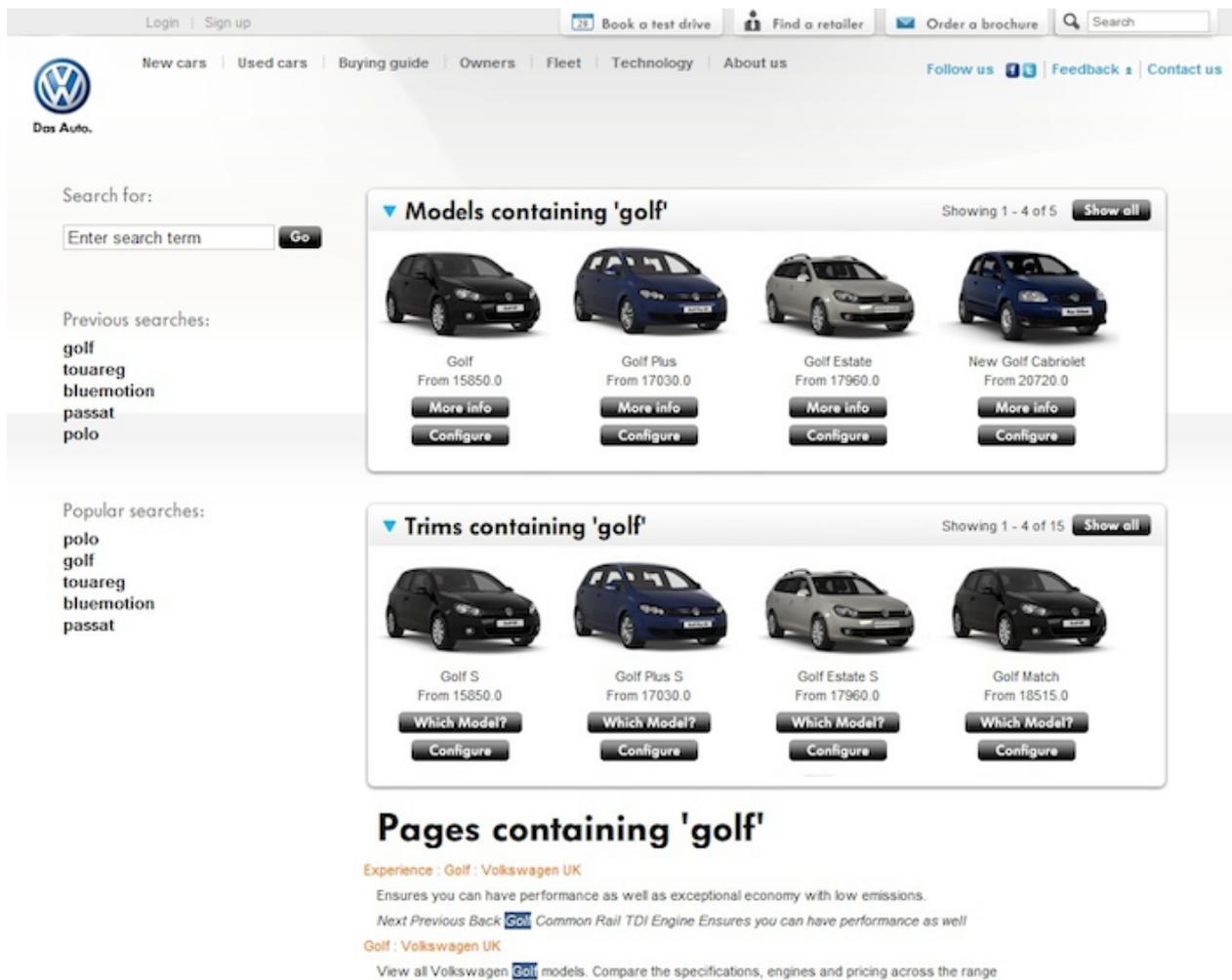


Figure 2. Example for the usage of search facets. (A [larger version of the figure](#) is also available.)

However it also provides a vast range of functional and non-functional opportunities for future development and the features can be seen as both foundational as well as behavioural. It is clear that the components and platforms delivered in this project will be reused across others.

Functional Benefits

One of the key advantages provided by the system is the degree of expressivity open to application developers. Whereas previously searches were syntactic, based on keywords and phrases, across unstructured and meaningless content (in the eyes of a search engine), we have moved to a model of semantics where meaning and aggregation can be derived and applied. By combining the expressivity and richness provided by semantics and structured data, we can start to apply semantics to unstructured content and data, enriching our knowledge base.

Eg:

Find me all the derivatives priced between x and y that have an engine power greater than z and come in red

Additionally by using ontology and knowledge representation we can search for data across a range of contexts, combining different sets of content in different contexts e.g:

Find me all the derivatives priced between x and y that have an engine power greater than z and come in red and have a user review of greater than 4 stars

Finally by harnessing the features of SPARQL 1.1 federated extensions we can seamlessly cross domains, including data from other 3rd party datasources in a single expression e.g:

Find me all the derivatives priced between x and y that have an engine power greater than z and come in red and have a user review of greater than 4 stars. Include results not only from VW, but also other car manufacturers, external car review sites and second hand car dealers.

Case Study: Contextual Search for Volkswagen and the Automotive Industry

Operational Benefits

A key deliverable of the contextual search project was a sophisticated search service delivered through a SPARQL endpoint. This is the highest standard achievable on the web and represents the pinnacle of web technology.

It provides a standardised web interface to content, negating the requirement for a developer to write new publishing code every time the organisation wants share data with a 3rd party. Furthermore, system administrators don't have to maintain multiple architectures or systems lending a separation of concerns.

It allows the organisation to reduce technical dependencies between projects and technical workstreams and opens a creative fiat for developers with different skillsets, technology stacks and frameworks to work with meaningful, rich and relevant data.

External benefits

Components of the project were extended to third parties, in particular the Used Car Locator application, a system incorporating processes, functionality, and content outside of the immediate domain. One such manifestation of this consisted of a separate and distinct website served into the main Volkswagen website via an IFRAME, containing data rich content not suitable for traditional website crawling. Additionally, the sheer quantity of content would distort search results obtainable a syntactic / unfaceted search application.

It was therefore decided to align their used car database using a SPARQL endpoint, in order to obtain identifiers for all models, trims and derivatives. Leveraging recommended vocabularies to annotate their web pages with RDFa, provided a means to publish RDF content gracefully without undue impact on existing processes or operations.

Additionally, it also allowed the search appliance to facet used car data within the existing search result as well as across other contexts or applications

Equally important was the ability to expose structured data not only to internal crawlers, but also third-party search crawlers such as Google and Yahoo, supporting the extrapolation of rich snippets and structured data from HTML.

Key Benefits of Semantic Web Technology

In summary the key benefits of using Semantic Web technology for Volkswagen were as follows:

- A standardised interface to data and content, accessible to developers with different skillsets, using different technologies within and without the organisation.
- Separation of concerns between information and application, both logically and physically.
- Increases value, reusability and accessibility of data.
- Very powerful federation features.
- Adoption and use didn't necessitate process or change management. It could be leveraged at any stage within the product lifecycle painlessly and gracefully, both internally and externally.

Future Projects

The SPARQL endpoint and related systems are also intended to support the delivery of digital content to retailer showrooms, mobile and handheld devices. Additionally, the datasets exposed via SPARQL will be increased to include NVS codes and new car stock along with registration lookup data, all combining and referencing historical model data thus cementing Semantic Web technology in the heart of the Volkswagen platform.

Conclusion

The milestones of this project act as a benchmark for the automobile industry, not only informing how Volkswagen organise their data, but everyone else. Volkswagen's long running tradition of setting and maintaining high standards within the automobile industry is only complimented by the achievements of this project.

"Besides improving search, the contextual search project has helped us transport complex datasets into a number of projects in record time. We really hope the ontologies will benefit not just us, but the automotive web community at large."

References

- [Dublin Core](http://purl.org/dc/terms)(http://purl.org/dc/terms)
- [Good Relations](http://purl.org/goodrelations/v1) (http://purl.org/goodrelations/v1)
- [Vehicle Sales Ontology](http://purl.org/vso/ns) (http://purl.org/vso/ns)
- [Ontology for Innovation](http://purl.org/innovation/ns) (http://purl.org/innovation/ns)
- [Car Options Ontology](http://purl.org/coo/ns)(http://purl.org/coo/ns)
- [Volkswagen Vehicles Ontology](http://purl.org/vvo/ns) (http://purl.org/vvo/ns)
- [StructWSF](http://openstructs.org/structwsf)(http://openstructs.org/structwsf)
- [Nutch](http://nutch.apache.org/)(http://nutch.apache.org/)
- [Google GSA](http://www.google.co.uk/enterprise/search/gsa.html) (http://www.google.co.uk/enterprise/search/gsa.html)
- [Virtuoso](http://virtuoso.openlinksw.com/)(http://virtuoso.openlinksw.com/)