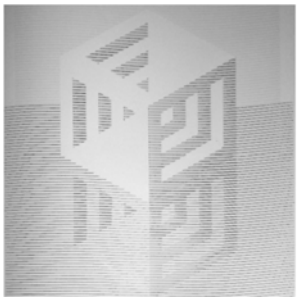
Semantic Web
Thesaurus Management
Linked Data
Ontology Search
Graph Databasis

RDFLink
SKOS Shuttle

Handout 2022

semweb

Semantic Web Technologies · Records Management
Software Development · ICT Consulting · Project Management



Semantic Web and Linked Data

1

Emerging Technology

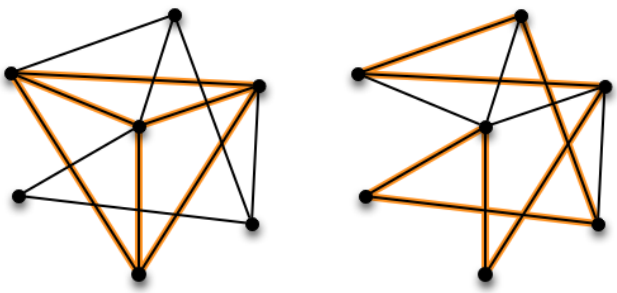
The concept of semantic web goes back to Tim Berners-Lee and is being standardized and developed by W3C. Semantic web is an emerging technology that allows the linking of data from various sources through the use of URIs. A similar principle is used by today's websites, where links to other websites are found. The data in semantic web are not linked to each other over websites but rather over so called triplets via URIs. A triplet always contains 3 elements in the form of a simple sentence: subject – verb/predicate – object. Everything that exists in a triplet is data or metadata.

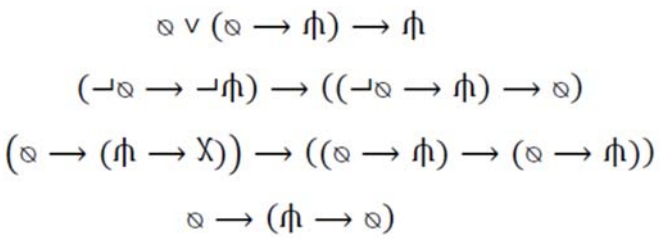
To describe a complex object or, for example, resource "X", a finite (a priori not firmly defined) sentence of triplets is used, whereby all elements have the same subject "X".

Each property of "X" is thereby defined through a triplet; "X" is given a property (verb/predicate) and a value (object). [Triplets were extended to so called quadruples ("quads") already in 2008. The fourth element in a quad is the context, in which the triplet statement should be valid.] Interestingly enough, resources such as "X" can also be linked through triplets, through which a semantic graph is created, whereby each node represents a subject (or an object) and the edges that link the nodes connote the properties (verb/predicate).

A semantic graph can contain thousands of similarly connoted concepts. Data in semantic web is available in graphs and is normally publicly available. This sort of public, freely available data in the form of semantic graphs is very easy to navigate through. It provides precise information, information relationships and enormously reduces the search time. According to the current acronym we speak of "LD data" or simply "LD" (linked data) data or also of "LD cloud" or RDF.

By the end of 2011, LD data was estimated to comprise 30 billion triplets and approximately 500 million semantic graph interconnections.
A resource can therefore be defined through a priori unknown number of triplets – not necessarily in the same semantic graph (!). The resource is consequently described in a semi-structured manner i.e. with a structure that can be variable in both time and space. The LD technology necessitates standard formats such as RDF – Resource Description Framework.





Why Semantic Web?

Semantic web is not only a visualization of LD but represents a series of technologies that act upon LD visualizations with logical operators and thereby allow to discover (validate, infer) “new” (intensional, implicit) data. So called “Reasoners” (computer programs that can process semantic web or LD representation formats) can find data within semantic graphs fast and safely, where otherwise a simple search query would deliver thousands of search results. Complex problems can be quickly and efficiently resolved through this approach. Semantic web technologies (precise representation and logical processing) are therefore indispensable for scientific companies because:

1) company data can be linked precisely yet flexibly (semi-structured),
2) knowledge units (standards, practices, rights…) about company processes can be modeled modularly, flexibly and sustainably (knowledge management),
3) company data on supply and delivery can be linked precisely to one another so that a homogenous, easily navigable representation of data is possible. The time required for search, report and response time is therefore reduced drastically.

Company knowledge can be provided in ontologies – but how?

Domain specific knowledge (standards, practices, processes, rights, empirical results) is usually held for decades in the form of structured documents. Especially expert knowledge should be appropriately transferred into LD form, where its special behavior patterns are documented. The resultant document – a semantic sub graph – is known as “ontology.” Ontology is a systematic representation of knowledge, using subjects and objects within a domain. Different branches use and share ontologies. Thus, ontologies describe special facts and can therefore be considered to be a special case of semantic graphs. For example, geneticists have been using ontologies for quite some time .

How do you get to the LD cloud?

Lead by the necessity of building a sustainable, clearly structured and easily navigable information platform, you want to “LDify” * a part of your company data in order to be able to:

1) carry out precise searches through easy application,
2) gain or maintain an overview over data,
3) link company data with other data precisely and flexibly.

*

“LDify” (i.e. translate from common data sets into RDF data sets, also “RDFy”

3

An actual condition analysis will first determine the type and form of your data. In the resultant specifications, the target services and the way they should be carried out shall be determined.

The major steps of the RDFifying process are:

a) careful selection of vocabularies to describe the data;

b) the data in question will be mapped onto semantic graphs; triplets are created here, which can be suitably hosted in your company

c) development of the specific applications that are necessary for your data processing.

To operate the semantic graphs as RDF data, opportune scaling measures are implemented that guarantee application speed.

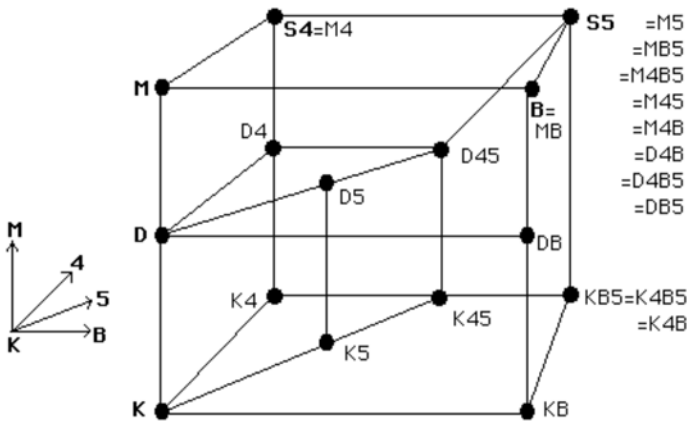
Does my company have to make all data public?

Although RDF data principally requires and supports publication, their barrier lie however in legal and competition law limitations. A company should not and must not publish all RDF data. However it can profitably make use of the connected semantic web technologies internally and with trusted partners.

Since RDF data requires web technology per se, the access to this data is placed in suitable URIs in the RDF repository. These URIs make use of a suitable rights model that protects the web area in question and thereby keep the data non-public. With the use of today's technology, access to a semantic graph – be it for internal or controlled use – is therefore suitably protected. If participant users or market forces wish to combine certain areas (e.g. suppliers), it is possible to simply authorize the openness of a semantic graph through the use of web technology.

What is the "price" of RDFifying?

To ensure the precise and flexible processing of RDF data associated with semantic web technology, it is necessary that RDF data be logically built and maintained according to standards. Future use of company data implies that a proper data mantainance be provided. Thus, the tasks of a knowledge engineer constitute here a very important role in the company.

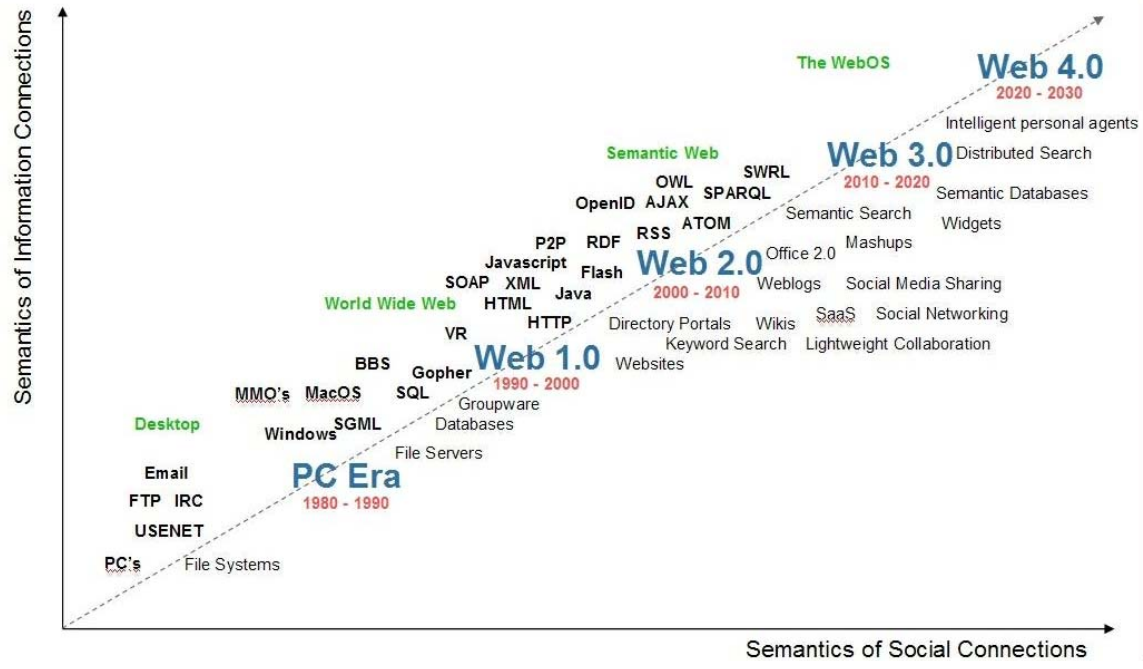


What are the risks of RDFifying?

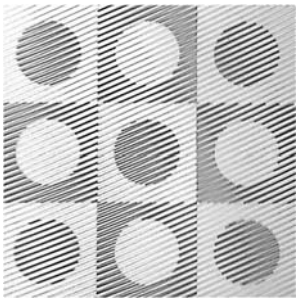
We see the risks of RDFifying fundamentally in an inconsistent state of RDF data over time. If the meta data portion is not properly maintained, the associated semantic graph becomes out of date, shows “logical holes”, or contains inappropriate concepts. RDF data organized to semantic web technology follow a logical structure – similar to a special bill of materials.
The existence of an error in the structure can therefore lead to large-scale losses upon acquisition requests. A further risk could lie in the (possible) logically weak modeling of data. In this case, linking or attaching RDF data that will be available in the future is only partially possible, if not completely impossible. Too weakly RDFifyed data chunks will necessitate remodeling. This can however also be the case for entire semantic sub graphs.

How do RDF clouds behave with regard to data archiving?

According to the principle of records management, only if their relevant record class should be archived due to legal, structural or timely requirements, data are destroyed or archived. Suitable reports enable the archiving of RDF content. Archiving turns the necessary data into a readable and durable form. Connecting archived company data to the RDF cloud is not further necessary.



semweb



Semantic Web in practice

Semantically finding the right thing

Ontology / Thesaurus Management
Find your way through large data and search spaces! Make use of the advantages of domain specific SKOS (Simple KnOwledge Systems) based Thesauri together with text-mining technologies and find previously unknown valuable results.

Semweb

Semweb efficiently accompanies you during the semantization of relevant parts of your company, especially during the integration of your data bases/CMS/DMS/ERP systems into a fast semantic search space, within and across sectors. Specific RDF (but also conventional) thesauri, encyclopedias such as DBPedia, as well as specific scientific ontologies enrich your information systems with meaningful semantic components. This leads to new advantages and functionalities such as:

Faceted search results
At the click of a mouse, you can successively set limits to the range of results in order to quickly find what you need in your data/search space, facet after facet (a facet is a search perspective or category, such as “economy” or “scientific article”).

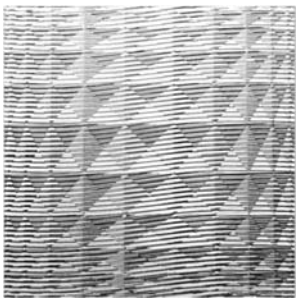
Autocomplete with linked thesauri
As you type in your search query – or partially type it in – you are provided with search suggestions that are acquired through the selected domain specific or domain encompassing Thesauri. Additional SKOS indications suggest abstractions, specifications, and synonyms of your search query and speed up the process even before the real search begins.

Query expansion
Thesauri allow a precise expansion of the search within limited areas, where the available encyclopedias are consulted according to the SKOS model. Not only are results of varying languages possible, but also results across linked encyclopedias.

More like this Search
Have you ever found a document and asked yourself if there are similar documents in the search area? Sophisticated text mining algorithms provide further search results according to their similarity to the reference document (area of application e.g. observance of the competition or scientific results).

Tagging
(NER – automatical named entity recognition)
Terms from a text in a natural language such as English or German are given a classification and meaning in a previous analysis and the document is indexed with this. Your information space will become more precise and appropriate search results will be found faster.

Advisory Systems
Guide potential and existing clients through their queries and provide consultation: an advisory system with chatbot technology – supplemented by semantic technology – can be the ideal solution. It can allow your clients to reach their destination automatically and save time.



LD Management

Linked Data are meshable data built on top of RDF Data.

What is RDF?

Linked open data (RDF) describes freely available data in the World Wide Web that can be identified through Uniform Resource Identifiers (URIs) and can thereby be directly accessed via HTTP. Data can also be linked to other data using again URIs. In the ideal case data are represented and linked using the "Resource Description Framework" (RDF) as well as standard technologies based on RDF like "Protocol and RDF Query Language" (SPARQL) and the web ontology language (OWL), thus making linked data part of the semantic web. This linked data constitutes a worldwide network that is also called "linked (open) data cloud" or "giant global graph". For all those cases where the focus is less on the open and free usability of the data we speak simply of "linked data".

Who profitably uses RDF?

Public institutions such as NGOs, municipalities, electronic libraries, archives, broadcasting corporations, hotel and airport systems use RDF with the purpose of allowing a uniform use of data via the web.

Who has implemented it so far?

In lod-cloud.net is a current example of institutions that have turned their information to RDF form. According to lod-cloud.net/state the RDF "cloud" amounted to approx. 30 billion records in 2011 with more than 500 million outgoing connections.

The challenge

Each institution maintains its data in its own data bases, intranets and repositories. The manner in which this knowledge is stored remains proprietary, i.e. in such formats that can only be processed by given commercial applications. The implementation of one's own data base into a RDF repository requires fast web technologies as well as special modeling knowledge and needs standards.

How does Semweb support me during RDFifying?

Semweb

assists you through every step of RDFifying – including support and operation of RDFified data.

We analyze your databases or relevant parts of it, and make modeling suggestions in several variants (if possible) on how your represented data knowledge can be modeled and structured using appropriate vocabularies. We implement the variant that is chosen by you, with the guarantee that the mentioned RDF repository will be operative within the agreed time frame.

semweb services

Thesaurus Management in the cloud

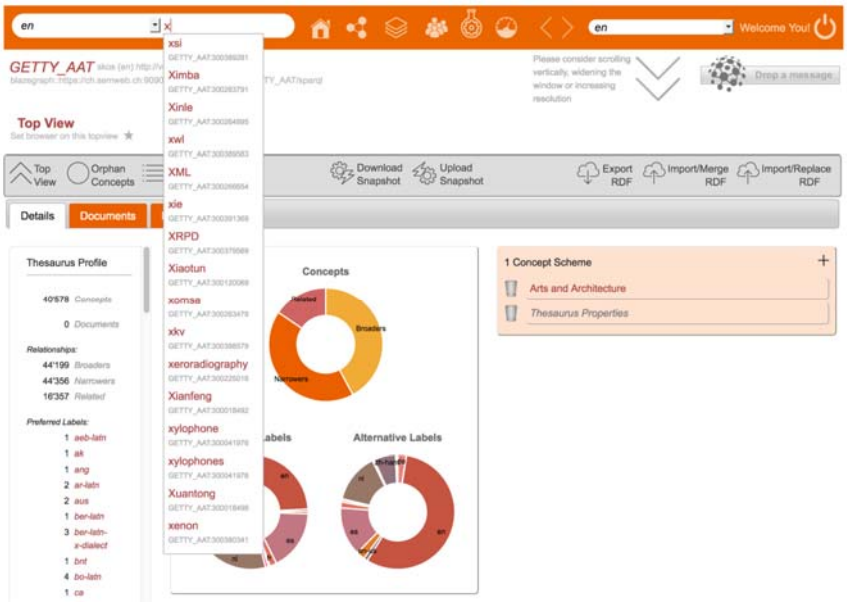
SKOS Shuttle

Thesaurus Management as a Service

Thesauri are controlled vocabularies whose concepts are connected through relations. They build a systematic representation of domains. Thesauri are a generalisation of taxonomies; they are de facto polyhierarchies where a concept is related to its synonyms, broader or narrower concepts (but even using any kind of relationships).

In Semantic Web thesauri are ideally organized according to the SKOS model. This allows for standard access to concepts and their interrelated concepts (interoperability of data sources).

The use of thesauri in Enterprises empowers concept maintenance and allows for fast and precise search processes. A further use of thesauri is in the **corpus analysis**, where – starting from a text corpus and a thesaurus – relevant information on term frequency, distribution and potentially new concepts for the same thesaurus are computed. Furthermore thesauri allow in linked data spaces the precise interlinking between key concepts.



Semweb supports you in building, maintaining and operating taxonomies, thesauri and ontologies with SKOS Shuttle or the TMS of your choice.

Benefit from this chance and get yourself a free account:

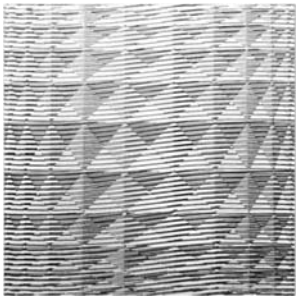
Step 1 Request your free account
Step 2 Define as many thesaurus projects as you need, from scratch or upon RDF import
Step 3 Navigate or export the Thesaurus data in up to 6 RDF formats.

Advantages of SKOS Shuttle in a nut shell:

- Uses internal but even external RDF repositories (e.g. Blazegraph™ or Stardog™)
- Accesses securely all information (SSL/DIGEST)
- Rights, Groups, User and Project Management
- Powerful autocomplete in expert/essential mode
- Any number of documents associable with every concept
- Collaborative time zone concept
- Historization of changes
- Systematics Assistant helps you with different descriptor systems
- Orphan Concept Analysis and direct “deorphanization” without SPARQL
- RDF Triple editing – if necessary
- RDF Import/Export in 6 formats
- RESTful API for a complete, secured access to your thesauri

Find out more about our services on semweb.solutions/skosshuttle

semweb



RDF Services

RDFLINK

- **Linked-Data from tabular data**
- **Generation of Linked-Data in the Cloud**
- **Linked-Data as a service**

Nowadays the biggest part of enterprise data is hosted in relational data bases systems. This causes operating silos which have to be joined by often expensive data federation and/or migration projects. Thanks to the inherent isomorphismus between relational and linked data and to the much higher interoperability and flexibility of the latters, it is advisable to switch to linked data as a general data basis in the enterprise.

With RDFLINK You develop a linked data model, then generate linked data from your relational data is now it is much more like a cakewalk.

Using RDFLINK enterprises can easily generate their linked data and fill them in to RDF repositories at a finger snap.

During one of its successful LD projects Semweb developped RDFLINK, which supports all the mapping stages from tabular to RDF data

Learn RDFLINK and get your first free full operable account on semweb.solutions/rdflink

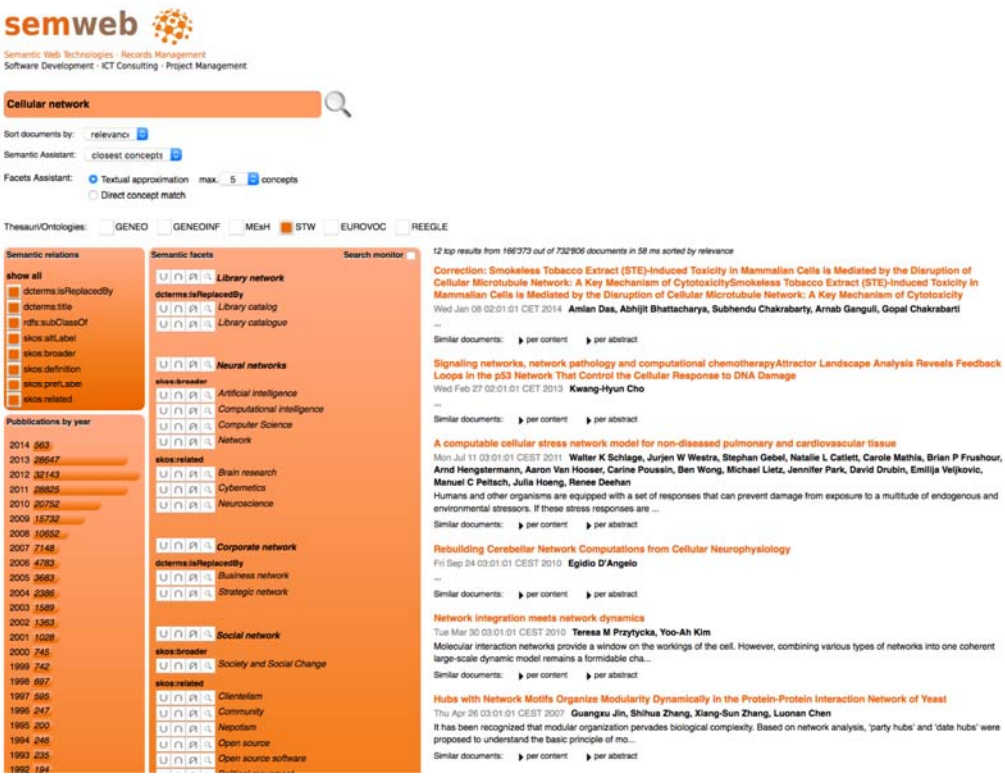
Ontology Assisted Search

Document search in a wide document space can often deliver too many results. Restricting the results to some few documents might be a time and resource consuming task. On the other way a search might deliver to few results, because neither synonyms nor categories were given. In the first case the search needs a „reduction“ while in the second case it needs an „expansion“ of the results.

Thanks to the use of ontologies (or even of thesauri) reflecting structured information on domains, the search can be effectively trimmed to deliver less (reduction) or more (expansion) results.

In the following demonstration based on medical documents some ontologies are predisposed to be used as a controlling tool in addition to your search query; their relations deliver upon your query semantical facets which can be used to reduce/expand the query results.
Enjoy playing around changing the query, adding/retaining semantic relations/categories or changing the way how semantical facets should influence the results.

Learn Ontology assisted search on
https://semweb.solutions/leistungen/rdfservices/en- semsearch/

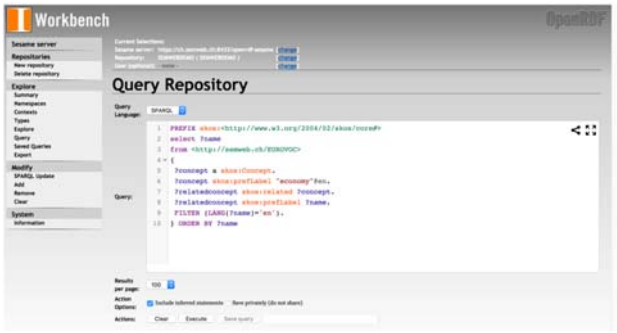


Graph Databases – RDF

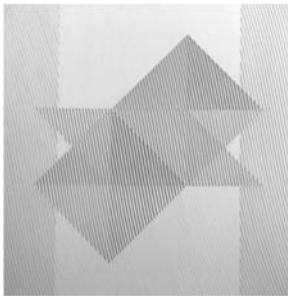
Semantic data are stored nowadays in RDF stores – **graph databases** – and are queried by humans or by programs by means of query languages like SPARQL to get logical connections among several semantic graphs or even among different semantic databases, giving a method to access data and meaning across several data sources.

Semweb supports you during the phases of selecting and integrating your RDF store. Currently we use some of the best RDF stores through their SPARQL endpoints: Bigdata™ (SSL), Stardog (SSL), GraphDB™ (SSL), MarkLogic™, Virtuoso™.

To explore some features of the graph database please consult https://semweb.solutions/leistungen/ rdfservices/en-rdfstores/



semweb



Paintings: Renato Ricci 1922-1995

Specialist for Semantic Web.

Launching, Planning, Realization and Coordination of complex Projects for Industry and Research.

semweb / Fabio Ricci

Business location	Via Spluga 3 23854 Olginate (Lecco) Italy
Executive	Fabio Ricci
Year of foundation	2007 [previously WebDMS, semweb GmbH]

Services	■ Universities Support ■ Consulting, Expertises, Planning, Project Leading and Realisation of complex projects ■ Software Development/Adjustments, Solution Architectures ■ Problem Solving, Rollout Management, Troubleshooting, Test Engineer, Training, IT Moderator, Application Manager
Focus/main activities	■ Semantic Web Technologies, Ontology Conception / Management ■ Records Management, Document Management ■ Thesaurus Management, Text Mining ■ Complex Projects ■ „Mission Impossible“
Support for	■ Problem Solving ■ Rollout Delays ■ Shortage of staff ■ Special knowledge needed ■ Re-Modelling of data structures and processes ■ Semantic Web für die Industrie (Smart Factory, CPS/CPPS) ■ Ontology Conception ■ Semantical Web Service ■ Enterprise-Wiki ■ Knowledge Recycling ■ System Integration

Telephone	+39 389 068 13 34
Contact	hello@semweb.solutions
Website	semweb.solutions

semweb

Semantic Web Technologies · Records Management
Software Development · ICT Consulting · Project Management