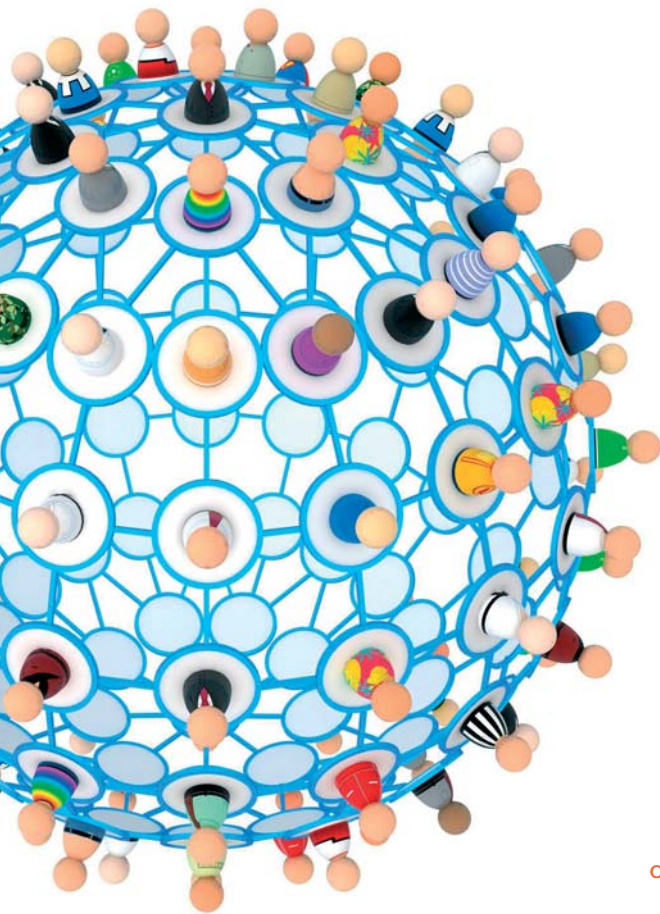# The OWLIM Family
of Semantic Repositories

The BBC Selected OWLIM

After evaluation of RDF engines, the BBC has selected OWLIM to replace MySQL in its Dynamic Semantic Publishing stack, first proven for the World Cup 2010.

OWLIM is Core Technology for the Press Association

Learn more about architectures that build on top of OWLIM, such as at the Press Association, and products that integrate it, such as GATE and TopBraid Composer.

100,000 queries for 1$ in Amazon's Cloud

A cluster of 100 Amazon EC2 nodes running OWLIM demonstrates a combined throughput of 5,000,000 SPARQL queries per hour.

New Features and Recent Benchmarks

OWLIM ver. 4 comes with rebranding of the family and scalability to 1 trillion nodes.
Geo-spatial queries, Support for SPARQL 1.1, OWL 2 RL and QL. The results from BSBM v.3

ontotext

http://www.ontotext.com/owlim

RDF Databases = Semantic Repositories

SEMANTIC REPOSITORIES

Semantic repositories or **semantic databases** are database management systems – their main function is to store and query structured data.

The essential difference compared to relational databases is that semantic repositories can infer non-explicit information using:

- **More expressive schema** definitions, namely ontologies that encode relevant semantics

- Inference mechanisms to **interpret stored data**

SEMANTIC DATA INTEGRATION

Semantic repositories provide an ideal platform for data integration because RDF is designed for the management of data created without centralized control:

- New data sources can be adopted with little effort

- Schema changes are easy to accommodate

EASIER QUERYING OF DIVERSE DATA

A major problem with relational databases concerns evaluating queries over heterogeneous datasets. In order to make a query that retrieves all the relevant information one needs to know the schemata of all the sources (e.g. tables, columns).
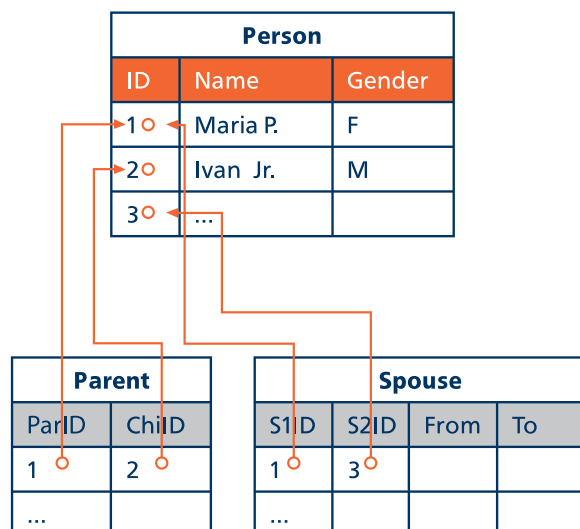
Semantic repositories allow one to:

- **Query data without knowing the vocabulary**, used to assert it, given even a course-grained schema maping

- **Retrieve relationships of unknown types**; a query pattern like "John ?x Marry", asking for arbitrary type of relationship between two objects, is impractical in relational DBMS

RDF REPRESENTS A GENERIC DATA MODEL

- The logical structure of data is not fixed in its physical representation

- Structure and semantics are interpreted, based on RDFS schemata and OWL ontologies

The diagram below illustrates the differences between data representation in a sample relational database model (on the right) and the corresponding RDF model (on the left).

Statement

Subject	Predicate	Object
myo:Person	rdf:type	rdfs:Class
myo:gender	rdfs:type	rdfs:Property
myo:parent	rdfs:range	myo:Person
myo:spouse	rdfs:range	myo:Person
myd:Maria	rdf:type	myo:Person
myd:Maria	rdf:label	"Maria P."
myd:Maria	myo:gender	"F"
myd:Maria	rdf:label	"Ivan Jr."
myd:Ivan	myo:gender	"M"
myd:Maria	myo:parent	Myd:Ivan
myd:Maria	myo:spouse	myd:John

Person

ID	Name	Gender
1	Maria P.	F
2	Ivan Jr.	M
3	...	

Parent

ParID	ChiID
1	2
...	

Spouse

S1ID	S2ID	From	To
1	3		
...			

OWLIM is a Robust Semanitc Repository

OWLIM IS AN EFFICIENT RDF DATABASE

OWLIM is a mature, native RDF semantic repository. Its performance, efficiency and robustness allows it to replace legacy DBMS in a very wide range of applications.
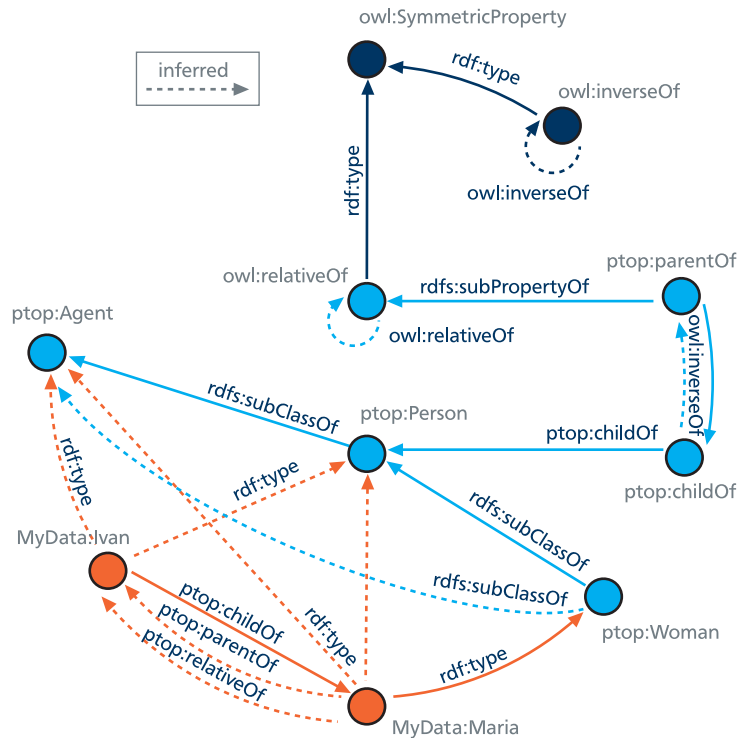
OWLIM is particularly suitable for:

- **Evolving data schema** and query needs
- **Analytical tasks** and Business Intelligence
- Integration of **heterogeneous and sparse data**

OWLIM IS ALSO AN INFERENCE ENGINE

It uses rule-based reasoning to support:

- RDFS, OWL Horst, **OWL 2 RL** and QL
- **Consistency checking**
- **Customisable semantics** via rules and axiomatic triples

MORE INTELLIGENT QUERY ANSWERING

OWLIM offers greater analytical power. A query can match criteria and return results based on data that differs from the query patterns, but bears relevant meaning.

For instance, a query pattern **"Maria relativeOf ?x"** can return "Ivan" as a match based only on the assertion **"Ivan childOf Maria".** In this case the relevant fact was asserted in the inverse direction and using a much more specific predicate (see the graph above).

Naive OWL Fragments Map
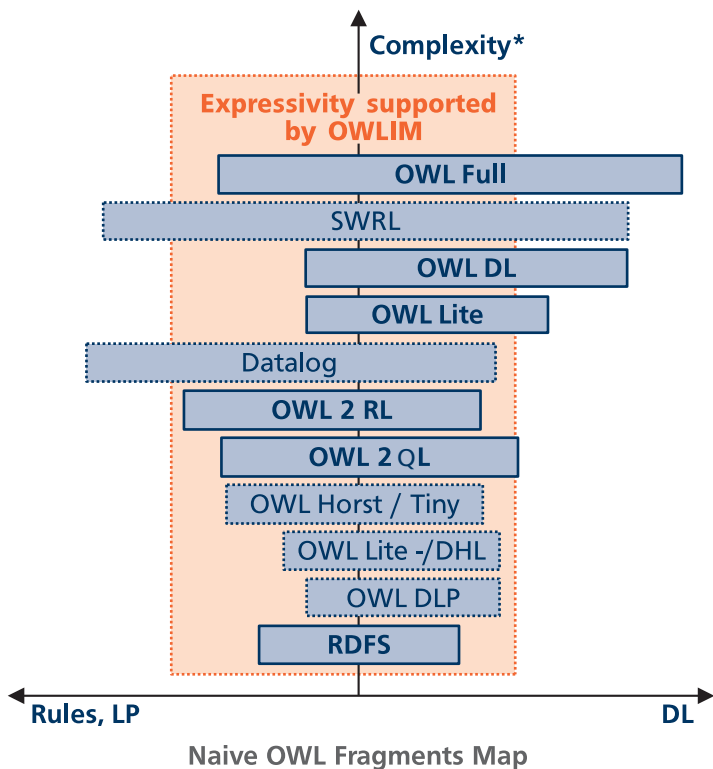
The OWLIM Family

USING OWLIM

END USERS CAN USE OWLIM THROUGH:

- OpenRDF Workbench (Sesame's Web UI routines)
- Ontology editors integrated with Sesame or Jena
- Forest - the front-end library used in FactForge

APPLICATIONS CAN:

- **Embed it as a library** and access it through the Sesame or Jena APIs
- Access it remotely as a **standalone server**
- Access it though a **SPARQL end-point**

OWLIM EDITIONS

OWLIM is a family of RDF semantic database engines that cover the requirements of a wide range of uses and application scenarios:

OWLIM Lite: A free light-weight database. The ideal entry point for researchers and those new to RDF applications.

OWLIM SE: A single-server RDF database platform with transparent reasoning support throughout the entire life cycle of the data and many advanced features, such as native support for geospatial data. Suitable for group-level information needs.
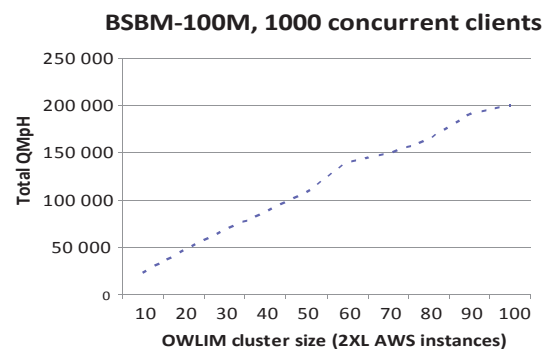
OWLIM Enterprise: All the features of OWLIM SE, in a replication cluster configuration and reliability features to take RDF data management to true enterprise grade.

EASY APPLICATION SET-UP

The distributions of OWLIM have been optimized to minimize the time and efforts for setting up an application:

- **Getting Started**: a sample application that can be used for bootstrapping a solution. One can load and query its own dataset by simple modification of Getting Started, without any Java programming and compilation needed
- **Quick Start Guide**
- **Easy Web server deployment:** OWLIM can be deployed simply by copying a WAR file

BSBM-100M, 1000 concurrent clients

Chart: Total QMpH vs OWLIM cluster size (2XL AWS instances). Y-axis from 0 to 250 000, X-axis from 10 to 100.

OUTSTANDING SPEED AND SCALABILITY

Benchmarking semantic repositories is a challenging task. Few reference results are commented below. More comprehensive information is available at http://ontotext.com/owlim/benchmarking-results. The key references are available on the back cover.

Best scalability: back in year 2009 OWLIM loaded and was able to efficiently query 20B statements.

Best loading speed and query performance: In the latest comprehensive independent evaluation (with BSBM v.3) OWLIM's loading speed was twice faster than the second best and it had best query performance among all the repositories which were able to handle updates and multi-client loads.

Best reasoning: In the last 2 years, all independent evaluations concerning reasoning in semantic repositories establish that OWLIM provides the fastest and most comprehensive inference.

http://www.ontotext.com/owlim/editions

Cluster and Cloud

OWLIM REPLICATION CLUSTER SCALE OUT

OWLIM Replication Cluster is an enterprise grade, distributed database system that provides scale-out handling of concurrent query requests: the **query processing rate scales linearly with the number of worker nodes**.

Updates are first tested on one worker, then pushed to all the other worker nodes.

Read requests are dispatched to the worker node with the shortest processing queue, thus providing simple, but practical load-balancing.
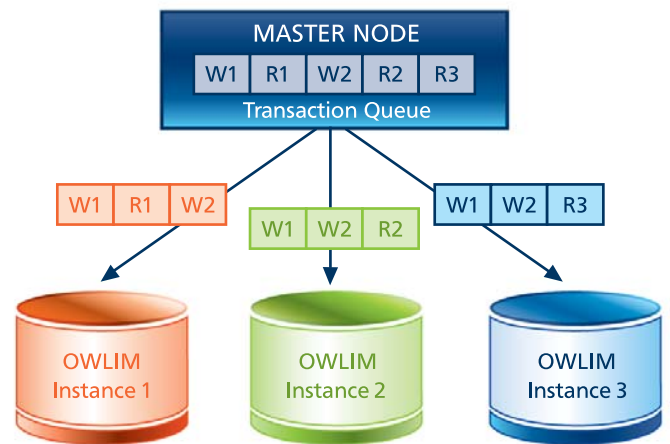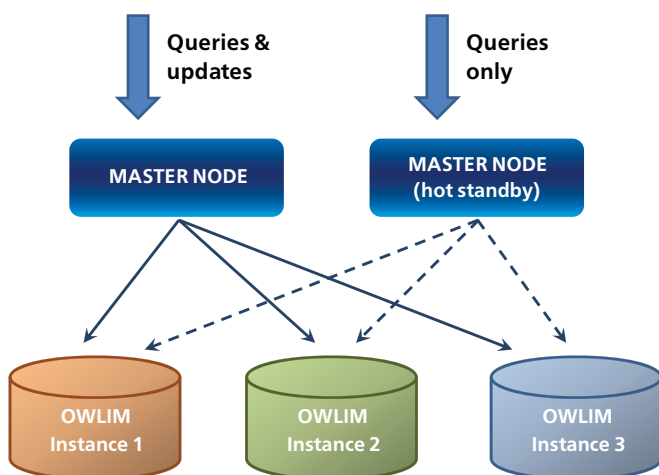
RESILIENCE, FAILOVER, MANAGEABILITY

A cluster is organised as one or more master nodes that manage one or more worker nodes. **Failover and load-balancing between worker nodes are automatic.**

No single point of failure - multiple (stand-by) master nodes ensure continuous cluster performance even in the event of a master node failure.

Online reconfiguration - the cluster deployment can be modified when running, which allows worker nodes to be added during peak times, or released for maintenance, online backup, etc.

$1 PER 100,000 QUERIES IN AMAZON CLOUD

In a BSBM v.2 (100M dataset) test, a cluster of 100 nodes (2XL Amazon EC2 instances) demonstrated:

- Throughput of **5M queries/hour** (200,000 QMpH)
- **Excellent horizontal scalability** - almost linear up to 100 nodes (see the chart on the left)
- **Low parallelisation overhead** - each node in the 100 node cluster had 90% utilisation
- 100,000 SPARQL queries can be answered per 1$ of Amazon EC2 infrastructure cost

BENEFITS OF USING OWLIM IN THE CLOUD

- **Pay-as-you-go cost model**, where customers can avoid upfront investment in hardware and software licenses.
- **Instant, on-demand provisioning** of OWLIM on a variety of hardware configurations.
- **Simple deployment and management**, where the customer does not need to deal with software installation, configuration and upgrades.

Deploying and using OWLIM is just a click away!

http://www.ontotext.com/owlim/cluster

OWLIM in Use

APPLICATION DOMAINS

Over the last 5 years OWLIM has been successfully used in projects and applications in a wide range of domains: from defense to commodities and market intelligence. Follows a list of some of the domains with comments about the typical applications.

LIFE SCIENCES
Integration of large-scale KB in the LinkedLifeData platform, consolidating 26+ biomedical databases into dataset containing more than 5 billion facts.

AstraZeneca uses OWLIM for semantic analysis and search in clinical study reports.

PUBLISHING AND MEDIA
OWLIM is being used for metadata-based content aggregation and publishing by BBC (see p.7), Press Association and other leading publishers.

TELECOMMUNICATIONS
Korea Telecom uses OWLIM for its IPTV infrastructure.

Other large telecommunication companies use it in their semantic business process management systems.

DIGITAL LIBRARIES
The National Archive of UK uses OWLIM to manage its Semantic KB, that provides semantic search for the UK Government Web Archive.

"BigOWLIM is an important and core part of our architecture which we will be building all our products around"

John O'Donovan,
Director, Technical Architecture and Development
Press Association, UK

TOOLS THAT INTEGRATE OWLIM

BUNDLED IN **GATE** AS AN ONTOLOGY SERVICE
GATE is the most popular text mining platform

INTEGRATED IN **PROFIUM METADATA SERVER**
Profium Metadata is used for content delivery by some of the largest news agencies

IN **KIM** PLATFORM AS A SEMANTIC REPOSITORY
KIM is a semantic annotation and search platform

IN **TOPBRAID COMPOSER** AS A REASONER
TBC is one of the most advanced RDF/OWL editors

AS THE DATA LAYER OF THE **LARKC** PROJECT
LarKC is probably the most ambitiuous large - scale reasoning project, http://www.larkc.eu/

IN **INFORMATION WORKBENCH** AS REPOSITORY
fluidOps' Information Workbench for linked data application development

AS SEMANTIC REPOSITORY IN BPENG'S **XBPR**
XBPR is a Business Process repository and semantic enterprise application integration platform

AS TRIPLE STORE IN **MERESCO**
MERESCO is an open multi-paradigm data management system developed in Python

"The Information Workbench™ integrates Ontotext's OWLIM semantic repository for an industrial strength, robust, high-performance and scalable implementation. The Replication Cluster on EC2 is a perfect solution for applications that require dynamic scaling to support an increasing number of clients and queries".

Peter Haase,
Senior Architect, R&D at fluidOperations

http://www.ontotext.com/owlim/usage

DYNAMIC SEMANTIC PUBLISHING

BBC's 2010 World Cup website was the first project running on its Dynamic Semantic Publishing architecture. This approach automates the aggregation and publishing of interrelated content objects according to an ontological domain model.

Dynamic Semantic Publishing provides improved navigation, content re-use and re-purposing, search engine rankings, journalist determined levels of automation ("edited by exception") and supports semantic advertisement placement.

The architecture is now being adopted at BBC SPORT and will be used for BBC Olympic 2012.

OWLIM IN THE ARCHITECTURE OF BBC

OWLIM serves as the RDF store in the architecture facilitating metadata-based management, inference and retrieval of content.

The metadata is added instantly to catch up with the stream of new content. Factual knowledge is updated constantly to reflect developments.

The main function of the repository is to provide selections of media assets (articles, photos, etc.) relevant to a specific subject. Those are used for dynamic generation of web pages on the subject.

REASONING: PURPOSE AND SCHEMA

Here inference exhances the matching between the content metadata and the subjects to take into consideration the semantics of all data.

Materialization-based reasoning ensures high efficiency of query handling. OWLIM's unique schema for fast retraction of inferred statements makes possible frequent and immediate updates of the factual knowledge

"A RDF triplestore and SPARQL approach was chosen over and above traditional relational database technologies due to the requirements for interpretation of metadata with respect to an ontological domain model."

WORLD CUP WEB SITE STATISTICS

- 800+ Dynamic aggregations/pages (Player, Team, Group, etc.), generated through SPARQL queries

- Average unique page requests/day: 2 million

- Average **SPARQL queries/day: 1 million**

- **100s repository updates/inserts per minute** with OWL 2 RL reasoning

- Multi data center fully resilient, clustered 6 node triple store

"We are happy to have chosen OWLIM, after evaluation of several RDF stores, to build a high-performance semantic stack for the World Cup 2010 site"

Jem Rayfield,
Senior Technical Architect, BBC

http://www.ontotext.com/publishing

OWLIM Key Features

Proven to Handle Millions of Queries per Day in Critical Applications
Supports OWL 2 RL and QL Reasoning, and Provides Unmatched
Linked Data Integration, Management, and Retrieval Capabilities

Based on published results and independent evaluations OWLIM is:

- The **most scalable** semantic repository in the World [1]

- The **most efficient** semantic repository in the World [2,3], in terms of speed with
  which it can load and query the data, considering its semantics

- The repository offering the **most comprehensive reasoning** support [3,4]
  It is the only semantic repository which provides OWL 2 RL and QL support [5]

Any combination of 3 of the following features puts OWLIM ahead of the competition:

- **Pure Java** implementation, delivering their full functionality and performance through
  both **Sesame** or **Jena**

- Support for all popular RDF syntaxes and query languages, including **SPARQL v.1.1**

- **Cluster** support brings resilience, failover and horizontally scalable parallel query
  processing for enterprise environments

- **Geo-spatial extensions** for the efficient evaluation of spatial query constraints

- **Optimized owl:sameAs handling**, which delivers dramatic improvements in
  performance and usability when data from multiple sources are integrated

- **High performance retraction** of statements and their inferences. OWLIM handles
  inference in a consistent and transparent manner during loading, updates, querying

- **Full-text search**, based on either Lucene or proprietary techniques

- **RDF rank**, similar to Google's PageRank, can be calculated for the nodes in an RDF graph
  and used for ordering query results by relevance

- 40-bit URI internal identifiers allow for handling RDF graphs with **1 trilliion nodes**
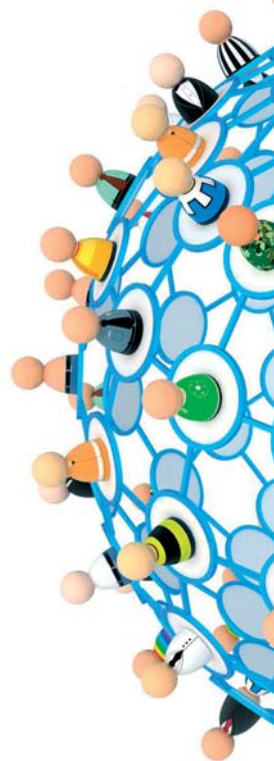
Notes

[1] Large Triple Stores. Wiki page supported by W3C. http://esw.w3.org/LargeTripleStores, considering repeatable, "full-cycle" experiments.

[2] BSBM V3 Results (February 2011). Bizer, Ch., Schultz, A. http://www4.wiwiss.fu-berlin.de/bizer/BerlinSPARQLBenchmark/results/V6/index.html.
     BigOWLIM demonstrated by far the best loading performance and also the best query performance among those repositories that can handle
     updates and multi-client query tasks

[3] "In our tests, BigOWLIM provides the best average query response time and answers maximum number of queries for both the
     datasets. ... it is clear to see that execution speed-wise BigOWLIM outperforms Allegrograph and Sesame for almost all of the dataset
     queries." - Thakker , D., Osman, T., Gohil, S., Lakin, P. (Press Association and the Nottingham Trent University).
     A Pragmatic Approach to Semantic Repositories Benchmarking. In Proc. of the 7th Extended Semantic Web Conference, ESWC 2010

[4] Stoilos G., Grau B. C., Horrocks I. "How Incomplete is your Semantic Web Reasoner?" In Proc. of AAAI 10, 2010.

[5] Implementations – OWL. http://www.w3.org/2007/OWL/wiki/Implementations

http://www.ontotext.com/owlim