## Reasoning in Semantic Wikis

Markus Krötzsch<sup>1</sup>

Sebastian Schaffert<sup>2</sup>

Denny Vrandečić<sup>1</sup>

<sup>1</sup> Institut AIFB, Universität Karlsruhe, Germany

<sup>2</sup> Salzburg Research, Austria

Published under the terms of CC-BY-NC 3.0 ( $\rightarrow$  details)



M. Krötzsch (AIFB Karlsruhe)

[Berners-Lee, Hendler, Lassila, 2001]:

"A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities"

"Computers will find the meaning of semantic data by following hyperlinks to definitions of key terms and rules for reasoning about them logically."

"Naturally, you want to check this, so your computer asks the service for a proof of its answer, which it promptly provides by translating its internal reasoning into the Semantic Web's unifying language."



Wikipedia classifies "Semantic Web" into the category "Buzzwords" (together with Web 2.0)

Wikipedia classifies "Semantic Web" into the category "Buzzwords" (together with Web 2.0)

"The Semantic Web is a web of data" [W3C, 2007]

AIFB(

Wikipedia classifies "Semantic Web" into the category "Buzzwords" (together with Web 2.0)

"The Semantic Web is a web of data" [W3C, 2007]

"Spoken cynically, [...] the Web has received toy engines that neither meet its requirements nor scale to its size." [Fensel, van Harmelen, 2007]

#### "The Contradiction of Web and Reasoning"



### Semantics in Wikis



- 3 Adding Expressivity
- Query Languages

### 5 Performance

6 Closing Remarks



## Outline

### Semantics in Wikis

- 2) Semantic MediaWiki and IkeWiki
- 3 Adding Expressivity
- Query Languages
- 5 Performance
- 6 Closing Remarks



# From Websites to Wikis

#### Wikis

- Content management systems for text (usually)
- Web-based (usually)
- Collaborative (usually)
- Quick and simple (well, compared to XHTML)

# From Websites to Wikis

#### Wikis

- Content management systems for text (usually)
- Web-based (usually)
- Collaborative (usually)
- Quick and simple (well, compared to XHTML)
- 1994: Portland Pattern Repository ("Ward's Wiki")
- ...
- Jan 15, 2001: Wikipedia is launched ("to add a little feature to Nupedia")

AIFB

# From Websites to Wikis

### Wikis

- Content management systems for text (usually)
- Web-based (usually)
- Collaborative (usually)
- Quick and simple (well, compared to XHTML)
- 1994: Portland Pattern Repository ("Ward's Wiki")

#### • ...

- Jan 15, 2001: Wikipedia is launched ("to add a little feature to Nupedia")
- Feb 12, 2001: 1,000 articles
- Sep 7, 2001: 10,000 articles
- Jan 22, 2003: 100,000 articles
- Mar 1, 2006: 1,000,000 articles
- Today: 1,987,473 articles, 10,000s of sites, 100s of engines



As wikis grow big, retrieving content becomes harder.

### Challenges

- Where can I find the answer to my question?
- What is the answer to my question?
- Which interesting information can the wiki offer to me?

→ Text-based search mainly solves content location.
 → Aggregating knowledge is a cumbersome manual task.
 → More "structure" needed



MediaWiki already has ways of structuring content:

- Links: browsing and searching
- Categories: classifying content pages for browsing
- Redirects: treat synonyms efficiently (no unique names ...)
- Some more (namespaces, templates, disambiguation, ...)

MediaWiki already has ways of structuring content:

- Links: browsing and searching
- Categories: classifying content pages for browsing
- Redirects: treat synonyms efficiently (no unique names ...)
- Some more (namespaces, templates, disambiguation, ...)

Main structuring mechanism: pages!

AIFB

# Many Approaches ...

A number of semantic wikis appeared:

- KendraBase
- Rhizome
- Platypus ٠
- Semantic MediaWiki (SMW)
- WikSAR
- KaukoluWiki
- IkeWiki
- MaknaWiki
- OntoWiki

Design choices: Text-annotation, ontology creation, or data collection? Closed domain or open community? SemWeb experts or not? Small or large scale?

### Semantics in Wikis



- 3 Adding Expressivity
- Query Languages
- 5 Performance
- 6 Closing Remarks



#### Semantic MediaWiki

- Extension module for MediaWiki (see Wikipedia)
- Free (GPL)
- Developed since 2005, 2–5 developers
- Current version: SMW 1.0 (release pending)
- Used

Documentation and download: http://ontoworld.org/wiki/SMW Further Reading: [Kr. et al., 2007]

 $\rightarrow$  try it out SMW on ontoworld.org



#### SMW adds properties:

- Associate pages with "values"
- Introduce new properties as needed
- Syntax: [[Property::value]]

Different kinds of values: numbers, other pages, strings, dates, locations, ...

**AIFB**(

#### SMW adds properties:

- Associate pages with "values"
- Introduce new properties as needed
- Syntax: [[Property::value]]

Different kinds of values: numbers, other pages, strings, dates, locations, ...

Properties can have different types:

- Properties are associated with types
- Types are mostly built-in, but possibly customisable
- Syntax: [[has type::Datatype]]

Why should anyone annotate pages?

Why should anyone annotate pages?

- "Which interesting information can the wiki offer to me?"
   → Browsing (→ online example)
  - Use semantics for finding "related" information
  - Simple interfaces: just clicking around
- "What is the answer to my question?"
  - $\rightsquigarrow$  Querying ( $\rightarrow$  online example)
    - Simple wiki-based query language
    - Inline queries: dynamically embedd results into pages



## Formal Semantics: Mapping to OWL

#### Formal grounding of SMW annotations: OWL DL mapping

SMW	DL	OWL
Simple Pages	Individual names	OWL-Individuals
Category pages	Class names	OWL-Classes
Property pages	Roles	OWL-Properties
Type pages	Individual names	OWL-Individuals

ObjectProperty, DatatypeProperty, or AnnotationProperty? Depends on type of SMW-property!

Simple annotations of a page "p":

- [[Category:C]]  $\mapsto p: C$  (or C(p))
- [R::v]  $\mapsto (p, v) : R \quad (or R(p, v))$

Also available as OWL DL export (OWL/RDF format)  $\rightarrow$  online example

M. Krötzsch (AIFB Karlsruhe)



## IkeWiki

#### IkeWiki

- Stand-alone wiki implementation
- Free (GPL)
- Developed since 2005, 1 developer (more to come)
- Current version: IkeWiki 1.99.39beta
- Used in projects

#### Screencast and download:

http://ikewiki.salzburgresearch.at/

Different focus than SMW:

- URIs used as identifiers in the wiki
- Free-form RDF editing possible
- Some differences in features and interfaces



### Semantics in Wikis



- 3 Adding Expressivity
- Query Languages
- 5 Performance
- 6 Closing Remarks



## Making Semantic Wikis more Expressive

So far only basic ABox statements...

So far only basic ABox statements...

Issues that must be taken into account:

- Expressivity vs. Performance: Web applications are performance critical
- Usability/debuggability: users must understand logical features
- Robustness: the system should be robust to vandalism and errors
- **Standardisation:** semantics well-documented, interchange possible, simplifies implementation



*"The problem is that the word 'cook' has no meaning, or semantic content, to the computer."* [Berners-Lee, Hendler, Lassila, 2001]



*"The problem is that the word 'cook' has no meaning, or semantic content, to the computer."* [Berners-Lee, Hendler, Lassila, 2001]

→ Let's consider a wiki of recipes:

- Pages about recipes and ingredients
- Recipes with textual descriptions and list of ingredients
- E.g., the page "Mango pickles" states that this recipe contains 300g of green mango.

How can we represent this in terms of binary properties?



*"The problem is that the word 'cook' has no meaning, or semantic content, to the computer."* [Berners-Lee, Hendler, Lassila, 2001]

→ Let's consider a wiki of recipes:

- Pages about recipes and ingredients
- Recipes with textual descriptions and list of ingredients
- E.g., the page "Mango pickles" states that this recipe contains 300g of green mango.

How can we represent this in terms of binary properties?

• (pickles, 300g\_gr\_mango) : contains

*"The problem is that the word 'cook' has no meaning, or semantic content, to the computer."* [Berners-Lee, Hendler, Lassila, 2001]

→ Let's consider a wiki of recipes:

- Pages about recipes and ingredients
- Recipes with textual descriptions and list of ingredients
- E.g., the page "Mango pickles" states that this recipe contains 300g of green mango.

How can we represent this in terms of binary properties?

- (pickles, 300g\_gr\_mango) : contains
- (pickles, gr\_mango) : contains (pickles, 300g) : amount



*"The problem is that the word 'cook' has no meaning, or semantic content, to the computer."* [Berners-Lee, Hendler, Lassila, 2001]

→ Let's consider a wiki of recipes:

- Pages about recipes and ingredients
- Recipes with textual descriptions and list of ingredients
- E.g., the page "Mango pickles" states that this recipe contains 300g of green mango.

How can we represent this in terms of binary properties?

- (pickles, 300g\_gr\_mango) : contains
- (pickles, gr\_mango) : contains (pickles, 300g) : amount (pickles, cayenne) : contains (pickles, 2Tsp) : amount



"The problem is that the word 'cook' has no meaning, or semantic content, to the computer." [Berners-Lee, Hendler, Lassila, 2001]

→ Let's consider a wiki of recipes:

- Pages about recipes and ingredients
- Recipes with textual descriptions and list of ingredients
- E.g., the page "Mango pickles" states that this recipe contains 300g of green mango.

How can we represent this in terms of binary properties?

- (pickles, 300g\_gr\_mango) : contains
- (pickles, gr\_mango) : contains (pickles, 300g) : amount (pickles, cayenne) : contains (pickles, 2Tsp) : amount
- (pickles, gr mango) : contains

(gr mango, 300g) : amount



"The problem is that the word 'cook' has no meaning, or semantic content, to the computer." [Berners-Lee, Hendler, Lassila, 2001]

→ Let's consider a wiki of recipes:

- Pages about recipes and ingredients
- Recipes with textual descriptions and list of ingredients
- E.g., the page "Mango pickles" states that this recipe contains 300g of green mango.

(pickles, 300g) : amount

(gr mango, 300g) : amount

How can we represent this in terms of binary properties?

- (pickles, 300g\_gr\_mango) : contains
- (pickles, gr\_mango) : contains (pickles, cayenne) : contains (pickles, 2Tsp) : amount
- (pickles, gr mango) : contains
- (pickles, ing01) : contains (ing01, gr\_mango) : ingredient (ing01, 300g) : amount

~> Semantic wikis should make this easy.



### Does Your Mum Really Need Logic?

M. Krötzsch (AIFB Karlsruhe)

Semantic Wiki Reasoning

Reasoning Web 2007 19 / 39

## Does Your Mum Really Need Logic?

"What can I cook with 3kg of green mango?"

## Does Your Mum Really Need Logic?

What can I cook with 3kg of green mango?"
 → data value ranges, unit support

# Does Your Mum Really Need Logic?

- What can I cook with 3kg of green mango?"
   → data value ranges, unit support
- Which recipes contain mango?"

- What can I cook with 3kg of green mango?"
   → data value ranges, unit support
- Which recipes contain mango?"
   → generalisation to "mango"
  - taxonomic approach: ing01 : GrMango, GrMango ⊑ Mango
  - role composition:

AIFB(

- What can I cook with 3kg of green mango?"
   → data value ranges, unit support
- Which recipes contain mango?"
   → generalisation to "mango"
  - taxonomic approach: ing01 : GrMango, GrMango ⊑ Mango
  - role composition:

Which recipes do not contain meat?"

- What can I cook with 3kg of green mango?"
   → data value ranges, unit support
- Which recipes contain mango?"
   → generalisation to "mango"
  - taxonomic approach: ing01 : GrMango, GrMango  $\sqsubseteq$  Mango
  - or role composition:

Which recipes do not contain meat?"
 → closed world reasoning, role composition

- What can I cook with 3kg of green mango?"
   → data value ranges, unit support
- Which recipes contain mango?"
   → generalisation to "mango"
  - taxonomic approach: ing01 : GrMango, GrMango ⊑ Mango
  - role composition:

- Which recipes do not contain meat?"
   → closed world reasoning, role composition
- Which recipes are unsuitable for mango allergics?"

- What can I cook with 3kg of green mango?"
   → data value ranges, unit support
- "Which recipes contain mango?"
   ~> generalisation to "mango"
  - taxonomic approach: ing01 : GrMango, GrMango ⊑ Mango
  - role composition:

- Which recipes do not contain meat?"
   → closed world reasoning, role composition
- "Which recipes are unsuitable for mango allergics?"

   → background ontology (What is a mango allergic? When can
  mango ingredients be substituted?)



## Semantics in Wikis

- 2 Semantic MediaWiki and IkeWiki
- 3 Adding Expressivity
- Query Languages
- 5 Performance
- Closing Remarks



SMW has configurable support for

- Class hierarchies
- Property hierarchies
- Datatypes
- N-ary data values
- Equality reasoning

# The SMW Query Language

Expressive features:

- Category (class) and property statements
- Conjunction and disjunction
- Data ranges
- Nesting

No variables, universal statements, negation, number restrictions.  $\rightsquigarrow$  Queries can be mapped to DL:

```
City \sqcap (∃population.ge(500000) \sqcup
∃located_in.(Country \sqcap ∃member_of.{EU}))
```



 $\rightsquigarrow$  SMW queries use only  $\sqcap$ ,  $\sqcup$ ,  $\exists$ ,  $\{a\}$  (nominals),  $\top$ , and data values

Instance retrieval possible in polynomial time

#### Idea

materialise answers inside out, computing result sets for subqueries first



# Complexity of SMW queries (2)

Can we generalise the above result?

# Complexity of SMW queries (2)

Can we generalise the above result?

## The DL $\mathcal{EL}++$

**Tractable** description logic, featuring  $\sqcap$ ,  $\exists$ ,  $\{a\}$  (nominals), role composition  $R \circ S \sqsubseteq T$ , data values (concrete domains, if "convex")



# Complexity of SMW queries (2)

Can we generalise the above result?

### The DL $\mathcal{EL}++$

**Tractable** description logic, featuring  $\sqcap$ ,  $\exists$ ,  $\{a\}$  (nominals), role composition  $R \circ S \sqsubseteq T$ , data values (concrete domains, if "convex")

→ but SMW also has  $\sqcup$  and non-convex concrete domains → generally ExpTime-complete! (even for logic  $\mathcal{FEL}$ :  $\sqcap$ ,  $\sqcup$ ,  $\exists$ )

AIFB

Can we generalise the above result?

## The DL $\mathcal{EL}++$

**Tractable** description logic, featuring  $\sqcap$ ,  $\exists$ ,  $\{a\}$  (nominals), role composition  $R \circ S \sqsubseteq T$ , data values (concrete domains, if "convex")

→ but SMW also has  $\Box$  and non-convex concrete domains → generally ExpTIME-complete! (even for logic  $\mathcal{FEL}$ :  $\Box$ ,  $\Box$ ,  $\exists$ )

### Observation

Disjunctions in SMW do not add real complexity:  $(A \sqcup B)? \mapsto A? \cup B?$ 

→ general approach: Horn Description Logics
 → source for theoretical results ...

AIFB

IkeWiki has configurable support for

- class and property hierarchies
- domain and range restrictions
- OWL DL inferencing if installed
- experimental Prolog binding (unstable)

SPARQL as a query language



- "SPARQL" → "SPARQL Protocol And RDF Query Language"
- Almost a W3C standard (hopefully)
- Based on graph pattern matching + filters

- "SPARQL" → "SPARQL Protocol And RDF Query Language"
- Almost a W3C standard (hopefully)
- Based on graph pattern matching + filters

AIFB

# Semantics and Complexity of SPARQL

#### Operational semantics based on SPARQL algebra:

AIFB

# Semantics and Complexity of SPARQL

### Operational semantics based on SPARQL algebra:

SPARQL is hard:

- Basic graph pattern matching is already NP-complete
- Matching SPARQL graph patterns is PSPACE-complete
- Filters are also tricky
- However: data complexity is LOGSPACE
- Note: no schema, just RDF with simple entailment

M. Krötzsch (AIFB Karlsruhe)

AIFB

Operational SPARQL algebra vs. model theoretic OWL semantics ~> What does SPARQL for OWL mean?

## Conjunctive Queries (CQs) for DLs

Conjunctive queries similar to Datalog, e.g.

 $Person(x) \land livesIn(x, y) \land bornIn(x, y)$ 

~ more expressive than DL querying

Remark: still not quite the same as SPARQL!



Conjunctive queries stricly harder than DL reasoning alone:

- 2-EXPTIME for *SHOQ* and *SHIQ* [Glimm, Horroks, Lutz, Sattler, 2007; Lutz, 2007]
- PSPACE for (restricted)  $\mathcal{EL}++$  [Kr., Rudolph, Hitzler, 2007]
- NP if TBox is empty (plain RDF)
- Complexity/decidability for  $\mathcal{SHOIN}$  (OWL DL) unknown!
- → Difficult in a Web application . . .

## Semantics in Wikis

- 2 Semantic MediaWiki and IkeWiki
- 3 Adding Expressivity
- Query Languages
- 5 Performance
- 6 Closing Remarks



## **Practical Performance**

#### Complexity $\neq$ Performance:

- "Hard tasks" might be feasible in practice
- "Easy tasks" might still perform poorly

#### Example

KB-complexity (and data complexity) for CQs on  $\mathcal{EL}++$  is *P*.



## Complexity $\neq$ Performance:

- "Hard tasks" might be feasible in practice
- "Easy tasks" might still perform poorly

#### Example

KB-complexity (and data complexity) for CQs on  $\mathcal{EL}++$  is *P*. Bounded by some multiple of  $k^{q^q}$  (*k* size of *KB*, *q* size of query)!

 $\rightsquigarrow$  "Polynomial" is often still too bad for practice

On a real site

- SMW takes less than 9% of overall processing time, but
- 40% 50% of this time is query answering.

(http://ontoworld.org/profileinfo.php)

#### SMW:

- built-in semantic store
- based on MySQL
- binding via storage abstraction layer

#### IkeWiki:

- Jena store as backend
- Support for RDFS and OWL inherited from Jena
- Binding via DIG interface

→ Clean interfaces enable easy extension. Try it.



Basic evaluations and experiences:

- IkeWiki for OWL: up to around 10,000 statements
- IkeWiki for RDFS: up to around 50,000 statements
- SMW with simple KB: tested with about 13,800,000 statements

#### **Disclaimer:**

Performance greatly depends on schema/queries! Even a hundred statements are too much if you can choose the query!

 $\rightsquigarrow$  Sensible restrictions on queries needed



## **Reasoners Wanted!**

#### A wiki-reasoner must ...

- ... tolerate large datatsets, even if reasoning is too slow (Jena in-memory crashes quite instantly with 1.3Mio triples),
- ... support basic datatype operations,
- ... support basic retrieval options (order by, limit, offset),
- ... support incremental updates,
- ... not impose non-local syntactical constraints on axioms,
- ... crash only when truly necessary

## **Reasoners Wanted!**

## A wiki-reasoner must ...

- ... tolerate large datatsets, even if reasoning is too slow (Jena in-memory crashes quite instantly with 1.3Mio triples),
- ... support basic datatype operations,
- ... support basic retrieval options (order by, limit, offset),
- ... support incremental updates,
- ... not impose non-local syntactical constraints on axioms,
- ... crash only when truly necessary

### A wiki-reasoner should ...

- ... support some syntactical KB inspection (string-matching?),
- ... provably implement a well-documented semantics,
- ... have some approach towards non-unique names,
- ... be free software (no, not as in beer)

## Semantics in Wikis

- 2 Semantic MediaWiki and IkeWiki
- 3 Adding Expressivity
- Query Languages
- 5 Performance
- 6 Closing Remarks



## Conclusion

In 2007, semantic wikis bear great challenges and opportunities for reasoning reasearch.

- SMW and IkeWiki need and use reasoning
- More expressivity is needed
- Complexity results are only the first step
- Extending SMW/IkeWiki is easy: software is free, we cooperate, and we hire ;-)
- (and, of course, you can also just use them)



"Knowledge representation, as this technology is often called, is currently in a state comparable to that of hypertext before the advent of the Web:"

"Knowledge representation, as this technology is often called, is currently in a state comparable to that of hypertext before the advent of the Web:"

"it is clearly a good idea,



"Knowledge representation, as this technology is often called, is currently in a state comparable to that of hypertext before the advent of the Web:"

# *"it is clearly a good idea, and some very nice demonstrations exist,*



"Knowledge representation, as this technology is often called, is currently in a state comparable to that of hypertext before the advent of the Web:"

# *"it is clearly a good idea, and some very nice demonstrations exist, but it has not yet changed the world."*



# Further reading I

- Tim Berners-Lee, James Hendler, and Ora Lassila. The Semantic Web Scientific American, May 2001. → Available online.
- Birte Glimm, Ian Horrocks, Carsten Lutz, and Uli Sattler. Conjunctive Query Answering for the Description Logic SHIQ Proc. IJCAI 2007 (→ Download)
- Markus Krötzsch, Sebastian Rudolph, and Pascal Hitzler. Conjunctive Queries for a Tractable Fragment of OWL 1.1 Proc. of ISWC'2007 (→ Download)

Markus Krötzsch, Denny Vrandečić, Heiko Haller, Max Völkel, and Rudi Studer. Semantic Wikipedia

Journal of Web Semantics, 2007, to appear. ( $\rightarrow$  Download)



# Further reading II



#### Carsten Lutz Inverse Roles Make Conjunctive Queries Hard DL-Workshop 2007 (→ Download)

Semantic MediaWiki Online Dokumentation and Installation Guide Available at http://ontoworld.org/wiki/SMW, 2007.

- Nigel Shadbolt, Wendy Hall, and Tim Berners-Lee. The Semantic Web Revisited IEEE Intelligent Systems, May/June 2006. (→ Download)
  - World Wide web Consortium Semantic Web Online at http://www.w3.org/2001/sw/ (Sept. 2007)

