AIFB

Approximate Reasoning for the Semantic Web
Part V
Approximate Resolution for OWL

Frank van Harmelen
Pascal Hitzler
Holger Wache

ESSLLI 2006 Summer School
Malaga, Spain, August 2006

AIFB

Contents Part V

- **KAON2 – resolution-based reasoning with OWL**
- Approximate reasoning with Screech

AIFB

The KAON2 OWL Reasoner

- Completely new deduction algorithms.
- *Not* tableaux-based.
- Reasoning via reduction of OWL DL to (positive) disjunctive datalog.
- Goal: Efficient ABox reasoning.
- Current performance similar to state-of-the-art tableaux reasoners. Better for some tasks.
- Binaries available from http://kaon2.semanticweb.org
  - Implementation by **Boris Motik**.
  - Theory by B. Motik, U. Hustadt, U. Sattler, R. Studer.
- Treats all of OWL DL except nominals.

AIFB

KAON2: Basic Ideas

- ABox reasoning (instance retrieval) is more important for practice than TBox reasoning.
- Resolution ideal for ABox reasoning (Prolog).
- Similarly good: Deductive Database techniques

- $\rightarrow$ Resolution proofs for OWL DL?
  - Naive approach does not always terminate.
  - Reason: Transformation to FOL yields existential quantifiers which are Skolemised to function symbols.
  - Termination of algorithms not guaranteed in presence of function symbols.

AIFB

KAON2: How to deal with termination issue

- Finitely many usages of existential quantifiers suffice for sound and complete reasoning.
  – How many and which ones?
- First process TBox: Derive (all necessary) logical consequences using ordered resolution.
  – Finite Set!
  – Then generation of further individuals via function symbols no longer necessary!
- Existential quantifiers can then be removed!

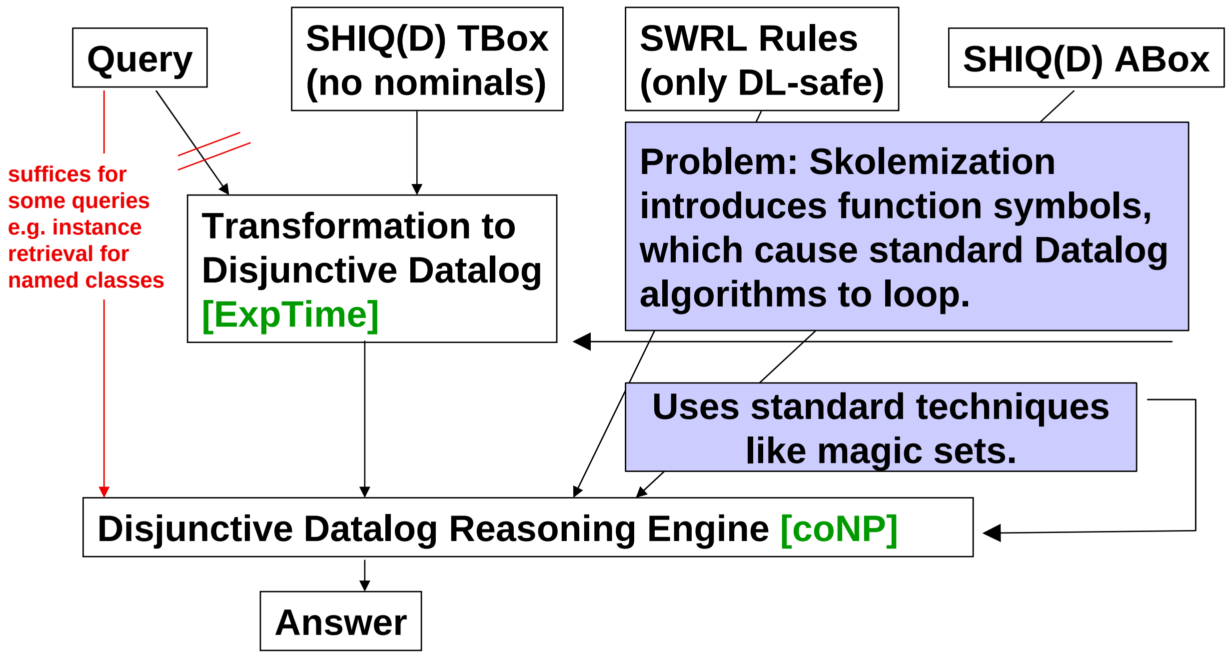
AIFB

KAON2: Inference mechanism

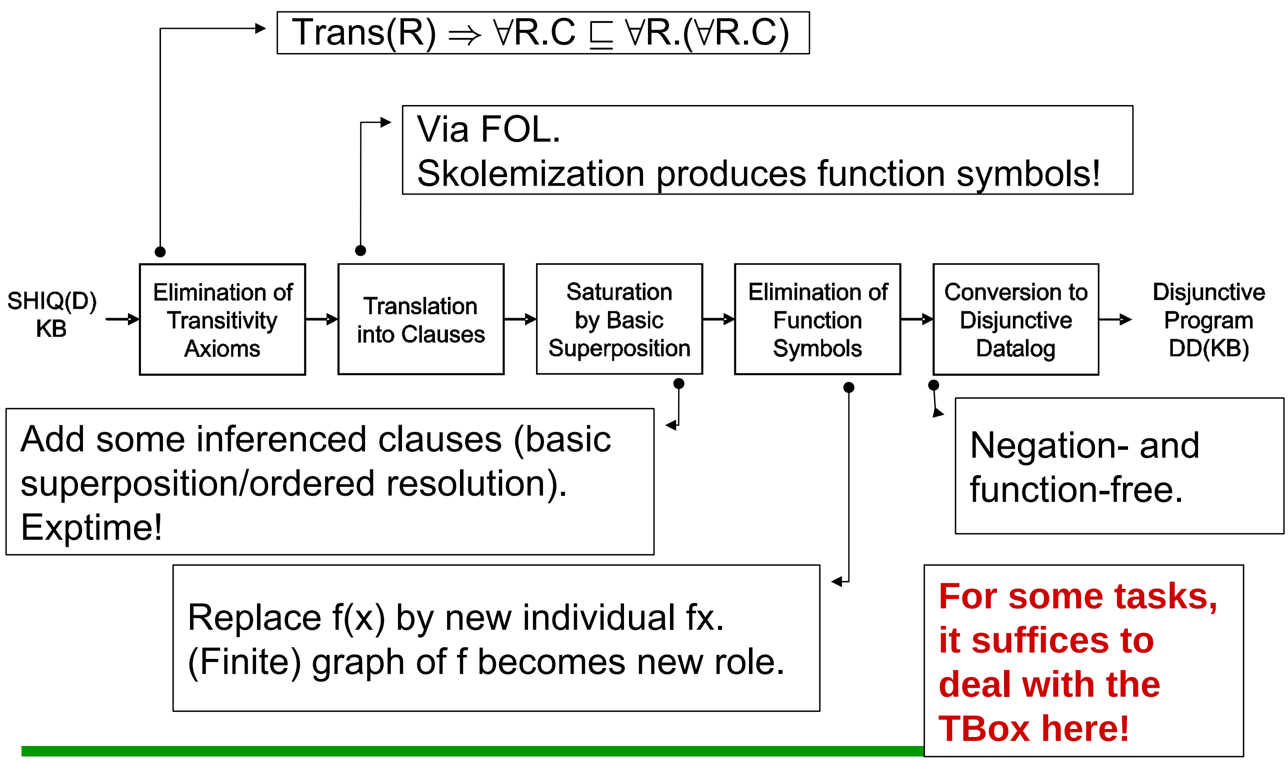
1. Translate TBox to function free clauses.
   (Exptime!)
2. Add ABox.
3. Employ standard reasoning methods for function-free clauses, e.g. magic sets.
   (NP-complete!)

- TBox needs to be processed only once!
- Algorithm is worst-case optimal!
- Data complexity is NP!

KAON2 Reasoner core architecture



KAON2 transformation algorithm



AIFB

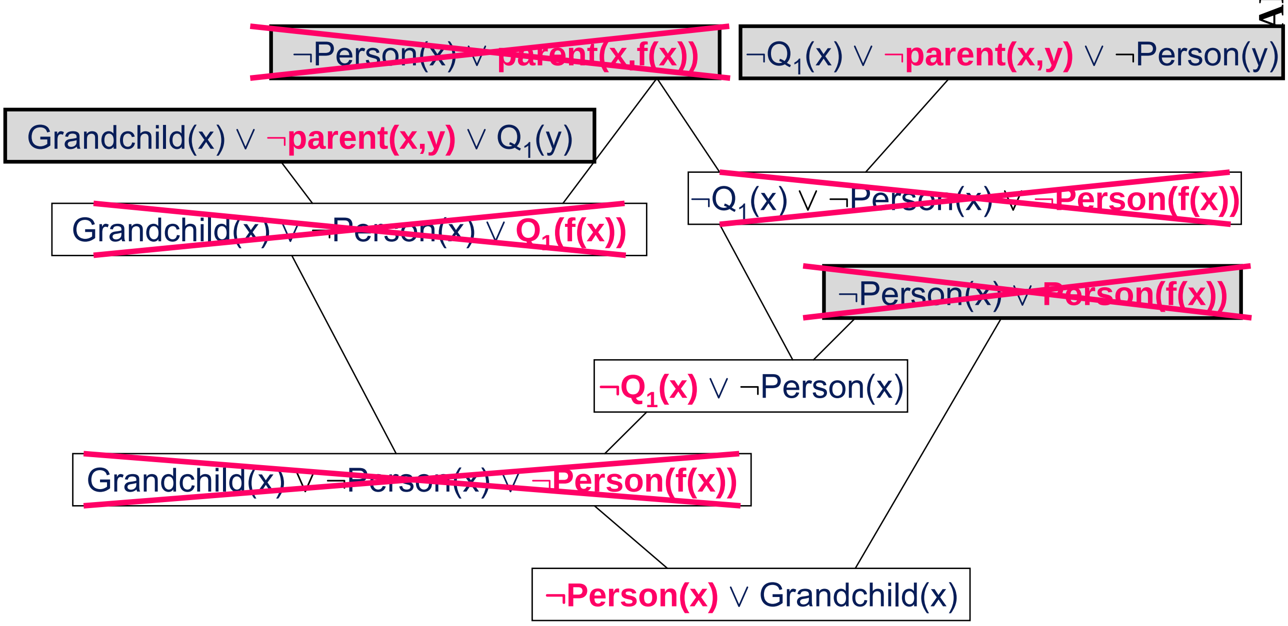
Simple example for transformation (ALC only)

structural transformation & clausification

KB	FOL KB
Person ⊑ ∃ parent.Person	¬Person(x) ∨ parent(x,f(x))
	¬Person(x) ∨ Person(f(x))
∃ parent.(∃ parent.Person) ⊑ Grandchild	Grandchild(x) ∨ ¬parent(x,y) ∨ Q_1(y)
	¬Q_1(x) ∨ ¬parent(x,y) ∨ ¬Person(y)
Person(a)	Person(a)

AIFB

Saturation



Translate to Datalog

AIFB

Result: Disjunctive datalog program

KB
Person ⊑ ∃ parent.Person ∃ parent.(∃ parent.Person) ⊑ Grandchild Person(a)
DD(KB)
Q_1(x), Person(y) ← parent(x,y) ← parent(x,y), Q_1(y), Grandchild(x) ← Q_1(x), Person(x) Grandchild(x) ← Person(x) Person(a)

AIFB

A Theorem (Hustadt, Motik, Sattler 2004)

Let KB be an $\mathcal{ALCHIQ}$(D) knowledge base, defined over a concrete domain D, such that satisfiability of finite conjunctions over D can be decided in deterministic exponential time. Then, the following claims hold:

1. KB is unsatisfiable if and only if DD(KB) is unsatisfiable.
2. KB ⊨ α if and only if DD(KB) ⊨ α , where α is of the form A(a) or R(a, b), and A is an atomic concept.
3. KB ⊨ C(a) for a nonatomic concept C if and only if, for Q a new atomic concept, DD(KB ∪ {C ⊑ Q}) ⊨ Q(a).

Performance evaluation

- Different architectures difficult to compare.
  - caching mechanisms
  - preprocessing steps
  - etc.
- Generally, KAON2 seems to do better on ABox reasoning tasks, in particular if ABox is large and TBox is of medium size.
- Generally, KAON2 appears to be inferior on TBox reasoning tasks.

KAON2 additional features

- Additional features
  - an API for programmatic management of OWL-DL and SWRL and F-Logic ontologies,
  - a stand-alone server providing access to ontologies in a distributed manner,
  - an inference engine for answering queries (including support for SPARQL),
  - a DIG interface, allowing access from tools such as Protégé
  - efficient access to instances via relational databases
- Download: http://kaon2.semanticweb.org/

AIFB

Part V Contents

- KAON2 – resolution-based reasoning with OWL
- **Approximate reasoning with Screech**

van Harmelen, Hitzler, Wache ● ESSLLI 2006 ● Malaga, Spain ● August 2006

AIFB

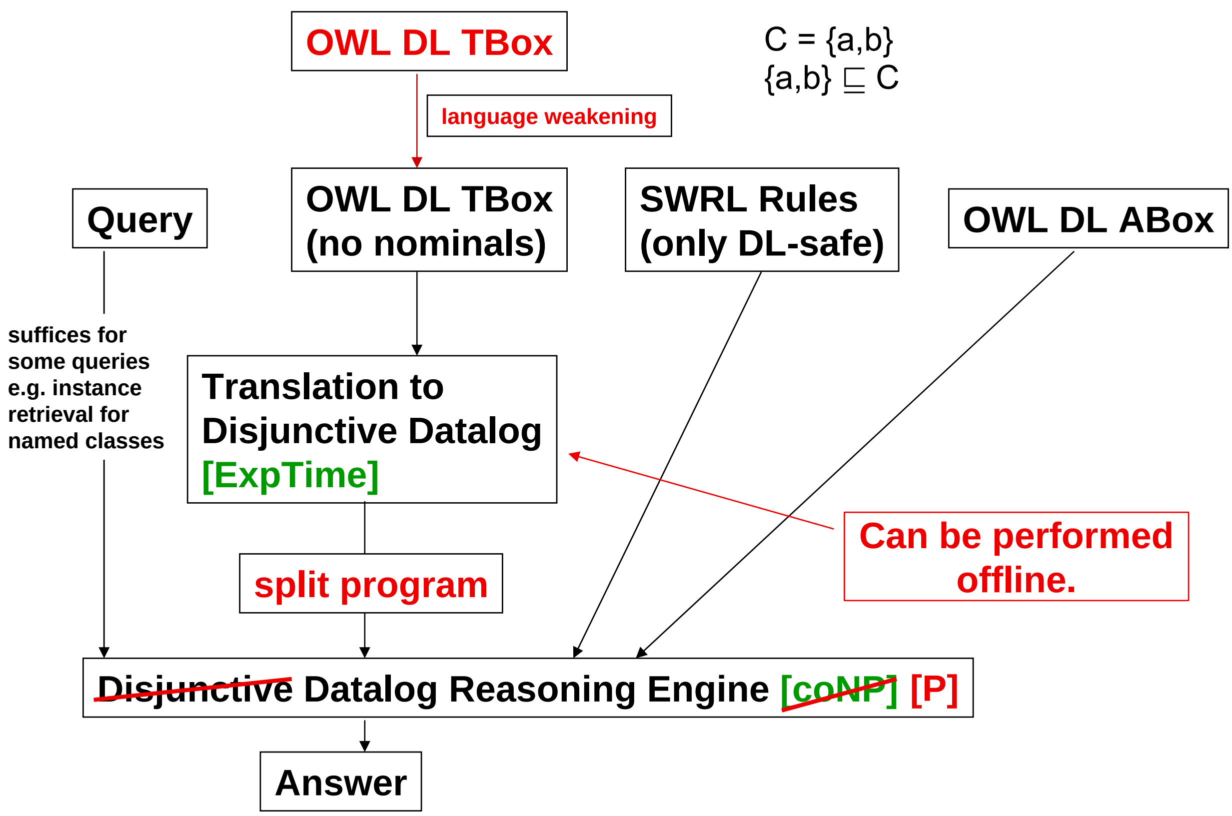
Problem Description

- Reasoning with OWL DL is hard. (Expressivity vs. scalability)
- For certain Semantic Web applications quick responses are more important than absolute accuracy of answering.
  e.g. SmartWeb scenario.
  - Answering of human queries in an open domain.
- We **trade soundess for time**, using **approximate reasoning**.

Approximate Reasoning

- do not confuse with *fuzzy* or *probabilistic* reasoning!
- speed up obtained by
  - modifying the underlying inference relation
  - in a semantically controlled and well-understood way.
- e.g. by decreasing the complexity class of a reasoning task
- e.g. by utilizing intimate knowledge of the bottlenecks in a reasoning algorithm.

Approximate reasoning with Screech for large ABoxes



Screech simple example

serbian $\sqcup$ croatian $\sqsubseteq$ european
eucitizen $\sqsubseteq$ european
german $\sqcup$ french $\sqcup$ beneluxian $\sqsubseteq$ eucitizen
beneluxian $\equiv$ luxembourgian $\sqcup$ dutch $\sqcup$ belgian

serbian(ljiljana).	serbian(nenad).	german(pascal).
french(julien).	croatian(boris).	german(markus).
german(stephan).	croatian(denny).	indian(sudhir).
belgian(saartje).	german(rudi).	german(york).

Screech simple example

beneluxian $\equiv$ luxembourgian $\sqcup$ dutch $\sqcup$ belgian
translates into the following four clauses:
~~luxembourgian(x) $\vee$ dutch(x) $\vee$ belgian(x) $\leftarrow$ beneluxian(x)~~
beneluxian(x) $\leftarrow$ luxemburgian(x)
beneluxian(x) $\leftarrow$ dutch(x)
beneluxian(x) $\leftarrow$ belgian(x)

split of first clause:
luxembourgian(x) $\leftarrow$ beneluxian(x)
dutch(x) $\leftarrow$ beneluxian(x)
belgian(x) $\leftarrow$ beneluxian(x)

$\vdash$ **luxembourgian(saartje)**
$\vdash$ **dutch(saartje)**
$\vdash$ **belgian(saartje)**

AIFB

Screech reasoning

- data complexity is **P**
- complete
- but unsound

AIFB

Screech Performance (not optimized yet)

Galen ontology
673 axioms, 175 classes
randomly populated with 500 individuals

267 disjunctions in 133 rules eliminated

Complete run:
- queried for the extensions of all 175 Galen classes
- resulting in 5809 classifications (Screech)
  - 5353 (i.e. **92.2%**) **correct**
- For 138 out of 175 classes: computed extension correct
- Average **time saved: 39.0%**

AIFB

Screech Performance example run

Time (DD)	Time (SPLIT)	Instances	Class Name
11036 ms	6489 ms	154/154	Biological_object
11026 ms	5959 ms	9/9	Specified_set
11006 ms	6219 ms	9/13	Multiple
11015 ms	5898 ms	16/16	Probe_structural_part_of_heart
11036 ms	7711 ms	4/4	Human_red_blood_cell_mature
11055 ms	5949 ms	24/58	Biological_object_that...

Table 2. Performance comparison for instance retrieval using disjunctive datalog (DD) vs. the corresponding split program (SPLIT), on the KAON2 datalog engine. *Instances* indicates the number of instances retrieved using DD versus SPLIT, e.g. class *Multiple* contained 9 individuals, while the split program allowed to retrieve 13 (i.e. the 9 correct individuals plus 4 incorrect ones). The full name of the class in the last row is Biological_object_that_has_left_right_symmetry.

AIFB

Screech conclusions

- Screech reasoner based on KAON2
- complete but unsound
- 40% speed-up with only 8% wrong answers

- future work
  - optimise further
  - tackle other parts of the KAON2 algorithms
  - combine with other approximate reasoning and optimization techniques

AIFB

One idea for improving Screech

- Use statistical knowledge about distribution of individuals among classes.
- E.g. there are few luxemburgians compared to dutch and belgians.

luxembourgian(x) $\vee$ dutch(x) $\vee$ belgian(x) $\leftarrow$ beneluxian(x)

split :

dutch(x) $\leftarrow$ beneluxian(x)

belgian(x) $\leftarrow$ beneluxian(x)

AIFB

Wrap-up: what we did

AIFB

Wrap-up: What we did

- Introduction Semantic Web
- OWL/Description Logics
  - Semantics
  - Tableau reasoning
  - Resolution-based reasoning
- General approximate reasoning
  - Cadoli-Schaerf
  - BCP, abstraction, knowledge compilation, Top-k,...
- Tableaux-based approximate reasoning for OWL
  - Cadoli-Schaerf
  - Approximate query answering
- Resolution-based approximate reasoning
  - Screech

AIFB

Wrap-up: Why you should do approximate reasoning

- We need expressive reasoning for the semantic web.
- Sound and complete algorithms are too slow.

- Hence: We need approximate reasoning!

- Lots of literature on approximate reasoning exists and has not yet been applied to the semantic web!

⇒ Go do it!

Discussion

- We think that approximate reasoning will help to make semantic web a reality.

- Objections?
- Questions?
- Remarks?

End of course

- Thanks for your interest.
- Feedback is most welcome!

AIFB

Additional semantic web applications

- **SmartWeb**
  See http://www.smartweb-project.org
- Semantic Mediawiki

AIFB

SmartWeb

- Film from
  http://smartweb.dfki.de/main_inf_de.pl?Bilder/index.html

AIFB

Additional semantic web applications

- SmartWeb
- **Semantic Mediawiki**
  See http://wiki.ontoworld.org/

AIFB

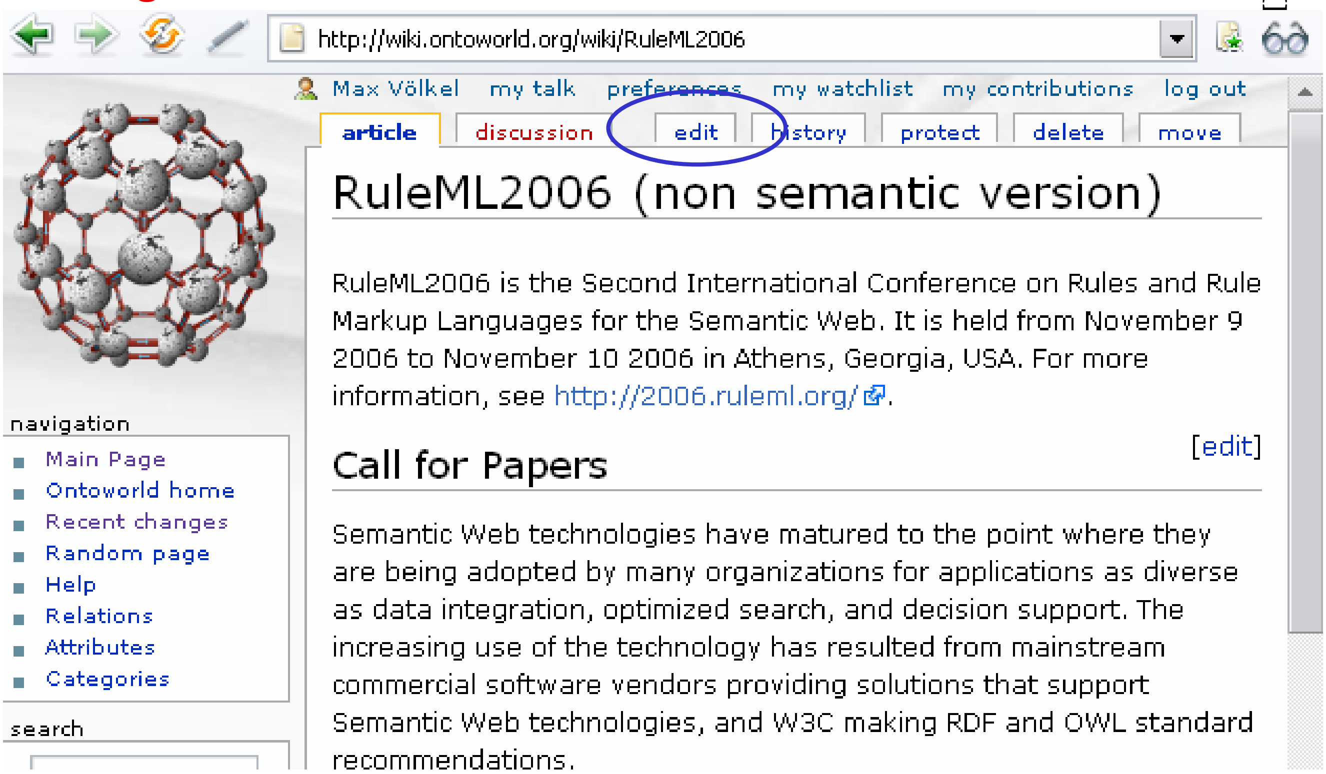
Semantic Mediawiki

- Enhancement of Mediawiki (used in Wikipedia)

- Simple knowledge representation techniques
- Added value for user

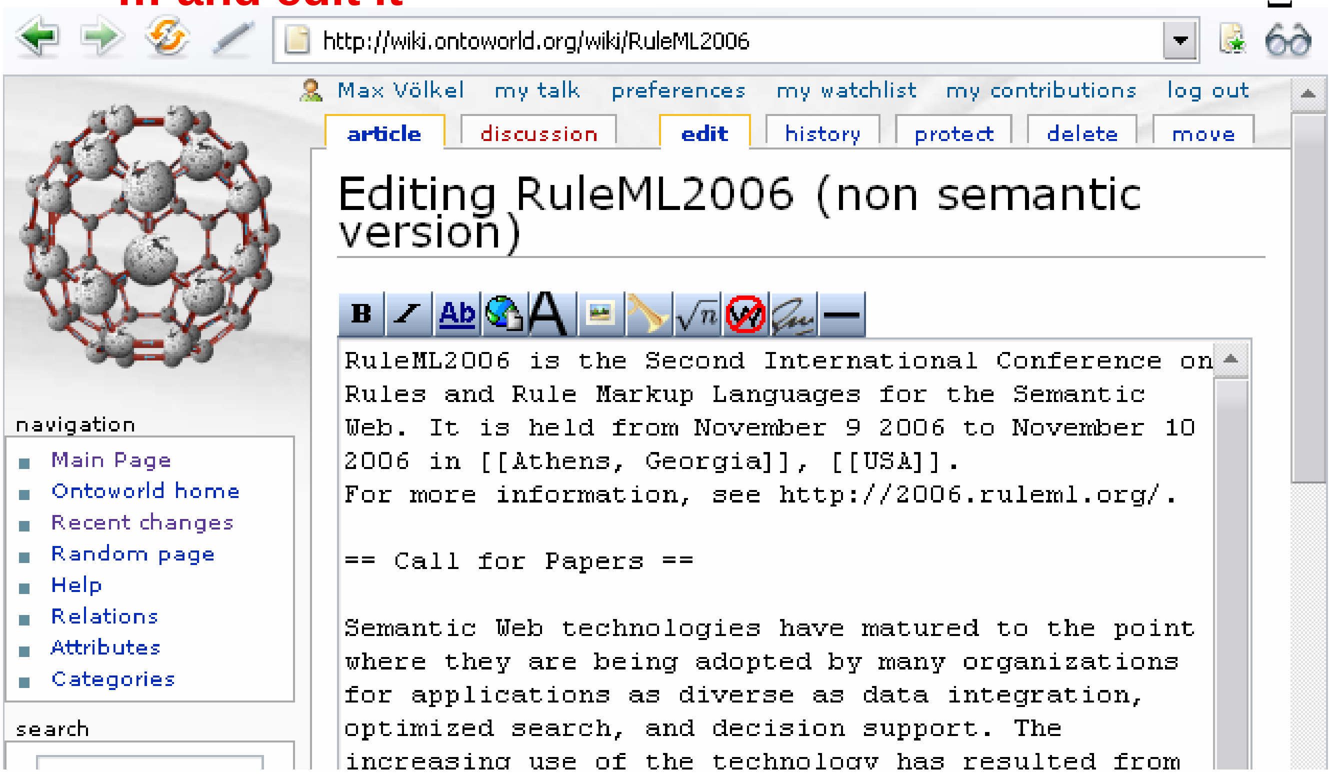
- In particular:
  - enhancement of querying
  - better data reuse

We go to the article on the RuleML2006 conference …



van Harmelen, Hitzler, Wache ● ESSLLI 2006 ● Malaga, Spain ● August 2006

… and edit it



article | discussion | edit | history | protect | delete | move

RuleML2006 is the Second International Conference on Rules and Rule Markup Languages for the Semantic Web. It is held from November 9 2006 to November 10 2006 in [[Athens, Georgia]], [[USA]].
For more information, see http://2006.ruleml.org/.

There is already an ordinariy link to the article of „Athens, Georgia“

article | discussion | edit | history | protect | delete | move

RuleML2006 is the Second International Conference on Rules and Rule Markup Languages for the Semantic Web. It is held from November 9 2006 to November 10 2006 in [[located in::Athens, Georgia]], [[USA]].
For more information, see http://2006.ruleml.org/.

Just say what the relation between this page (RuleML2006) and „Athens, Georgia“ is.

AIFB

From links …

… in [[Athens, Georgia]], [[USA]]. …

… to typed links

… in [[located in::Athens, Georgia]], [[USA]]. …

AIFB

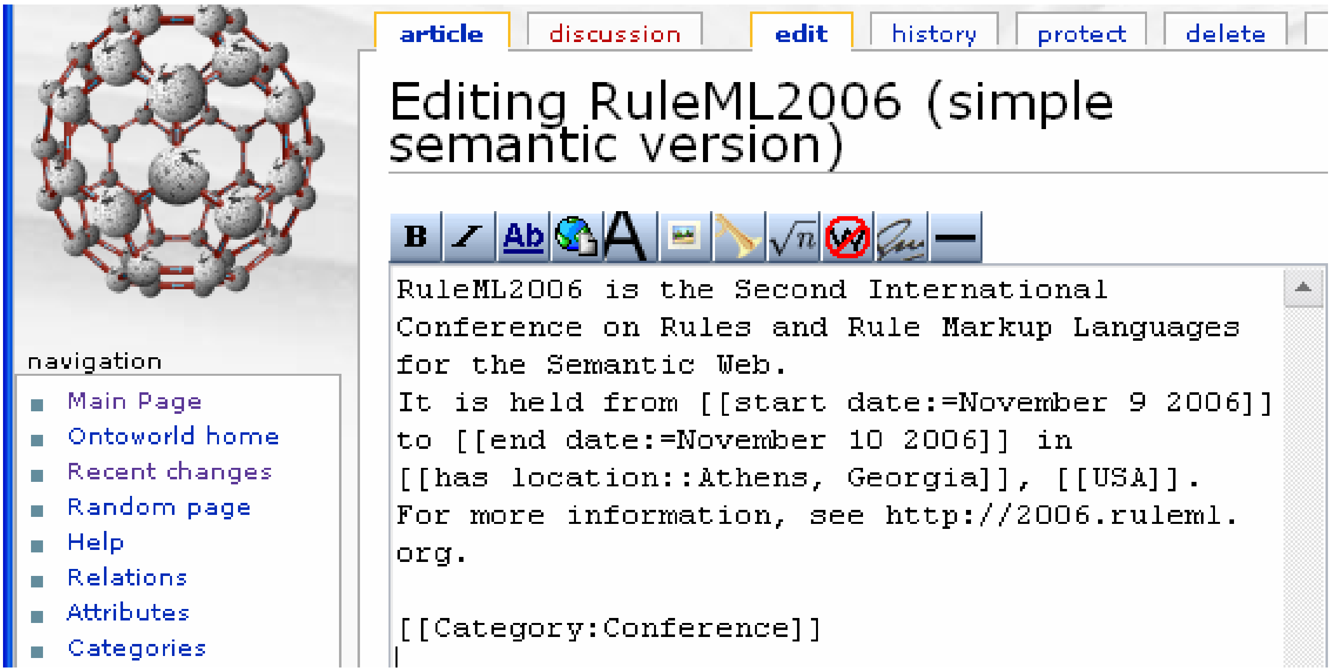
From values …

… It is held from November 9 2006 to November 10 2006 in…

… to attributes

… It is held from [[start date:=November 9 2006]] to [[end date:=November 10 2006]] in…

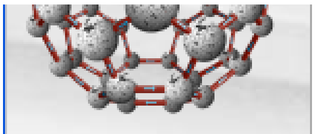
Save.



It looks exactly the same as before

AIFB

What the humans see, when they scroll down



RuleML2006 is the Second International Conference on Rules and Rule Markup Languages for the Semantic Web. It is held from November 9 2006 to November 10 2006 in Athens, Georgia, USA. For more information, see http://2006.ruleml.org.

navigation
- Main Page
- Ontoworld home
- Recent changes
- Random page
- Help
- Relations
- Attributes
- Categories

search

Go Search

toolbox
- What links here
- Related changes
- Special pages
- Printable version
- Permanent link

for demonstration purposes, we have
- RuleML2006 (non semantic verison)
- **RuleML2006 (simple semantic version)**

Relations to other articles — Click + to find similar articles.

RuleML2006 (simple semantic version) Has location Athens, Georgia +

Attributes of RuleML2006 (simple semantic version) — Click + to find similar articles.

Start date: 2006-11-09 +
End date: 2006-11-10 +

Editing help on relations and attributes View as RDF

AIFB

What the humans see, when they scroll down

Relations to other articles — Click + to find similar articles.

RuleML2006 (simple semantic version) Has location Athens, Georgia +

Attributes of RuleML2006 (simple semantic version) — Click + to find similar articles.

Start date: 2006-11-09 +
End date: 2006-11-10 +

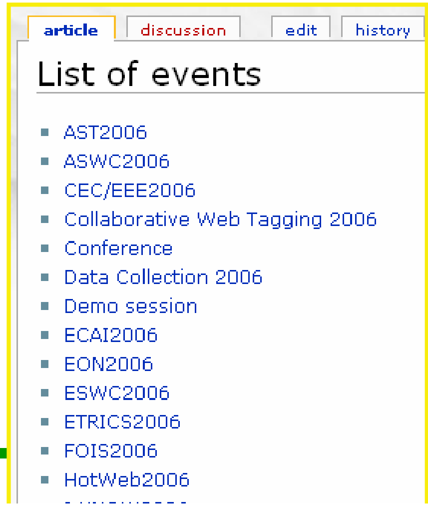
Editing help on relations and attributes View as RDF

AIFB

Benefits for Wikipedians: <ask> for your data

- Inline queries allow for questions like ...
  - ... movies from the 70s starring Sean Connery
  - ... list of events (all conferences and workshops)

```
<ask format="ul" link="all">
     [[Category:Event]]
</ask>
```



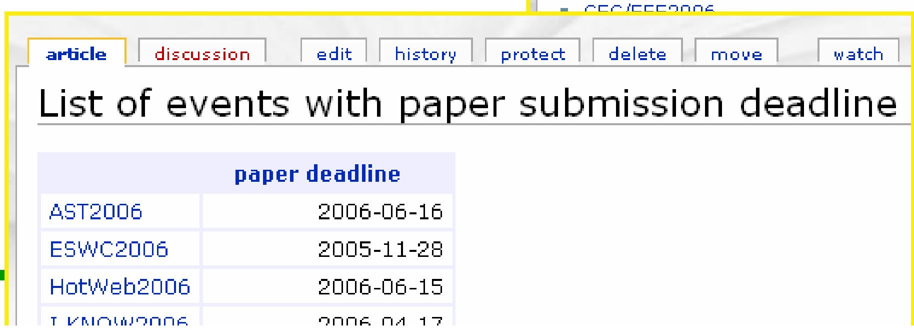
van Harmelen, Hitzler, Wache ● ESSLLI 2006 ● Malaga, Spain ● August 2006

AIFB

Benefits for Wikipedians: <ask> for your data

- Inline queries allow for questions like ...
  - ... movies from the 70s starring Sean Connery
  - ... list of events with their deadline

```
<ask format="ul" link="all">
     [[Category:Event]]
     [[paper deadline:=*]]
</ask>
```



AIFB

Benefits for Wikipedians: <ask> for your data

```
<ask format="ul" link="all">
  [[Category:Event]]
  [[paper deadline:=>June 1 2006]]
  [[paper deadline:=<December 31 2006]]
  [[title:=*]]
  [[paper deadline:=*]]
  [[Category:Topic Semantic Web query
languages]]
</ask>
```

	title	paper deadline
RuleML2006	Second International Conference on Rules and Rule Markup Languages for the Semantic Web	2006-06-05

van Harmelen, Hitzler, Wache ● ESSLLI 2006 ● Malaga, Spain ● August 2006

AIFB

Applications

- Automatic tables and lists
  - E.g. Countries sorted by area, population, alphabet, ...
- **Maintenance** with hand crafted checks
  - Does every country have one capital?
- Integration in applications
  - ```
    latte = wikipedia.get("Latte Macchiatto");
    print latte["contains"]
    ```
- Visualization and browsing
- ... And many unexpected ones
    ```

AIFB

Who is using Semantic MediaWiki?

- Wikicompany ~ 9.000 pages, 227 user accounts
- The Bible Wiki ~ 40.000 pages, 156 user accounts
- WWW2006 Social Wiki ~ 40 pages, 34 user accounts
- Semantic eGovernment Community Portal, ~ 30 pages
- ScubaWiki, ~ 170 pages, 33 user accounts
- NeOn Wiki ~10 pages
- this wiki = Wiki.Ontoworld.org ~400 pages, 158 user accounts
- SourceryForge ~1864 pages, 272 user accounts

Demo and more information at
http://wiki.ontoworld.org/

AIFB

Semantic Mediawiki

- Lightweight use of metadata (no *semantics* in the stronger sense)
- *Already* added value for the user!

- Simple: Introduce background knowledge by means of *ontologies*.
  - Beware of scalability problems!
  - ... and others ...