

Metadata, sémantika a sémantický web

Ing. Vilém Sklenák, CSc.

„There are things we know that we know. There are known unknowns that is to say, there are things that we now know we dont know but there are also unknown unknowns. There are things we do not know we dont know.“

„Existuje známé známé, tedy věci, o nichž víme, že víme. Víme také, že jsou známé neznámé, to jest - víme, že jsou některé věci, které nevíme. Existují ale rovněž neznámé neznámé. Takové, o nichž nevíme, že nevíme.“

Neznámá známá metadata

Pojem je všeobecně známý, ale co je správně?

1. data o datech
2. informace o informacích
3. **význam** nebo sémantika dat
4. **zdroj** poskytující informace o jiném zdroji
5. popisná informace o **webovém zdroji**

A co „ví“ Google? **define:metadata**

Jsou vždy metadata metadaty?

- Data mohou být jednak metadaty, jednak prostými daty.
- Rozhoduje způsob použití.
- Bibliografický záznam:
 1. metadata pro uživatele
 2. data pro knihovní systém

Přednosti metadat

- Umožňují abstrahovat od detailů reprezentace, zachycují informační obsah nezávisle na původní formě dat.
- Umožňují reprezentovat doménové znalosti popisem informační oblasti dle příslušnosti výchozích dat – úsudky, odvozování. . .

Klasifikace metadat

Lze uplatnit různá kritéria. Například TDKIV:

- administrativní m.
- archivační m.
- m. pro právní nároky
- m. pro užití
- popisná m.
- strukturální m.
- technická m.

resp.

- samostatná m.
- vložená m.

Klasifikace metadat

Z hlediska úrovně abstrakce popisu obsahu:

- **Syntaktická m.** – detaily o zdroji (dokumentu), katalogizace, kategorizace.
- **Strukturální m.** – struktura dokumentu, navigace, zobrazení, XML schémata.
- **Sémantická m.** – kontextově relevantní informace, doménově specifické elementy
- **Ontologie** – nejvyšší forma metadat, klíčový princip sémantického webu.



Web se mění

- **Lokalizace zdrojů** – od jednoduchého použití klíčových slov k sémantickým technikám
- **Uživatelé** – zdroje použitelné jak pro člověka, tak pro strojové zpracování
- **Webové služby** – zatím je web spíše v pasivním postavení = místo pro nalezení věcí, ale...

Sémantika

Co říká encyklopedie CoJeCo:

1. oddíl jazykovědy
2. součást sémiotiky zkoumající vztah mezi formou a významem znaku
3. informatika pravidla přisuzující význam přípustným kombinacím symbolů

Co „ví“ Google? [define:semantics](#)

Sémantiky mohou být různé

- Implicitní s.
- Neformální s.
- Formální s.

Implicitní sémantika

```
<zbozi>  
<nazev>Neco</nazev>  
<dodavatel>Firma</dodavatel>  
<cena>190.00</cena>  
</zbozi>
```

- existuje-li konsensus – impl. s. rovnou ve webových aplikacích (např. shopping agenty).
- problém – mnohoznačnosti – „Co je cena?“.

Neformální sémantika

- explicitní sémantika vyjádřená neformálním způsobem
- omezená možnost přímého strojového zpracování
- „obchází“ se přímou implementací do softwaru

Například: <h2>, Dublin Core, ...

Formální sémantika

- explicitní sémantika vyjádřená formálně
- umožňuje přímé zpracování včetně automatické inference
- určující vlastnost sémantického webu = strojově zpracovatelný **obsah**:
 - sémantika jako součást aplikace
 - robustnější je oddělení
 - procedurální nebo deklaratovní přístup



Hey, I know this ontology

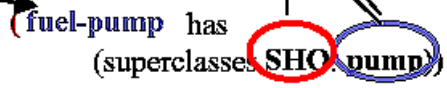
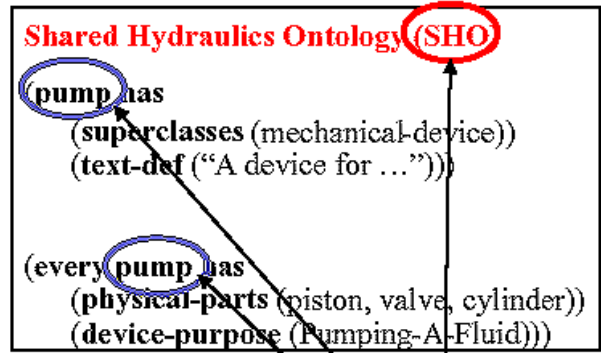
The purpose of this review is to remind operators of the existence of the Operations Manual Bulletin 861, which provides information regarding flight operations with low fuel quantities, and to provide supplementary information regarding main tank boost pump low pressure indications.

797 FUEL DUMP LOW PRESSURE INDICATIONS

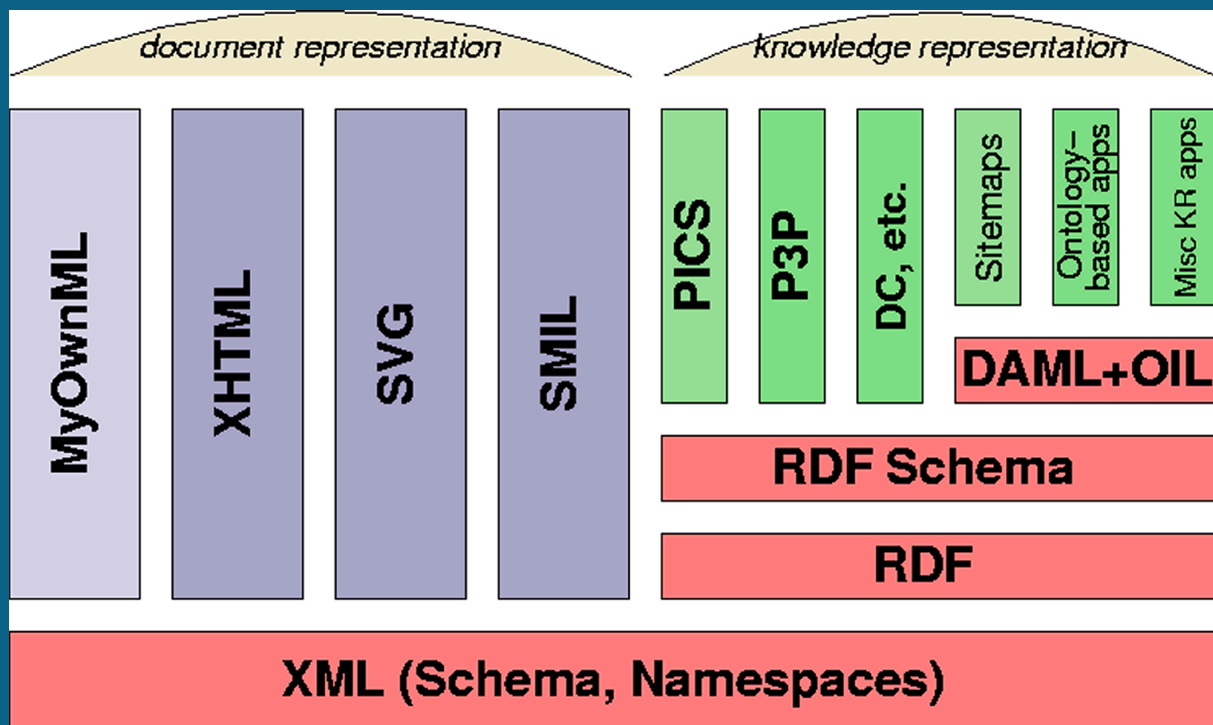
When operating 797 airplanes with low fuel quantities for about

<concept id=fuel-pump>FUEL PUMP</concept>

Semantic Markup
What does it mean?



Sémantický web



„Dolování metadat“

- lexikografická analýza
- vizuální vzory
- využití ontologií

Úvod závěru

- Vyhledávání informací „volá“ po sémantické podpoře.
- Vyhledávací stroje jsou zdokonalovány a pořád dokáží uživatele příjemně překvapit.
- Efektivnost je však v porovnání se sémantickým vyhledáváním nižší. . .

Efektivnost vyhledávání klíčová slova vs. ontologie

- F score
- obecné dotazy
KW: 20–44 %, ON: 92–94 %
- specifické dotazy
KW: 71–83 %, ON: 83–91 %

Pokračování závěru

- Sémantický web je vize.
- Její naplňování je během na dlouhou trať.
- Rozpor s očekáváním okamžitých efektů jak ze strany uživatelů, tak firem.
- Nyní je stabilizovaný stav v otázce **jak** metadata vyjadřovat.
- Další vývoj – způsoby získávání metadat.

Vize?

August 2009: How Google beat Amazon and Ebay
to the Semantic Web