

Uživatel:Slepi/Sémantická wiki

Wiki jsou skvělým nástrojem pro sdílení znalostí mezi lidmi. Tyto znalosti jsou ukládány ve formě formátovaného textu a multimediálních souborů, takže jsou snadno přístupné člověku (*čtenáři*). Nicméně počítač (*software*) si s těmito informacemi neví moc rady, neumí je automaticky kategorizovat, třídit, seskupovat ani zpracovávat. Tyto činnosti proto musí vykonávat člověk sám.

První příklad: Pokud si otevřete kategorii kategorii Biochemie, uvidíte spoustu článků, které se zabývají tématem blízkým oboru biochemie. Stejně je to u portálu Biochemie, který navíc obsahuje články seřazené do skupin podle témat. Všechny tyto *seznamy článků* vytvořil člověk.

Druhý příklad: Představte si, že se učíte na zkoušku z onkologie a chcete si nechat vypsát všechny typy nádorů, které postihují dělohu. Podíváte se tedy na stránku Nádory dělohy. Je to seznam článků, které na toto téma existují. Existují však ještě další, které v seznamu nenajdete (např. Nádory děložního hrdla). Je to tím, že tyto shrnující články spravuje *člověk*. Pokud byste chtěli vyhledávat *všechny* články, které se týkají nádorů v peritoneální dutině, jste úplně ztraceni.

Tady mohou pomoci data a sémantika, které jednotlivým článkům dají význam nejen pro člověka (*čtenáře*), ale i pro počítač. A ten pak může automaticky vyhledávat, kategorizovat, vytvářet souhrny nebo myšlenkové mapy. Nyní můžete ve vyhledávacím políčku vyhledávat názvy článků nebo jejich prostý text. Představte si svět, kde do vyhledávacího políčka zadáte myšlenku nebo dotaz a WikiSkripta vám rovnou dají odpověď (resp. seznam článků, kde ji naleznete).

Možnosti zpracování dat ve wiki

Sémantika a sémantická wiki

Sémantika je nauka zabývající se významem. Zejména se její bádání týká významů jednotlivých slov (resp. významové složky jazykového znaku, tj. vztahů mezi znaky a označovanými předměty, jevy či událostmi). Je to jedna z nauk jazykovědy (*lingvistiky*).

Sémantická wiki je wiki, která kromě nástrojů pro tvorbu textu a vzájemnou spolupráci autor obsahuje i sémantické funkce. Tyto funkce označí jednotlivá data (informace) tak, aby jim porozuměl i software. Zkrátka převede informace čitelné pro člověka do dat čitelných pro počítač.

Co v tomto případě považujeme za *data*? Jsou to páry **vlastnost – hodnota** přiřazené k určité konkrétní stránce (článku). Příklady takových párů přiřazených k článku uvádím v tabulce:

článek	vlastnost	hodnota
Pseudomonas aeruginosa	Gramovo barvení	G-
Akutní pankreatitida	ICD-10-CM	K85

Vlastně se jedná o data, která nejčastěji vídáme v infoboxech. Takto to vypadá triviálně a možná zbytečně. Ale zkuste si vyhledat všechny gramnegativní tyčky, které způsobují infekce močových cest! To obyčejná wiki neumí, musíte si projít (*člověkem vytvořený*) portál nebo kategorii, který vám dá (*člověkem vytvořený*) seznam bakterií, možná i (*člověkem vytvořenou*) patičku článků se seznamem gramnegativních tyček, ale najít si v člancích zmínku o močových cestách musíte sami. Vidíte, že v přípravě dat (tvorba portálů a infoboxů) i v hledání dat (studium) je spousta lidské práce, kterou by ale mohla zastat sémantická nadstavba. Pojďme na příklady, kdy to může pomoci!

Příklady

- Sémantická wiki třeba sama ví, který článek se týká biochemie. Portály a kategorie tak už nemusí udržovat redaktoři nebo uživatelé. Vše je kategorizováno automaticky a původní kategorie takřka nejsou potřeba.
- Sémantická wiki dokáže vyhledat všechny články, které popisují G- bakterie způsobující záněty močových cest. Podobný vyhledávací dotaz skončí v obyčejné wiki nezdarem a informace si student musí vyhledat sám článek po článku.
- Sémantická wiki dokáže sama vygenerovat (a to nejen vyplnit, ale přímo i vytvořit!) šablonu pro infobox včetně vrozených sémantických funkcí. Dokonce umožňuje vytvořit přehledný formulář, který běžnému uživateli pomůže ve vyplnění šablony.
- Sémantická wiki dokáže vytvořit strukturu pro článek na dané téma (např. zmíněné bakterie). Články pak mají všechny shodnou strukturu, která je navíc čitelná pro zpracování počítačem.
- Sémantické funkce využívá např. dříve velmi populární WikiApiary (https://wikiapiary.com/wiki/Main_Page).

Mezi sémantická rozšíření patří například Semantic MediaWiki nebo Cargo.

Data a Wikidata

Další možností, jak využít data obsažená ve wikiprojektech, jsou Wikidata. Je to ručně (příp. *roboticky*) spravovaná databáze dat, které se pak automaticky vkládají do infoboxů. Už nemusíte u každé bakterie ručně zadávat, jestli je G+ nebo G-, protože Wikidata to již (díky přispěvatelům) ví (<https://www.wikidata.org/wiki/Special:WhatLinksHere/Q>

857288). Pomocí dat mohou také roboti automaticky aktualizovat různé seznamy.

Wikidata jsou sekundární databáze, která sbírá strukturované údaje pro potřeby projektů založených na softwaru MediaWiki. Pro zpracování dat se využívá rozšíření Wikibase (<http://wikiba.se/>).

Rozšíření Semantic MediaWiki

V dalším textu se budu zabývat rozšířením *Semantic MediaWiki*. Budu se snažit popsat způsob práce s rozšířením a předvést ukázky zpracování dat. Tož pojďme instalovat.

Instalace Semantic MediaWiki

Možností instalace je několik. Vybral jsem si dost konzervativní neserverový způsob, protože je to nejjednodušší k použití na nějakém veřejném hostingu bez přístupu k serveru.

Do složky `/extensions/` jsem si stáhnul rozšíření ve formátu `.tar.gz` z GitHubu (<https://github.com/SemanticMediaWiki/SemanticMediaWiki/releases>). Po rozbalení se ukáže složka rozšíření mezi ostatními.

Nyní je nutné vložit do souboru `LocalSettings.php` následující řádky:

```
require_once "$IP/extensions/SemanticMediaWiki/SemanticMediaWiki.php";
enableSemantics('example.org');
```

Poté jsem už mohl přistoupit na stránku `Special:SMWAdmin`, na které stačí stisknout tlačítko `Initialise or upgrade tables`, které nainstaluje a upraví tabulky MediaWiki.

Nyní již extension funguje, což ověřím přidáním nějaké hodnoty do článku:

```
Sémantika je [[obtížnost::těžká]] [[kategorie::věda]].
```

Ukládání dat

Syntax ukládání dat je jednoduchá:

```
[[jméno vlastnosti::hodnota vlastnosti]]      (zároveň zobrazí odkaz na stránku "hodnota vlastnosti")
{{#set:jméno vlastnosti=hodnota vlastnosti}}  (nezobrazí nic, jen uloží data)
```

Uvedený kód zobrazí hodnotu vlastnosti jako odkaz. Vlastnosti lze procházet a upravovat na stránce `Special:Browse`.

Takže konkrétní příklad:

```
Pseudomonas aeruginosa je [[Gramovo barvení:gramnegativní]] tyčka.
Pseudomonas aeruginosa je gramnegativní tyčka. {{#set:Gramovo barvení=gramnegativní}}
```

Důležité je zmínit, že "Pseudomonas aeruginosa" a "Gramnegativní" jsou v prvním případě wikičlánky, zatímco "Gramovo barvení" je vlastnost zobrazitelná na stránce `Property:Gramovo barvení`.

Zobrazení dat

Všechna zadaná data dovedu zobrazit na několika místech:

- `Special:Browse` dovede v datech vyhledávat a zobrazí je podle článků
- `Property:Name` zobrazí data související s danou vlastností
- `Special:Types` zobrazí data danéh typu
- `Special:Properties` zobrazí všechny vlastnosti uložené ve wiki

Pokud chci nějaká dat zobrazit ve článcích, můžu použít příkaz `show`:

```
== Záněty močových cest ==
Tyto záněty způsobuje Pseudomonas aeruginosa, což je {{#show:Pseudomonas aeruginosa|?Gramovo barvení}} tyčka.
```

Příkaz zobrazí hodnotu vlastnosti "Gramovo barvení" z článku "Pseudomonas aeruginosa", což jak jsme si řekli výše, je "gramnegativní". V textu se tedy objeví slůvku "gramnegativní".

Takhle to moc lákavě nevypadá, ale což takhle si nechat zobrazit všechny bakterie automaticky?

```
== Záněty močových cest ==
Tyto záněty způsobují následující patogeny:
{{#ask: [[způsobuje onemocnění::uroinfekce]] }}
```

Nebo dokonce:

```
== Záněty močových cest ==
Tyto záněty způsobují následující bakterie:
{{#ask: [[Category:Bakterie]] [[způsobuje onemocnění::uroinfekce]] }}
```

Uvedený kód zobrazí všechny bakterie (všechny články s odkazy), v nichž je uložena vlastnost "způsobuje onemocnění" a obsahuje hodnotu "uroinfekce". Takže už žádné složité šablony se seznamem bakterií, vše se vypíše automaticky!

Šablony a formuláře

Tohle všechno zatím trochu zavání programováním a složitým wikikódem. Ale my potřebujeme něco hezkého, čtivého a jednoduchého. Od toho jsou tu šablony a formuláře. A na ně budeme potřebovat rozšíření Page Forms.

Instalace Page Forms

Rozšíření *Page Forms* se dá stáhnout ze stránek MediaWiki. Archiv se rozbálí do složky `/extensions/` a do `LocalSettings.php` vložíme startovací kód:

```
wfLoadExtension( "PageForms" );
```

A je to.

Šablony a infoboxy pomocí Page Forms

Každý zná infoboxy. Vytvořit infobox je dlouhá drbárna s nejistým výsledkem. Přitom *Semantic MediaWiki* umí vytvořit šablonu infoboxu během pár minut. Stačí přejít na stránku `Special:CreateTemplate` a zadat jméno a parametry šablony. Celá šablona včetně dokumentace je pak vygenerována automaticky.

Ukázkou může být šablona `Entry` (<https://directory.fsf.org/wiki/Template:Entry>) na stránkách Free Software Foundation (FSF). Uživatel pak může jakýkoli článek tvořený touto šablonou jednoduše editovat pomocí formuláře. Ukázkou je opět článek o příkazu `sed` (<https://directory.fsf.org/wiki?title=Sed&action=formedit&debug=true>) na stránkách FSF.

Formuláře pomocí Page Forms

Vytvoření formuláře k šabloně je extrémně jednoduché. Na stránce `Special:CreateForm` stačí zadat jméno formuláře a šablonu, kterou chceme formulářem plnit, a formulář se vytvoří sám.

Uvedl jsem příklady z FSF, proto odkazuji na formulář ke stejnojmenné šabloně `Entry` (<https://directory.fsf.org/wiki/Form:Entry>). Tímto formulářem lze založit i úplně nový článek.

Editace článku prostřednictvím formuláře

Výše jsem odkázal na zajímavou věc, editaci článku pomocí formuláře. Šlo o článek o příkazu `sed` (<https://directory.fsf.org/wiki?title=Sed&action=formedit&debug=true>) na stránkách FSF. Jak docílit toho, aby vedle možností `Editovat` a `Editovat zdrojový kód` byla i možnost editovat článek v podobě formuláře?

Lze to zařídit parsovacím příkazem buď v článku, nebo ve vkládané šabloně:

```
<span style="display:none">{{#default_form:Infobox}}</span>
```

"Infobox" naradíme skutečným jménem formuláře a značky omezující zobrazení jsou tam z toho důvodu, že parsovací funkce zobrazuje v článku zbytečný text "Tato kategorie využívá formulář Infobox."