

Technická univerzita Liberec

**Fakulta mechatroniky, informatiky a mezioborových
studií**

Studijní program: N2612 – Elektrotechnika a informatika

Studijní obor: 1802T007 – Informační technologie

**Zhodnocení Wikipedie jako zdroje
informací**

**Valuation of Wikipedia as
information source**

DIPLOMOVÁ PRÁCE

Vypracovala: Hana Teubelová

Vedoucí práce: Mgr. Milan Keršláger

Rok: 2010

Prohlášení

Byla jsem seznámena s tím, že na mou diplomovou práci se plně vztahuje zákon č. 121/2000 Sb. o právu autorském, zejména § 60 – školní dílo.

Beru na vědomí, že Technická univerzita v Liberci (TUL) nezasahuje do mých autorských práv užitím mé diplomové práce pro vnitřní potřebu TUL.

Užiji-li diplomovou práci nebo poskytnu-li licenci k jejímu využití, jsem si vědoma povinnosti informovat o této skutečnosti TUL; v tomto případě má TUL právo ode mne požadovat úhradu nákladů, které vynaložila na vytvoření díla, až do jejich skutečné výše.

Diplomovou práci jsem vypracovala samostatně s použitím uvedené literatury a na základě konzultací s vedoucím diplomové práce s konzultantem.

V Liberci dne

.....

Hana Teubelová

Poděkování

Děkuji Mgr. Milanu Keršlágerovi za vedení mé diplomové práce a za podnětné návrhy, které ji obohatily.

Hana Teubelová

Název práce:

Zhodnocení Wikipedie jako zdroje informací

Autor: Hana Teubelová

Obor: N2612 – Elektrotechnika a informatika

Druh práce: Diplomová práce

Vedoucí práce: Mgr. Milan Keršlágner

Konzultant: —

Abstrakt: Wikipedie je v současnosti velmi známou a oblíbenou internetovou encyklopedií. Mnoho lidí se na ní každodenně obrací, ale jen málokdo dokáže objektivně zhodnotit, zda-li se dá brát jako věrohodný zdroj informací. Čtenáři se seznámí s Wikipedií jako encyklopedií pro vyhledávání informací, ale také se dovědí o možnostech jejího rozšíření či opravování a jak se články naložit, aby po editování spíše projektu neuškodili než-li ho vylepšili. Práce se také zabývá problémem vandalství a navrhuje možná vylepšení jak celého projektu, tak i přístupu jednotlivců. Praktická část je věnována robotům ve Wikipedii. Pro mou práci mají za úkol porovnat anglickou a německou Wikipedii s českou podle počtu článků, vnitřních odkazů a pahýlů. Další robot je zaměřen na opravu chyb ve Wikipedii. Čtenář zde nalezne postup k vytvoření a nasazení takového bota na Wikipedii.

Klíčová slova: Wiki, Wikipedie, encyklopedie, robot

Title:

Valuation of Wikipedia as information source

Author: Hana Teubelová

Abstract: Wikipedia is very known internet encyclopedia in the present. Lots of people are searching here for information every day, but only few of them are able to say if it is a credible information source. Readers learn more about Wikipedia as encyclopedia and also about its extension and correction possibilities, or how to work with the articles. Work deals also with the vandalism problem and suggests some improvements of the whole Wikipedia project. Practical part follow the robots in Wikipedia. For my work they have to compare english, german and czech version of Wikipedia according to the number of articles, inner references and stubs. The next robot is correcting mistakes in Wikipedia. In this work readers can find out how to create and run such a robot on Wikipedia.

Key words: Wiki, Wikipedia, encyclopedia, robot

Obsah

Úvod	9
1 Co je Wikipedie	10
1.1 Historie anglické verze Wikipedie	10
1.2 Historie české verze Wikipedie	11
1.3 Software UseModWiki	12
1.4 Software MediaWiki	13
1.5 Jak psát do Wikipedie	13
2 Články na Wikipedii	18
2.1 Druhy článků	18
2.2 Hodnocení článků	20
2.3 Uživatelé Wikipedie	21
2.4 Vandalismus na Wikipedii	23
2.4.1 Druhy vandalství	24
2.4.2 Boj proti vandalům	24
2.5 Projekty pro korekci Wikipedie	25
3 Použité technologie	27
3.1 C#	27

3.2	DotNetWikiBot Framework	28
3.3	Microsoft Visual Studio 2005	28
3.4	Konfigurace	29
4	Porovnání úrovně poskytovaných informací na Wikipedii	30
4.1	Porovnání Wikipedie s ostatními encyklopediemi	30
4.1.1	Knižní vydání	30
4.1.2	Elektronické encyklopedie	31
4.2	Automatické porovnání anglické a německé Wikipedie s českou	33
4.2.1	Zjištěné hodnoty	34
4.2.2	Vyhodnocení zjištěných dat	34
5	Rozšíření a korekce článků	36
5.1	Vytvoření robota na Wikipedii	36
5.1.1	Instalace MediaWiki 1.6.12	36
5.1.2	Popis funkce robota	40
5.2	Návrh postupu rozšíření Wikipedie	42
5.2.1	Přístup k uživatelům	42
5.2.2	Podpora projektů na Wikipedii	44
5.2.3	Vyhledávání	44
5.2.4	Doplnění pahýlů	45
	Závěr	46
	Seznam použitých zdrojů	48
	Obsah CD	50

Úvod

Každý z nás hledá informace jiným způsobem. Ti, kteří mají přístup k internetu, je často vyhledávají prostřednictvím různých virtuálních encyklopedií. Jednou z nejrozšířenějších se postupem času stala Wikipedie, mnohojazyčná webová encyklopedie, jež je neustále rozšiřována dobrovolnými přispěvateli z celého světa.

Mnozí z nás mají však s touto encyklopedií i neblahé zkušenosti v podobě neúplných, zcela mylných článků, nebo v ní nelze některé informace dohledat.

I já se k této encyklopedii často obracím jako ke zdroji informací, které využívám nejen pro studium. Většinou je ještě potřeba článek porovnat s jeho cizojazyčnou verzí, pro důvěryhodnost informací. Proto bych vás chtěla v této diplomové práci nejen s Wikipedií seznámit, ale také objektivně články ve Wikipedii zhodnotit, doplnit a opravit.

Kapitola 1

Co je Wikipedie

Wikipedie je multijazyčná encyklopedie, do které mohou uživatelé libovolně přispívat. Název Wikipedie vznikl složením slov wiki a encyklopedie. Wiki se používá pro označení webů, ve kterých může kdokoliv přidávat, či upravovat jejich obsah.

Wiki dokumenty se vytváří pomocí jednoduchého značkovacího jazyka za použití webového prohlížeče.

1.1 Historie anglické verze Wikipedie

První wiki stránky se datují do roku 1995. Naprogramoval je Ward Cunningham a byly nazvány WikiWikiWeb. Tento web mohl být kýmkoliv upravován. Cunningenova myšlenka nazvat své stránky WikiWikiWeb vznikla po navštívení Havaje, kde slova wiki wiki vyjadřují autobusovou kyvadlovou dopravu na letišti v Honolulu.[6]

Předchůdcem Wikipedie se však stala jiná encyklopedie, Nupedia, založena Jimmym Walesem. Rodák z Albamy, narozen roku 1966, se poprvé s počítačem dostal do styku na střední škole, poté také na Auburn University a University of Alabama. Postupem času pracoval jako manažer v Chicagu. Jeho úspěch mu umožnil věnovat se jeho zájmům v počítačích a především v internetu. Po spuštění portálu Bomis, který byl zaměřený na pop kulturu, začal v roce 2000 realizovat myšlenku otevřené internetové encyklopedie, Nupedia. Ta byla spuštěna v roce 2000 a ukončena v roce 2003. Tato encyklopedie však nemohla být veřejně upravována, kvůli srovnatelné kvalitě

s ostatními encyklopediemi. Do Nupedie psali pouze experti z daných odvětví a nejlépe s titulem Phd.. Jako editor Nupedie byl zaměstnán Larry Sanger, který o rok později navrhl, aby se využilo nástrojů wiki. [8]

V roce 2001 byla, současně s Nupedií, poprvé spuštěna anglická verze Wikipedie, encyklopedie do níž má přístup kdokoliv, kdo má zájem přidávat nová hesla. Každá stránka může být kýmkoliv upravována, a to z důvodu co nejvěrnějšího obsahu. Někteří uživatelé však mají snahu o jejich znehodnocení a upravují články do nesmyslné podoby, vytvářejí stránky s fiktivními informacemi a provádějí operace, které Wikipedii nepřidávají na dobrém jménu.

Wikipedie nejprve byla založena na systému UseModWiki, wiki softwaru napsaným Cliffordem Adamsem v programovacím jazyce Perl. Později však přešla na vlastní software s otevřeným kódem, na počátku wiki engine, napsaný v PHP s databází MySQL vytvořený Magnusem Manskem, který nyní používá i mnoho dalších wiki. Nyní je to kód MediaWiki napsaný Lee Danielem Crockerem.

V roce 2005 fungovala celá Wikipedia na 39 serverech umístěných na Floridě. Nyní je to 300 serverů na Floridě, 26 v Amsterdamu a 23 v Koreji.

Anglická verze Wikipedie je největší světovou encyklopedií se svými více jak 11 milióny stránek. Německá verze je na místě druhém. Proto většinu informací naleznete v anglické či německé verzi Wikipedie.

Jimmy Wales založil v roce 2003 nadaci Wikimedia, jež neprovozuje jen Wikipedii, ale i další projekty jako Wikislovník a tezaurus, úložiště volných dokumentů WikiSource, sdílenou sbírku multimediálních souborů Wikimedia Commons, otevřené zprávy Wikizprávy, Wikicitáty a Wikiknihy.[1]

1.2 Historie české verze Wikipedie

Česká verze Wikipedie byla založena 3. května 2002 Miroslavem Malovcem, který se stal až do prosince roku 2003 správcem Wikipedie. V této době však používala software UseMod. Při přechodu na MediaWiki se nezachovaly žádné články zveřejněné pod softwarem UseMod.

Dalším správcem se stal Vít Zvánovec, který stanovil dalších 11 správců. V roce 2005 byl tento správce nedobrovolně odvolán.[3]

Správcem musí být zkušený uživatel Wikipedie, který dokáže ostatním uživatelům poradit v jakékoliv situaci a navíc je mu umožněno provádět údržbářské činnosti. Musí mít zveřejněných alespoň 250 článků a alespoň tři měsíce musí být aktivní v projektu. Pokud správce vytrvá ve své funkci alespoň 6 měsíců, může požádat o práva byrokrata. Ten může měnit uživatelům úroveň uživatelských práv.

10. dubna 2004 se konalo první wiki-setkání. V této době se česká Wikipedie začala hojně rozrůstat, jak počtem článků, tak i počtem přispěvovatelů.

Wikipedisté se scházejí na místě zvaném Pod Lípou, kde vedou diskuse o obecnějších věcech týkajících se Wikipedie. Sem se můžete obracet se svými problémy, žádostmi a oznámeními.

V roce 2004 česká Wikipedie dosáhla hranice 3 tisíce článků a všechna softwarová hlášení byla přeložena do Češtiny. 18. ledna 2009 Wikipedie obsahuje přes 100 tisíc článků, 58 533 registrovaných uživatelů a 1 933 je aktivních uživatelů, kteří přispěli do Wikipedie v posledních 30 dnech.

Česká verze Wikipedie je na 20. místě v počtu článků z více než 250. jazykových verzí. Je 24x menší než anglická verze a 8x menší než německá. V roce 2006 připadlo na 1 tisíc Čechů 2,5 článku, v roce 2009 je to 11,8 článků.[2]

1.3 Software UseModWiki

Tento software pro wiki byl vytvořený Cliffordem Adamsem v letech 1999 – 2000 v jazyce Perl. Do roku 2002 byl používán ve všech verzích Wikipedie. Jednotlivé stránky v tomto systému byly uloženy v jednoduchých textových souborech. Produkt je licencovaný jako svobodný software pod GNU General Public License.

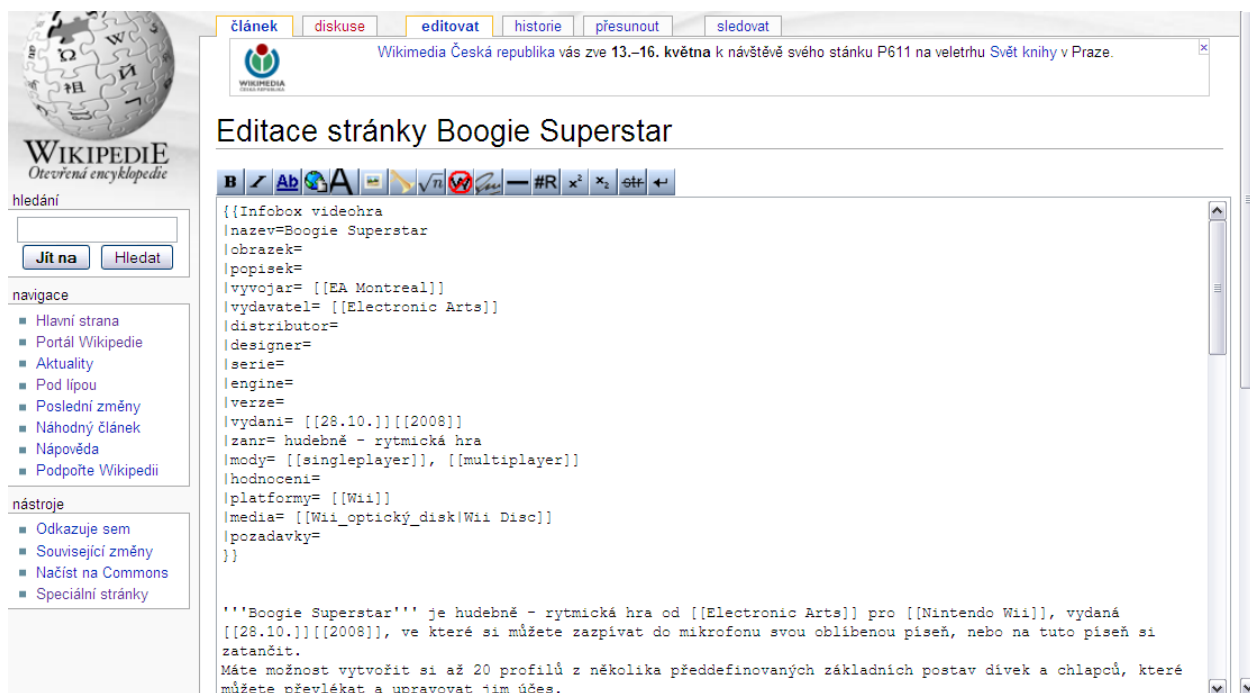
1.4 Software MediaWiki

MediaWiki je wiki systém, na kterém nyní běží Wikipedia a další projekty nadace Wikimedia. Je to svobodný software pod licencí GNU General Public License. Autor Magnus Manske napsal software v PHP za použití databáze MySQL, nebo PostgreSQL. Při uložení stránky MediaWiki ji zapíše do databáze, ale přitom zachovává předchozí verze, čímž je lze vrátit do původního stavu v případě vandalismu, nebo spamování. MediaWiki je pro uživatele velmi vstřícná, lze v ní napsat jakýkoliv text a pomocí \LaTeX i složité vzorce.[4]

1.5 Jak psát do Wikipedie

Psaní a editování článků ve Wikipedii má svá daná pravidla. Autor článku by měl mít dostatečně objektivní přístup a potřebné ověřené informace k tématu, o kterém píšete. Protože je Wikipedie otevřená encyklopedie, může být při případných pravopisných chybách, či nepřesnostech článek opraven dalšími uživateli. Každý článek se stává po zapsání do Wikipedie její stránkou. Článek by měl mít podobu encyklopedického hesla.

Základním funkčním prvkem wiki stránky je tlačítko editovat, které vám otevře editační okno se zdrojovým textem stránky. Ten po jeho upravení můžete uložit, čímž se jeho úpravy ihned zveřejní. V editačním oknu se neupravuje přímo HTML kód, ale zdrojová podoba wikitextu, jehož základní syntax je vypsán na další stránce.



Obrázek 1.1: Editovaná stránka ve Wikipedii

Při každém psaní nového článku by autor neměl zapomenout na infobox, nebo obrázek k danému článku, pokud ho má k dispozici, krátký úvod, vlastní text článku rozdělený do sekcí a podsekcí, a literaturu, odkud při psaní čerpal. Na závěr doplní seznam kategorií a mezijazykové odkazy.[5]

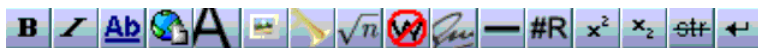
Pokud autor napíše velmi krátký článek, nebo takový, který dostatečně nepopisuje název článku, je nazván pahýlem a měl by být doplněn. Jako pahýly by neměly být označeny články, které dokáže upravit pouze odborník. Subpahýly jsou neúplné informace o daném tématu, zpravidla několika slovné. Bývají mnohem kratší než pahýly a v této formě by neměly setrvat delší dobu. Pokud autor nechce, aby byl subpahýl smazán, měl by ho co nejdříve doplnit alespoň na pahýl, nebo korektní článek.

Syntaxe v systému MediaWiki

Wikipedie používá systém MediaWiki, jenž je navržen pro velké množství uživatelů a stránek bez omezení funkčnosti a rychlosti. Při psaní do Wikipedie je proto potřeba používat danou syntaxi.

Syntax v systému MediaWiki	Výstup
"slovo"	<i>slovo</i>
''slovo''	slovo
'''slovo'''	<i>slovo</i>
~~~~	vrátí jméno uživatele
~~~~~	vrátí jméno uživatele a datum/čas
~~~~~	vrátí datum/čas
== hlavička sekce ==	dělí text do sekcí
=== podsekce ===	značí podsekce
==== menší podsekce =====	značí malé podsekce
*slovo	<ul style="list-style-type: none"> <li>• slovo</li> </ul>
**slovo	<ul style="list-style-type: none"> <li>• slovo</li> </ul>
***slovo	<ul style="list-style-type: none"> <li>• slovo</li> </ul>
*:	pokračování předešlé odrážky
*	ukončení jednoduché odrážky, další * značí začátek další jednoduché odrážky
#slovo	<ol style="list-style-type: none"> <li>1. slovo</li> </ol>
##slovo	<ol style="list-style-type: none"> <li>1. slovo</li> </ol>
:	odsazení
::	dvojnásobné odsazení – přidáním dalších přidáme další odsazení
[[odkaz]]	značí hypertextový odkaz
Normální text<small>malý text</small>	Normální text <small>malý text</small>
Normální text<big>velký text</big>	Normální text <b>velký text</b>
x₁	$x_1$
x¹	$x^1$

Tabulka 1.1: Syntax v systému MediaWiki



```
{{Infobox videohra
|navez=Boogie Superstar
|obrazek=
|popisek=
|vyvojar= [[EA Montreal]]
|vydavatel= [[Electronic Arts]]
|distributor=
|designer=
|serie=
|engine=
|verze=
|vydani= [[28.10.]] [[2008]]
|zanr= hudebně - rytmická hra
|mody= [[singleplayer]], [[multiplayer]]
|hodnoceni=
|platformy= [[Wii]]
|media= [[Wii_optický_disk|Wii Disc]]
|pozadavky=
}}
```

'''Boogie Superstar''' je hudebně - rytmická hra od [[Electronic Arts]] pro [[Nintendo Wii]], vydaná [[28.10.]] [[2008]], ve které si můžete zazpívat do mikrofону svou oblíbenou píseň, nebo na tuto píseň si zatančit.

Máte možnost vytvořit si až 20 profilů z několika předdefinovaných základních postav dívek a chlapců, které můžete převlékat a upravovat jim účes.

Při prvním přihlášení se naučíte základní kroky, jež děláte pomocí remotu, kde s ním děláte různé úkony do rytmu a tím ovládáte svou postavičku.

Pokud chcete přidat další taneční kroky, oblečení, účesy, nebo písně, musíte se zúčastnit soutěžní show. Tam se vždy utkají čtyři hráči ve třech kolech. Vy máte možnost si vybrat odvětví a písně, ve kterých se chcete utkat. Po dokončení show dostanete podle úspěšnosti v soutěži body, za které si kupujete jednotlivé balíčky pro Pop, Urban nebo Electro dance, v nichž jsou obsaženy právě další taneční kroky, oblečení, účesy a písně.

```
== Písně v Boogie Superstar ==
*Bleeding Love - [[Leona Lewis]]
*Bullseye - [[Aly & AJ]]
```

Obrázek 1.2: Editace stránky

# Boogie Superstar

**Boogie Superstar** je hudebně - rytmická hra od [Electronic Arts](#) pro [Nintendo Wii](#), vydaná [28.10.2008](#), ve které si můžete zaspívat do mikrofону svou oblíbenou píseň, nebo na tuto píseň si zatančit. Máte možnost vytvořit si až 20 profilů z několika předdefinovaných základních postav dívek a chlapců, které můžete převlékat a upravovat jim účes. Při prvním přihlášení se naučíte základní kroky, jež děláte pomocí remotu, kde s ním děláte různé úkony do rytmu a tím ovládáte svou postavičku.

Pokud chcete přidat další taneční kroky, oblečení, účesy, nebo písně, musíte se zúčastnit soutěžní show. Tam se vždy utkají čtyři hráči ve třech kolech. Vy máte možnost si vybrat odvětví a písně, ve kterých se chcete utkat. Po dokončení show dostanete podle úspěšnosti v soutěži body, za které si kupujete jednotlivé balíčky pro Pop, Urban nebo Electro dance, v nichž jsou obsaženy právě další taneční kroky, oblečení, účesy a písně.

<i>Boogie Superstar</i>	
<b>Vývojař(i)</b>	<a href="#">EA Montreal</a>
<b>Vydavatel(é)</b>	<a href="#">Electronic Arts</a>
<b>Platforma(y)</b>	<a href="#">Wii</a>
<b>Datum vydání</b>	<a href="#">28.10.2008</a>
<b>Žánr</b>	hudebně - rytmická hra
<b>Módy</b>	<a href="#">singleplayer</a> , <a href="#">multiplayer</a>
<b>Média</b>	<a href="#">Wii Disc</a>

## Písně v Boogie Superstar

[\[editovat\]](#)

- [Bleeding Love](#) - [Leona Lewis](#)
- [Bullseye](#) - [Alv & A.I](#)

Obrázek 1.3: Náhled editované stránky



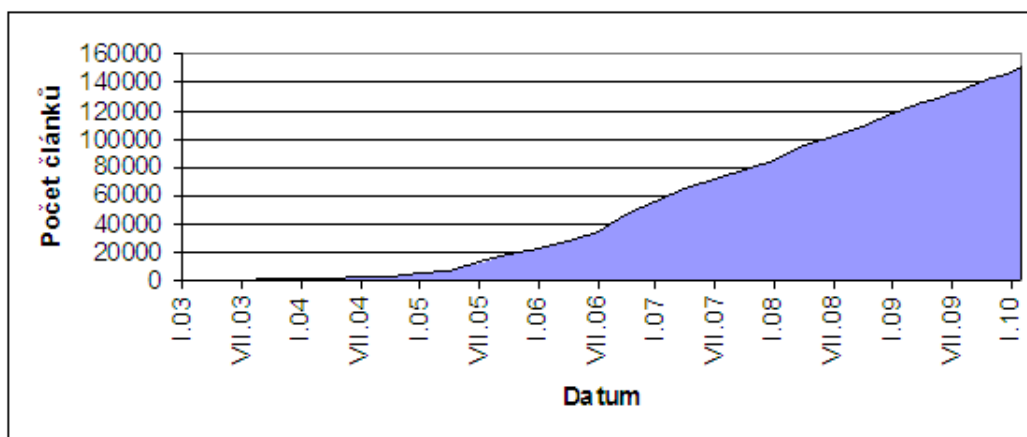
# Kapitola 2

## Články na Wikipedii

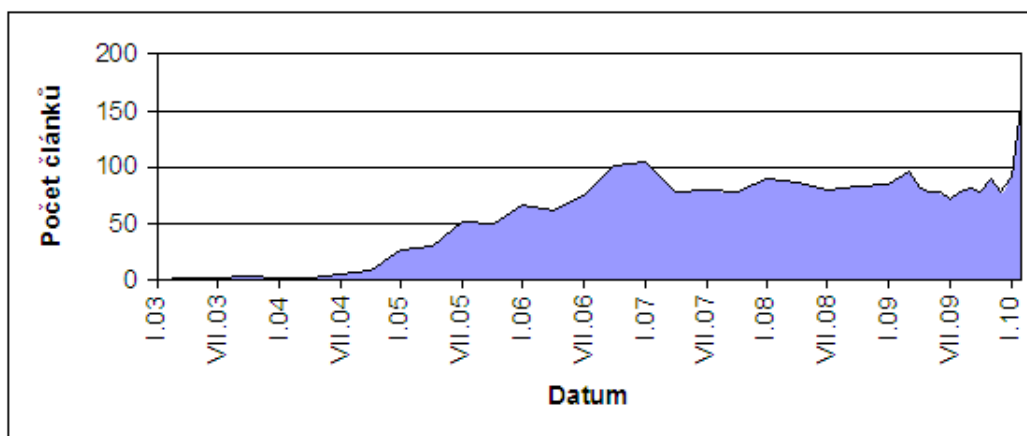
### 2.1 Druhy článků

Ve Wikipedii je článek definován jako stránka obsahující encyklopedickou či almanachovitou informaci. Ne všechny stránky na Wikipedii lze zařadit mezi články. Článkem nelze nazvat hlavní stranu, rozcestníkové stránky nebo stránky, které přesměrovávají na jiné stránky. Ostatní články lze rozdělit:

- článek – subpahýl – není považován za encyklopedicky přínosný, je velmi krátký a nejedná se o vandalismus
- článek – pahýl – chybí mu určitá informace, aby mohl být brán za kompletní encyklopedické heslo
- článek – základní úroveň encyklopedického hesla a obsahuje všechny informace vystihující heslo
- dobrý článek – splňuje kritéria vydaná WikiProjektem: Kvalita
- nejlepší článek – nepatrné množství článků, které splňují aktuální požadavky komunity nebo kritéria skupiny wikipedistů, jež se těmito články zabývají, na vzhled nejlepších článků



Obrázek 2.1: Vývoj počtu článků [11]



Obrázek 2.2: Vývoj počtu článků přidaných za jeden den [11]

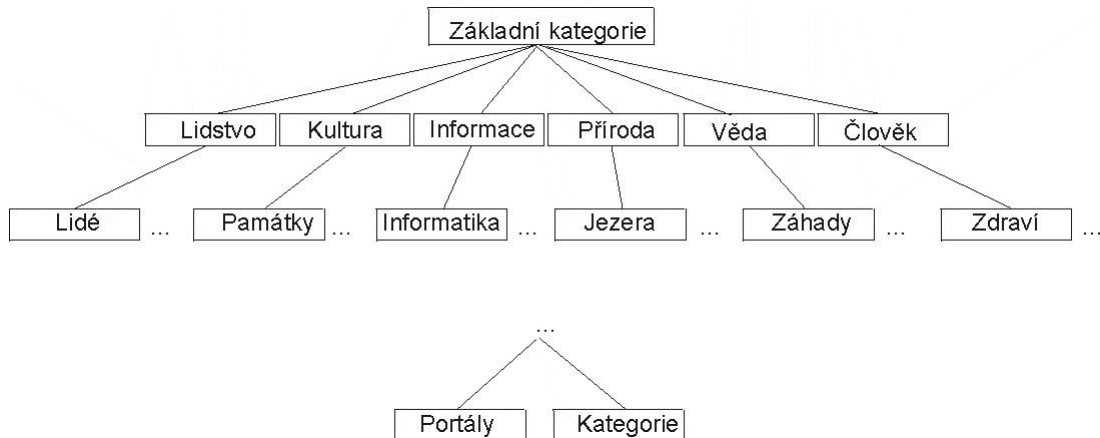
Vytvořit článek na Wikipedii lze několika způsoby:

- Kliknutím na červený odkaz, který můžete najít kdekoliv na Wikipedii.
- Pokud vyhledávač ve Wikipedii nenajde vámi zadaný název článku, zobrazí se text "Vytvořte stránku".
- Vytvořením vnitřního odkazu na stránce tak, aby se začlenil do textu. Poté se otevře editace článku při kliknutí na tento červeně označený odkaz.
- Vložením do adresního řádku prohlížeče <http://cs.wikipedia.org/wiki/Téma>, kde Téma je jméno článku, který chcete založit. Objeví se prázdná stránka dosud neexistujícího článku.

## 2.2 Hodnocení článků

V dubnu roku 2006 zkoumal články a tématické oblasti ve Wikipedii wikipedista Cinik. Třetina článků byla vyhodnocena jako dobrá, další třetina jako pahýly a zbytek byly subpahýly, nebo články vyžadující zásadní úpravy. 20 % všech témat byly matematické, zemědělské a přírodovědné stránky, 13 % astronomie a geografie, 4 % historické a 17 % životopisné články, 10 % sport, náboženství a umění, 9 % stavby a technika, 11 % ostatní témata, 4 % rozcestníky a 11 % přehledové články s letopočty a daty. Kvalitativně se nejvýše zařadily články z oblasti přírodních věd a techniky, naopak nejnižší se umístili články z oblasti humanitních věd vyjma náboženství, tedy z psychologie, sociologie, politiky, filosofie a práva.

Po necelých čtyřech letech jsem podobný průzkum provedla znovu. Zaměření článků jsem zjišťovala dle tematických portálů do kterých článek patří. Každý článek může patřit pod více portálů, ale také nemusí být pod portálem žádným. Vycházela jsem z kategorií, které jsou uvedeny vždy na konci stránky. Každá z kategorií může patřit do jiných kategorií, či portálů. Nejvyšším stupněm jsou základní kategorie.



Obrázek 2.3: Hierarchie kategorií

Protože každý článek může patřit pod více portálů, je zde uvedeno v procentech, kolik článků z celkového počtu je uvedeno pod daným portálem. 53 % článků je zařazeno pod portálem Historie, 43 % článků pod portálem Ekonomie, 47 % pod Astronomií, 46 % pod Fyziku, 40 % Právo, 36 % Válku, 28 % Umění a stejně má i Biologie, 23 % Kulturu, 22 % Filosofii, 16 % Jazyk, 15 % Itálii, 14 % Literaturu, 13 % Finsko, 11 % Dopravu. Zbylé portály obsahují pod 10 % článků.

Hodnocením kvality 100 nejsledovanějších článků české Wikipedie zkoumá i projekt české Wikiverzity Hodnocení kvality (top 100). Výzkum je prováděn členy WikiProjektu:Kvalita, jenž by rád na základě této studie vylepšil a zkvalitnil nejhorší články ze 100 nejsledovanějších na Wikipedii. Články jsou hodnoceny dle:

- NPOV – nezaujatý úhel pohledu
- Literatura
- Reference
- Úplnost
- Struktura
- Výklad
- Jazyk
- Formát
- Červené odkazy
- Úvod
- Obrázky

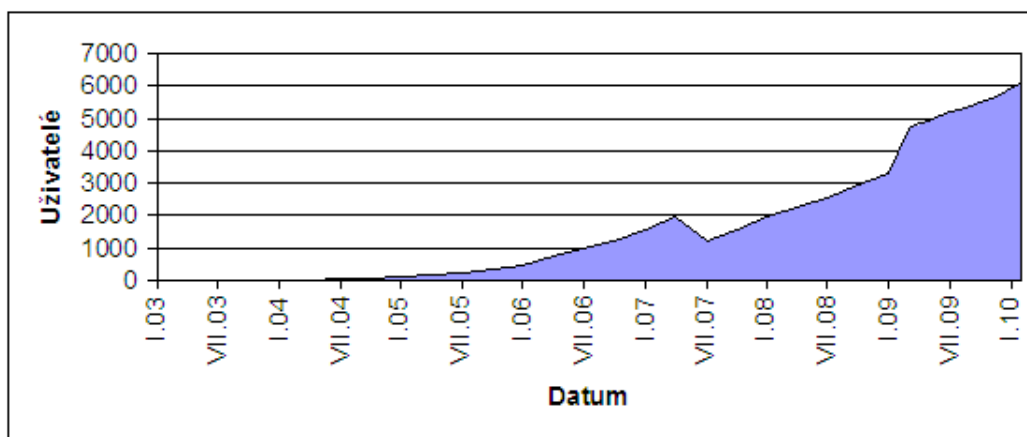
## 2.3 Uživatelé Wikipedie

Základem funkčnosti Wikipedie jsou její uživatelé. Dělí se do několika skupin podle rozsahu svých oprávnění. Právě oni dělají z Wikipedie velmi rozsáhlou encyklopedii s milióny článků ze všemožných odvětví, zabývají se řešením vandalismů, ale také radí, jak s Wikipedií zacházet.

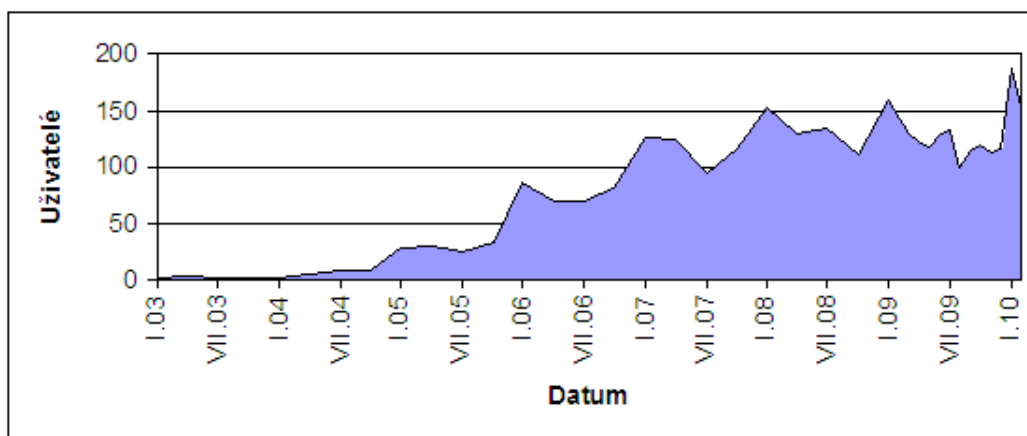
Rozdělení uživatelů:

- Registrovaní uživatelé
- Aktivní uživatelé – v posledních 30 dnech provedli nějakou operaci

- Roboti – automatické programy působící na Wikipedii přes její webové rozhraní
- Správci – uživatelé mající povolen přístup k některým omezeným funkcím, neboť jsou považováni za důvěryhodné
- Byrokraté – zároveň správci, ale navíc jsou jim propůjčena technická práva:
  - povýšit jiné uživatele na úroveň správce nebo byrokrata
  - přejmenovat uživatelské účty
  - udělit či odebrat uživateli status bota
- Revizoři – mohou ověřovat z jaké IP adresy edituje přihlášený uživatel a vypsat všechny uživatele používající IP adresu. Tyto informace jsou určitou dobu uloženy v databázi pro zabránění vandalismu. Minimálně 2 revizoři obsluhují českou Wikipedii, aby každý z nich mohl být kontrolován jiným revizorem.
- Stevardi – mohou libovolně měnit uživatelská práva libovolného uživatele na libovolném projektu nadace Wikimedia. Modifikují uživatelské přístupy nebo mají přehled o uživateli, kteří zneužili přístup do Wikipedie. Stevardi se volí jednou ročně.
- Importéři – importují stránky z jiných wiki
- Schválení uživatelé – od jejich registrace uplynuly 4 dny a současně přispěl alespoň 10 příspěvků. Tento uživatel získává práva
  - Editace polozamčených stránek
  - Nahrávání existujících souborů
  - Přesun stránek
  - Vykonávání akcí spojených s CAPTCHA (Turingův test používaný na webu pro automatické rozpoznání skutečného uživatele od robota) bez nutnosti jejího řešení



Obrázek 2.4: Vývoj počtu uživatelů [11]



Obrázek 2.5: Vývoj počtu uživatelů přidaných za jeden den [11]

## 2.4 Vandalismus na Wikipedii

Stránky, které může kdokoliv upravovat bývají velmi často napadány vandaly. Do této skupiny se řadí však i obyčejní uživatelé, kteří udělali při editování nepovolenou chybu, nebo si jen na zkoušku založili stránku se smyšleným obsahem. Více škody však nadělají ti, jenž promyšleným způsobem znehodnocují stránky Wikipedie. Proti zakázaným editacím bojují správci mající práva na smazání vandalských článků, ale i blokování uživatelů.

## 2.4.1 Druhy vandalství

- Spam – vkládání odkazů, které nesouvisí s daným článkem
- VandalBot – automatizované vkládání spamu nebo úmyslné ničení článků pomocí skriptu
- Dětský vandalismus – mazání obsahu stránek, přidávání textu nesouvisejícího s článkem
- Hloupé vtípky – vkládání textu, jenž vypadá velmi věrohodně, nebo hoaxů
- Lstivý vandalismus – seriózně vypadající editace, které jsou přesto nepravdivé a navíc podpořeny jinými seriózními editacemi vandala
- Vandalismus formou přesouvání stránek – vandal přesouvá existující stránky na nová místa pod nesmyslnými či urážlivými jmény
- Napadání uživatelských stránek – urážky, jež jsou vkládány na stránky určitého uživatele (druh pomsty)

Různé pokusy nováčků nemusí vždy být výsledně vyhodnoceny jako vandalství. Každý uživatel je na své vandalství upozorněn a má možnost své počínání vysvětlit. Pokud uživatel na své vytvořené stránce najde označení NPOV, porušení nezaujatého úhlu pohledu, má opět možnost se k tomuto nařčení vyjádřit. [12]

Největším problémem řešení konfliktů mezi správci a nařčenými uživateli bývá fakt, že většinou dialog mezi těmito osobami končí vysvětlením uživatele, ale správce dále mlčí. Nastává otázka, jak se zbavit nařčení a předpokladů, že je uživatel vandalem. Tento fakt jsem si osobně vyzkoušela na stránce Hewlett-Packard. Text je takřka celý přeložen z anglické stránky, kde s ním nemají sebemenší problém. Na české stránce byl však text označen za NPOV. Správce však na vysvětlení obsahu článku nereagoval.

## 2.4.2 Boj proti vandalům

Každý uživatel má možnost v článku vrátit editaci. To je jedna z věcí, kterou může kdokoliv při přistižení vandalství s článkem provést. Proti vandalům především

bojují správci. Mnoho problémů se řeší vložením šablon do článku[12]:

- {{subst:Experimenty}} – první upozornění na vandalismus
- {{subst:Experimenty2}} – opakovaný vandalismus i přes upozornění, přidávání nesmyslů
- {{subst:Experimenty2o}} – opakovaný vandalismus i přes upozornění, odstraňování obsahu
- {{subst:Experimenty3}} – opakovaný vandalismus i přes mnohá upozornění, přidávání nesmyslů. Poslední varování
- {{subst:Spam}} – vkládání nesouvisejících odkazů (v několika jazycích)
- {{subst:Multispam}} – vkládání nesouvisejících odkazů (v několika jazycích)
- {{subst:KomentářPOV}} – vkládání komentářů porušujících NPOV
- {{subst:KomentářPOV2}} – opakované vkládání komentářů porušujících NPOV, důraznější upozornění

## 2.5 Projekty pro korekci Wikipedie

Na Wikipedii existuje několik projektů pro vylepšení článků na Wikipedii a zvýšení její přehlednosti, především pak pro stanovení určitých pravidel při vytváření stránek. Obecně se nazývají Wikiprojekty. Seznam všech projektů naleznete na Wikipedii pod názvem Kategorie:WikiProjekty.

Většina projektů se zaměřuje na stránky z jednoho oboru, které se snaží upravit do specifické podoby. Další WikiProjekty však řeší otázky celé Wikipedie:

- WikiProjekt Check Wikipedia – vyhledává chyby v člancích, které je potřeba upravit.
- WikiProjekt Kvalita – účastníci projektu se snaží co nejvíce zkvalitnit články na české Wikipedii. Navíc projekt hodnotí kvalitu již existujících článků tím, že označuje články za dobré či nejlepší a snaží se udržet jejich úroveň.[10]



- WikiProjekt Autorské právo – má za úkol řešit problémy s autorskými právy na Wikipedii a vydává určitá pravidla a doporučení, jak toto právo neporušit.
- WikiProjekt Zdrojování – doplňuje zdroje na stránky Wikipedie. Článek, ve kterém nejsou uvedeny zdroje, bývá zpochybněn. Neozdrojovat článek a využívat zdroj může být také protiprávní.
- WikiProjekt Wikipedie v médiích/Wikipedie na Facebooku – se snaží popularizovat Wikipedii na této sociální síti. Fanoušky se snaží pravidelně zásobovat informacemi o dění na české Wikipedii.

# Kapitola 3

## Použité technologie

Při rozhodování jaký programovací jazyk bude použit pro tvorbu programů pracujících s Wikipedií, jsem zvolila C#. Především proto, že je to vysokoúrovňový objektově orientovaný jazyk a pro jeho kompatibilitu s DotNetWikiBot frameworkem, který je též napsán v C#, kde si ho lze podle subjektivních požadavků změnit.

### 3.1 C#

Vysokoúrovňový objektově orientovaný jazyk C# vyvinula společnost Microsoft. Vychází z jazyků C++ a Java, a syntaxi čerpá z jazyka C. Nejčastěji se využívá k tvorbě webových aplikací, databázových programů, nebo softwaru pro mobilní zařízení jako jsou mobilní telefony či PDA.

Byl schválený standardizačními komisemi ECMA (ECMA-334).[13] Tato mezinárodní norma stanoví pravidla a zavádí výklad programů vytvořených v jazyce C++, kde upřesňuje syntaxi a omezení, sémantická pravidla či přenositelnost zdrojového kódu programu full-feature.[ecma]

Vlastnosti C#:

- neexistuje vícenásobná dědičnost ani globální a proměnné metody
- používají se metody `get(accesor)` a `set(mutator)`
- je typově bezpečnější než C++, existují pouze bezpečné konverze jako rozšiřování

integerů, nebo konverze z odvozeného typu na rodičovský

- rozlišuje velká a malá písmena – Case sensitive
- podporuje paralelní spouštění kódu pomocí tzv. multithreadingu

## 3.2 DotNetWikiBot Framework

DotNetWikiBot Framework byl vytvořen pro zjednodušení práce s jinak složitými úkoly ve wiki webech. Tento framework je napsán v jazyce C# a založen na platformě Microsoft .NET Framework 2.0. Lze stáhnout v .zip souboru a nevyžaduje instalaci.

Umožňuje vytváření webových robotů a dalších programů, které zpracovávají informace o wiki stránkách založených na softwaru MediaWiki. DotNetWikiBot se používá pro programování v C# a .NET. Je objektově orientovaný, proto je přístup k objektům a jejich metodám velmi jednoduchý.

V DotNetWikiBot Frameworku existuje 5 typů objektů :

- Site – Definuje wiki stránky. V programech je použit pro načtení stránek dle URL.
- Bot – Definuje instance bota, některá konfigurační nastavení a pomocné funkce.
- Page – Definuje jednotlivé wiki články. V programech se používají metody pro načtení obsahu článku, jeho editování a ukládání.
- Pagelist – Definuje soubor wiki článků. Je použit pro načtení všech článků Wikipedie.
- WikiBotException – Vytváří vlastní aplikace výjimky.

## 3.3 Microsoft Visual Studio 2005

Pro tvorbu všech programů bylo použito vývojové prostředí Microsoft Visual Studio 2005. Je to kompletní sada vývojářských nástrojů pro vytváření ASP.NET, XML

Web Services, desktopových a mobilních aplikací. Z důvodu použití stejného integrovaného vývojového prostředí, sdílejí Visual Basic, Visual C ++, Visual C# a Visual J # nástroje při vytváření smíšených jazykových řešení.

Jako všechna vývojová prostředí obsahuje Visual Studio 2005 editor kódu podporující zvýraznění syntaxe a automatické dokončování, debugger pracující jak na úrovni kódu, tak na úrovni stroje a další nástroje jako designer formulářů pro tvorbu aplikací grafického uživatelského rozhraní, designer webu, tříd a databázových schémat.

### 3.4 Konfigurace

Všechny programy byly spouštěny na notebooku s operačním systémem Windows XP Professional. Počítač je založen na procesoru AMD X86 s celkovou fyzickou pamětí 1536MB.

# Kapitola 4

## Porovnání úrovně poskytovaných informací na Wikipedii

### 4.1 Porovnání Wikipedie s ostatními encyklopediemi

Při vyhledávání informací se často musíme obracet na encyklopedie s danou tematikou. Velkou výhodou mají encyklopedie v elektronické podobě, v nichž se snadněji vyhledává než v těch, které mají podobu knižní.

#### 4.1.1 Knižní vydání

Knižní vydání encyklopedií je často provázeno značnou hmotností a nepřehledností svazku. Pokud člověk vyhledává v knihách nemalých rozměrů informace, jejichž názvy v rejstříku nenachází, může to být velkým problémem. Zato tyto knihy jsou ve většině případů sestaveny úspěšnými odborníky, kteří se danou tematikou zabývají takřka celý svůj život. Potom se na tyto informace, pokud porozumíte autorově vyjadřování, můžete spolehnout. Abychom se ale vůbec ke knižní encyklopedii dostali, musíme se smířit s její vysokou cenou obsahující honoráře pro tyto odborníky.

## 4.1.2 Elektronické encyklopedie

Lepší varianta encyklopedií je proto elektronická. Jednou z velmi známých je Ottův slovník naučný, největší česká encyklopedie, původně tištěná mezi 19. a 20. století. Naleznete v ní přes 250 tisíc hesel ze všech možných odvětví, stačí jen zadat do vyhledávání. Dále si můžete prostřednictvím křížových odkazů vyhledávat různé souvislosti. Je opět napsána odborníky, takže informacím lze věřit. Tuto encyklopedii si však musíte zakoupit za nemalé peníze a dodávání nových informací zde nefunguje. Nejshůdnější možností hledání informací je proto elektronická podoba encyklopedií, bezplatná a ještě k tomu s neustálým rozšiřováním.

### **Knol**

Knol je dílo Googlu a je vnímán coby snaha konkurovat projektům jako je právě Wikipedie. Google však prezentuje Knol jako doplnění dalších encyklopedií. Jeho název je odvozen od slova knowledge – vědomosti. Projekt byl představen v prosinci 2007 a do provozu byl postaven v červenci 2008. [7] Podle Googlu má garantovanou odbornou kvalitu článků, ty jsou totiž psány jednotlivcem či celým týmem, jenž má za práci nárok na odměnu v podobě příjmu z reklamy, kterou mohou zveřejnit u svého článku. Jeho tvůrce však nese odpovědnost za v něm uvedené informace. Ale v reálu je Knol pro mnoho uživatelů jen hosting jejich výtvorů.

U Knolu se však pro vložení článku musíme zaregistrovat, kdežto ve Wikipedii se náš příspěvek může zapsat pod IP adresou. Wikipedie je propracovanější ve vkládání nestandardních objektů. V Knolu se pohodlně dají vložit jen obrázky a text, ale problémem mohou být tabulky, které někteří autoři vkládají jako obrázky a tím kazí celkový vzhled jejich stránky. Nový článek musí obsahovat nadpis, krátký abstrakt, alternativní nadpis a samotný text, který se dělí na kapitoly, stejně jako ve Wikipedii.

Každý autor si sám rozhodne, zda-li obsah článku bude tvořit zcela sám, všichni uživatelé zcela libovolně, nebo kdokoliv s jeho souhlasem.

Jednotlivá hesla se v Knolu mohou opakovat, tudíž k jednomu heslu naleznete další duplicitních stránky. Ty však můžete známkovat od 1 až 5, jako ve škole, nebo je okomentovat. Poté při vyhledávání se jednoduše obrátíte na článek, který je nejlépe

ohodnocen, či komentován. Ve Wikipedii nenaleznete známkování, ale špatné články jsou rovnou přepracovány a upraveny, což často pro vysvětlení hesla stačí. Wikipedia se snaží vyvarovat duplicitě stránek, což je jejím dalším plusem.

Knol má zatím jen několik stovek článků, z většiny na velmi dobré úrovni. Ale pokud použijete pro vysvětlení hesla vyhledávač Google, dozvíte se mnohem víc.

## **Britannica**

Tato encyklopedie poprvé vyšla v knižním vydání v letech 1768 – 1771 ve Skotsku. Po roce 1901 koupili práva na vydávání američtí podnikatelé Horace E. Hooper a Walter M. Jackson a všech 29 svazků 11. vydání vyšlo v roce 1911. Z této verze čerpá anglická Wikipedie. Od roku 1929 vychází každý rok aktuální verze. Třetí generace Britannicy vychází od roku 1974 a je to 15. vydání. Od 90. let vychází Britannica také na CD.

V roce 2007 bylo v anglické Wikipedii kolem 2 milionů záznamů, v Britannice kolem půl milionu. Největší spor se však vede o přesnosti článků. Časopis Nature publikoval výzkum, který udává, že obě encyklopedie jsou přibližně stejně přesné. Britannica tento výzkum však zpochybnila. Právě velkou výhodou Wikipedie je, že se články opravují velmi rychle, ale také to, že je bezplatná.

## **Citizendum**

Wiki encyklopedie Citizendum založena Larrym Sangerem se na Wikipedii snažila navázat, přetvořit Wikipedii bez jejích nedostatků. Nelze zde anonymně přispívat, ani editovat pod přezdívkou. Při vytváření stránky musíte dodržovat striktní pravidla. Editoři musí uvést své odborné znalosti, podle kterých budou jejich úpravy posuzovány. Citizendum je zatím mnohem méně známé než Wikipedie a množství článků je proti Wikipedii takřka zanedbatelné. Plusem je však eliminování vandalismu.

## 4.2 Automatické porovnání anglické a německé Wikipedie s českou

Porovnávání je zjišťováno od 5. října 2009. Celá Wikipedie má kolem 7,5 milionu článků, z toho v anglické Wikipedii nalezneme přes 3 miliony článků, v české kolem 140 tisíc a v německé přes 970 tisíc. Nejen počtem článků se však verze liší. Významným rozdílem je obsah jednotlivých stránek.

Pomocí programu napsaného v jazyce C# a DotNetWikiBot Frameworku jsem porovnávala tyto verze Wikipedie. Především jsem se zaměřila na počet slov a vnitřních odkazů na stránkách. Porovnávány jsou jen ty stránky, které existují v českém i anglickém, nebo německé jazyce. Tyto údaje jsem jednou měsíčně aktualizovala, abych zjistila, jakým směrem se Wikipedia vyvíjí.

Dalším úkolem programu je zjištění počtu pahýlů na Wikipedii. Je to označení článků, kterému chybí určitá informace, aby mohl být brán za kompletní encyklopedické heslo. Posuzování pahýlů je však velmi subjektivní a proto nebudu jejich počet srovnávat s jinými verzemi Wikipedie. V některých verzích totiž ani pahýly neexistují, proto i velmi krátké články jsou brány jako plnohodnotné.



## 4.2.1 Zjištěné hodnoty

Význam zkratk v tabulce:

- S ČJ : AJ – Poměr stránek dle počtu slov v ČJ a AJ
- O ČJ : AJ – Počet stránek dle počtu vnitřních odkazů v ČJ a AJ
- S ČJ : NJ – Poměr stránek dle počtu slov v ČJ a NJ
- O ČJ : NJ – Počet stránek dle počtu vnitřních odkazů v ČJ a NJ
- P AJ – Počet stránek v češtině, které lze přeložit do angličtiny
- P NJ – Počet stránek v češtině, které lze přeložit do němčiny

Datum	S ČJ : AJ	O ČJ : AJ	S ČJ : NJ	O ČJ : NJ	P AJ	P NJ
5.10.2009	17.457 : 78.995	21.168 : 75.284	19.029 : 60.708	25.330 : 54.407	96.452	79.737
7.11.2009	18.076 : 80.496	21.873 : 76.699	19.413 : 61.795	25.827 : 55.381	98.572	81.208
21.11.2009	18.272 : 80.952	22.194 : 77.030	19.607 : 62.197	26.049 : 55.755	99.224	81804
19.4.2010	19.711 : 83.201	24.897 : 78.015	20.831 : 64.400	27.802 : 57.429	102.912	85.231

Tabulka 4.1: Porovnání Wikipedie v českém, německém a anglickém jazyce

Datum	Počet pahýlů
7.11.2009	60.011
21.11.2009	60.329
19.4.2010	68.296

Tabulka 4.2: Počet pahýlů na české Wikipedii

## 4.2.2 Vyhodnocení zjištěných dat

Na počátku měření z více než 140 tisíc stran šlo přeložit do anglického jazyka 96 452 a právě tyto stránky byly porovnány. 17 457 jich mělo více slov na české stránce

než-li na anglické. Do němčiny šlo přeložit 79 737 stran a z toho 19 029 jich mělo v Češtině více slov. Na tyto údaje má nemalý vliv počet lidí, kteří ve svém rodném jazyce hovoří anglicky, nebo německy.

Ve druhém měření, o měsíc později, se ukázalo, že lze přeložit z českého do anglického jazyka přes 2 tisíce a do jazyka německého 1,5 tisíce stránek více. Tato skutečnost může být způsobena novými stránkami v Češtině vycházejících z anglických stránek a naopak, ale i postupným doplňováním odkazů na stránky v jiných jazycích. Větší počet slov v českém jazyce oproti anglickému má o 500 stran více, v anglickém jazyce je to však o 1500 stránek.

V dubnu 2010 byl počet stránek oproti listopadu o 20 tisíc vyšší, ale jen počet pahýlů vzrostl o 8 tisíc. Nové články bohužel na důvěryhodnosti Wikipedie nepřidávají a editoři by se měli především zaměřit na vhodné doplnění pahýlů. Počet přeložitelných stránek vrostl o 6 tisíc u anglické i německé Wikipedie, přesto se však počet slov i vnitřních odkazů rychleji zvyšuje u cizojazyčných Wikipedií.

# Kapitola 5

## Rozšíření a korekce článků

### 5.1 Vytvoření robota na Wikipedii

Pro řešení nejen vandalismů, ale i opravování chyb vytvořených méně zkušenými uživateli je dobré vytvářet roboty. Ti řeší konkrétní chyby přímo editací stránky. Pro běžného uživatele nejsou příliš dostupné informace o vytvoření použitelného robota a to ani na samotné Wikipedii. V následujícím textu se dozvíte jak vytvořit robota krok po kroku za použití DotNetWikipedia frameworku.

Jedna z mála informací, co vám Wikipedia o robotech sdělí, je potřeba schválení funkcí robota správci na Wikipedii. Roboti musí být hlavně co nejpečlivěji odladěni, pokud běží bez lidského dozoru. Činnost robota by se měla konzultovat nejlépe na stránce Pod Lipou. Pokud bude dělat větší množství editací, měli byste na základě souhlasu komunity požádat o příznak robota, který editace skryje v seznamu posledních změn. Příznak bota mohou udělovat byrokraté.

#### 5.1.1 Instalace MediaWiki 1.6.12

Každý robot by měl mít řádně otestované funkce, aby nenadělal více škody než-li užitku. Proto si pro zkušební provoz vašeho bota nainstalujte program MediaWiki, kde svého robota důkladně otestujte.

- MediaWiki si stáhněte na oficiálních stránkách. Dejte si pozor na podporovanou

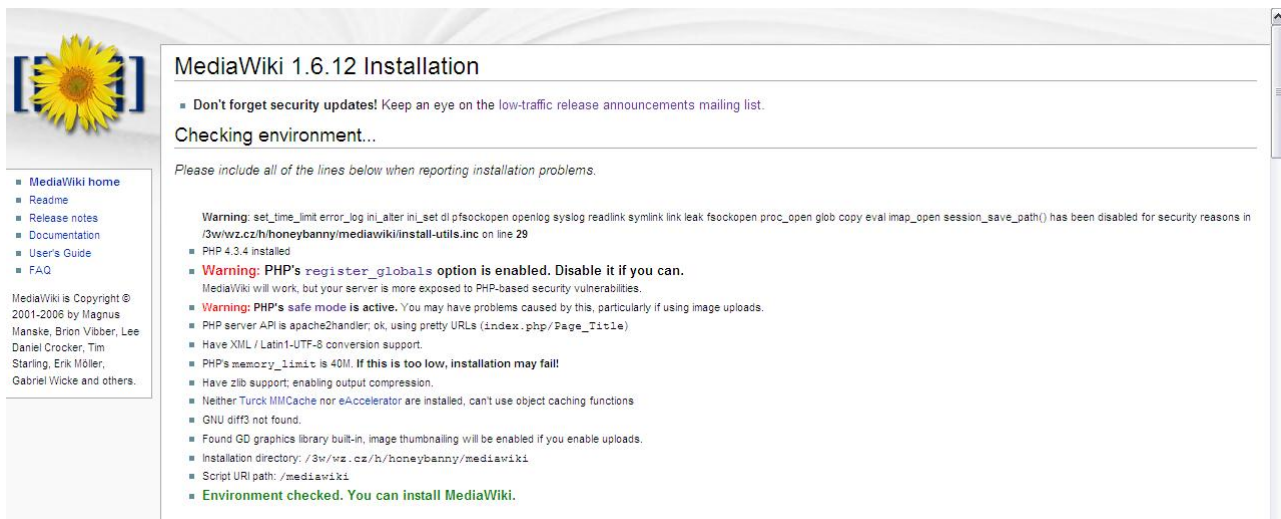
verzi PHP. Nejnovější verze 1.15.1 nelze spustit na PHP 5.3.1 pouze na 5.3.0 nebo 5.3.2. Pokud bude verze PHP nižší, stáhněte starší verzi MediaWiki. V tomto případě se MediaWiki instaluje na stránky serveru Webzdarma.cz s PHP 4, proto je použita verze MediaWiki 1.6.12.

- Stažený soubor je zazipovaný, rozbalte ho.
- V administračním systému si vytvořte novou databázi.
- Nyní naváš Web nakopírujte celý soubor MediaWiki.
- Do prohlížeče zadejte adresu kořenového adresáře a potvrďte. Tím spustíte titulní stranu a první krok instalace.



Obrázek 5.1: Instalace MediaWiki

- Stiskněte odkaz „setup the wiki“ pro načtení instalační stránky.
- K instalaci můžete přistoupit až poté, co zadáte požadované údaje. Stránka, kde na ně budete dotázáni, se zobrazí při správně nastaveném webu. To je potvrzeno hláškou „You can install MediaWiki“.



Obrázek 5.2: Instalační stránka MediaWiki

**Site config**

Wiki name:  **Must not be blank or "MediaWiki"**  
 Preferably a short word without punctuation, i.e. "Wikipedia".  
 Will appear as the namespace name for "meta" pages, and throughout the interface.

Contact e-mail:   
 Displayed to users in some error messages, used as the return address for password reminders, and used as the default sender address of e-mail notifications.

Language:    
 Select the language for your wiki's interface. Some localizations aren't fully complete. Unicode (UTF-8) used for all localizations.

Copyright/license:  No license metadata  
 GNU Free Documentation License 1.2 (Wikipedia-compatible)  
 A Creative Commons license - choose  
 A notice, icon, and machine-readable copyright metadata will be displayed for the license you pick.

Admin username:   
 Password:  **Must not be blank**  
 Password confirm:   
 An admin can lock/delete pages, block users from editing, and other maintenance tasks.  
 A new account will be added only when creating a new wiki database.

Shared memory caching:  No caching  
 Memcached

Obrázek 5.3: Instalační stránka MediaWiki

- Wiki name – zadejte jméno vaší wiki
- Contact email – váš kontaktní email (email administrátora)
- Language – vyberte si jazyk
- Copyright/license – zvolte „No license metadata“
- Admin username, Password a Password confirm – jméno, heslo administrátorského účtu a potvrzení hesla
- Shared memory caching – zvolte „No caching“
- Memcached servers – nemusíte vyplňovat

- E-mail (general) – emailové funkce, zvolte „Enabled“
  - User-to-user e-mail – povolení uživatelům komunikace přes emaily
  - E-mail notification – emailová notifikace změn na stránkách
  - E-mail authentication – při „Enabled“ povolíte odesílání autentifikačních emailů
  - Database type – typ použité databáze
  - SQL server host – název databázivého serveru
  - Database name – název vaší databáze
  - DB username, DB password a DB password confirm – název uživatele databáze, heslo a potvrzení hesla
  - Database table prefix – vyplňte v případě použití databáze pro více systémů najednou
  - Database charset – kódování databáze
  - Superuser account, Superuser password – účet superuživatele, nemusíte vyplňovat
- Správnost vyplněných údajů zkontrolujte a potvrďte „Install Mediawiki“
  - Pokud je instalace dokončena uvidíte „Installation successful!“
  - V adresáři config naleznete LocalSettings.php. Tento soubor přesuňte do kořenového adresáře.

Instalace by měla být dokončena. Vaše wiki stránky spustíte zadáním adresy webu a kořenového adresáře do prohlížeče.

Při instalaci se však může stát, že v jejím průběhu nastanou chyby. Řiďte se informacemi, co se zobrazí v horní části stránek.

V mém případě se při instalaci systému vyskytly problémy s vytvořením tabulek v databázi. Tento nedostatek jsem vyřešila zkopírováním MySQL příkazů ze souboru tables.sql ve složce maintenance/mysql5

Nyní můžete začít s naplněním vašeho wiki webu články, na kterých budete zkoušet funkce vašeho robota.

## 5.1.2 Popis funkce robota

Robot vyhledává chyby:

- V celé Wikipedii
- Na stránkách získaných z projektu Check Wikipedia

### Oprava chyb v celé Wikipedii

Nedostatky, které nejsou zahrnuty do projektu Check Wikipedia je potřeba vyhledávat na každé stránce. Jednou z těchto chyb je vnitřní odkaz shodný s jeho popisem. Interwikilink, jak se tento odkaz také nazývá, je označen z obou stran dvojitými hranatými závorkami. Popis odkazu se zobrazí na stránce jako text, jenž vás po kliknutí přesměruje na článek, nesoucí název odkazu, uložený na Wikipedii. Odkaz a popis jsou od sebe odděleni svislou čárou.

`[[automobil|automobil]]`

Takový zápis vnitřního odkazu je však nekorektní a měl by být nahrazen samotným odkazem bez popisku.

`[[automobil]]`

Obdobný problém nastává i pokud je odkaz kořenem popisku.

`[[automobil|automobilové]]`

V tomto případě se koncovka přidává až za ukončující hranaté závorky odkazu. Koncovka a odkaz nesmí být oddělené mezerou.

`[[automobil]]ové`

## WikiProjekt Check Wikipedia

Projekt Check Wikipedia se snaží o snížení počtu chyb na Wikipedii. Program prochází všechny stránky a vyhledává nedostatky způsobené uživateli, které je potřeba odstranit. Za celým tímto projektem stojí německý vystudovaný kartograf Stefan Kühn pracující jako lektor databází Oracle a žijící v Drážďanech. Ten objevil německou Wikipedii v roce 2003, kdy měla pouhých 9000 článků. Po čtyřech měsících spolupráce s Wikipedií se stal správcem a poté, až do roku 2008, byrokratem.[9]

Význam tohoto projektu spočívá ve vyhledání chyb robotem a jejich opravou uživateli. Nalezené nedostatky se dělí dle priority:

- Vysoká priorita
- Střední priorita
- Nízká priorita

## Oprava chyb nalezených projektem Check Wikipedia

Mnoho problémů nelze jednoduše opravit robotem, ale je potřeba individuálního přístupu a vyhodnocení celé situace.

Mezi chyby s nejvyšší prioritou patří i opakovaná reference. Pokud je stejná reference zapsána několikrát v článku, měla by se pojmenovat a dále zapisovat její zkrácenou verzi. Jméno reference musí nějakým způsobem popisovat referenci, což by bylo nejvhodnější vytvářet ručně. Proto program se bude snažit pojmenovat referenci prvním slovem za mezerou, které není číslo, neobsahuje znaky jako lomítka, tečky, čárky a podobné, a dále nesmí obsahovat řetězce "www." či http://.

```
<ref>"The 20th Century in Eretz Israel." s. 37</ref>
```

převédeme na



```
<ref name ="20th">"The 20th Century in Eretz Israel." s. 37</ref>
```

a dále používáme



<ref name="20th"/>.

Check Wikipedia takovou stránku vyhledá a vypíše tvar konkrétní reference.

article 	info 
Amalgám	<ref>[[Miroslav Šuta]]: [http://lhned.cz/2-16480620-000000_d-ff Evropská strategie eliminace rtuti], Odpady, 14. 7. 2005</ref>
Balčik	<ref>http://www.nsi.bg/otrasal.php?otr=19&a1=376&a2=377&a3=378#cont National Statistical Institute 2009</ref>
Bar Giora	<ref>"The 20th Century in Eretz Israel." s. 37</ref>
Gambie	<ref> Africa South of the Sahara 2003. 32nd Edition. London: Europa Publications, 2002. Str. 430. </ref>
Ha-Šomer	<ref>"The 20th Century in Eretz Israel." s. 86</ref>
Jaguar	<ref>(podle publikace "Jaguar (Kompletní historie od r. 1922 do současnosti)", Heiner Stertkamp)</ref>
Jay-Z	<ref>http://www.xximag.com/online/?p=2471</ref>
John Mayer	<ref>[http://www.grammy.com/GRAMMY_Awards/Winners/Results.aspx?title=&winner=John%20Mayer&year=0&genreID=0&hp=1 Oficiální web cen Grammy] (anglicky)</ref>
Kavarna	<ref>http://www.nsi.bg/otrasal.php?otr=19&a1=376&a2=377&a3=378#cont National Statistical Institute 2009</ref>

Obrázek 5.4: Opakovaná reference v projektu Check Wikipedia

Robot načte stránku, na které se reference nachází, její jméno udá jako konkrétní referenci, z možnosti důvodu kolize s jinými referencemi, a opakující se reference nahradí daným tvarem se jménem. Před spuštěním robota projekt Check Wikipedia našel přes 900 stránek s vícenásobnými referencemi.

Pokud je externí odkaz formátován jako interní, je uzavřen do dvojitých místo jednoduchých hranatých závorek, pak je to chyba se střední prioritou.

```
[[http://www.tour-eiffel.fr/teiffel/uk/ web]]
```

má být ve tvaru

```
[http://www.tour-eiffel.fr/teiffel/uk/ web]
```

Robot tento problém opravuje podobně jako u předešlé situace.

## 5.2 Návrh postupu rozšíření Wikipedie

### 5.2.1 Přístup k uživatelům

Wikipedie je založena na vytváření a úpravách článků uživateli. Čím více uživatelů přispívá, tím více článků Wikipedie obsahuje. Proto jsou obyčejní uživatelé nejdůležitější složkou celé Wikipedie.

Wikipedie je pátým nejpopulárnějším webem na světě a měsíčně ji navštíví přes 325 milionů lidí. Podle údajů z deníku The Wall Street Journal v poslední době Wikipedie přichází o své přispěvovatele a tím se snižuje korektnost celé Wikipedie. V roce 2009 opustilo řady uživatelů Wikipedie 49 tisíc lidí. O rok dříve to bylo pouhých 4900 přispěvovatelů. Nové články zakládá čím dál méně lidí a taktéž i úpravy, které jsou velmi potřebné pro přesnost a pravdivost článků nejsou v dostatečné míře. Tím klesá důvěryhodnost celé Wikipedie, která by měla být neustále zvyšována.

Wikipedie se snaží problém vyřešit i soutěží o nejlepší příspěvky, které budou odměněny věcnými cenami. To se však nejeví jako nejlepší řešení.

Problém dle mého názoru tkví v přístupu jak k nováčkům, tak i dlouhodobějším přispěvovatelům. Česká Wikipedie má nepřehledně popsané postupy, jak docílit napsání plnohodnotného článku. Pokud neumí nováček napsat infobox, bez něhož je článek pahýl, dá mu dost práce najít podrobnosti, jakým způsobem jej vytvořit a jakými informacemi naplnit. Proto informace o vytvoření korektního článku společně se syntaxem Wikipedie by měl být nováčkům dostupný nejlépe na jedné stránce.

Poté, co správce ohodnotí váš článek jako neúplný, NPOV (porušení nezaujatého úhlu pohledu) a dalšími označeními pro autora článku hanlivými, už toto hodnocení sám nesmaže. Nastává otázka, zda-li tuto šablonu smíte po úpravách smazat sami. Proto v české Wikipedii naleznete přinejmenším 60 tisíc pahýlů. Pokud se pokusíte diskutovat se správcem, který označení vaší stránce udělil, odpovědi se nedočkáte. Tento přístup správců by se měl jistě změnit. Pokud opravdu kontrolují každou editaci, jak správci tvrdí, pokud opětovně ukládáte tutéž stránku s novými úpravami, místo používání náhledů, měli by také šablony, které jsou již bez smyslu, odmazávat. Tím by docílili i vyšší důvěryhodnosti Wikipedie. Je těžké pojmenovat jako důvěryhodnou encyklopedii s 60 tisíci špatných článků.

Dalším problémem je vandalismus. Nejen nováčci jsou takto velmi často neprávem pojmenováváni. Pokud si uživatel všimne například nějakých dualit, špatného obsahu a podobných nepřesností, tyto údaje smaže, je ihned označen za vandala. Proto správce by si měl dobře rozmyslet, za co takové označení uživateli uděluje a především se seznámit se situací.

Uživatelé jsou lidé, co přispívají ze své dobré vůle a měli by být, jinak než materiálně, ohodnoceni. Každého poctivého autora přestanou bavit věčná falešná obvinění a raději se své dobrovolné práce vzdá. Pokud nezmění správci svůj neomalené a nekompromisní způsoby, alespoň na české Wikipedii, uživatelů bude neustále ubývat.

### 5.2.2 Podpora projektů na Wikipedii

Na Wikipedii existuje několik projektů pro její zkvalitnění. Běžný uživatel se však těchto projektů zúčastní minimálně a to z důvodu špatné dostupnosti. Tyto projekty se často schovávají za dlouhými názvy, málokde jsou publikovány a především vypadají jakoby se vytvářeli pouze pro správce nebo velmi zkušené uživatele.

Jako běžný uživatel jsem se snažila opravit chyby nalezené projektem Check Wikipedia. Chyby zde zveřejněné by se měli na příslušné stránce opravit a na stránce projektu upozornění na opravené chyby smazat. K mému překvapení jsem byla jednoduše označena za vandala. Samozřejmostí bylo, že jsem s těmito opravami skončila.

Správci mají za úkol zkontrolovat každou úpravu, proto řešením by mohl být seznam uživatelů opravujících chyby nalezené různými projekty a názvy článků, které by měly být dle projektu opraveny. Poté by měl každý správce povinnost vyhledat informace v tomto seznamu a porovnat je se skutečností.

### 5.2.3 Vyhledávání

Při hledání hesla jste velmi často odkázáni internetovým vyhledávačem na článek ve Wikipedii s příslušným tématem. Pokud však používáte vyhledávání přímo na Wikipedii můžete se setkat s několika problémy.

- Pokud hledáte např. šablony, projekty apod., které nejsou encyklopedickými hesly, nenajdete je v jiném tvaru než Wikipedie:Šablony a podobně.
- Při vyhledávání jsou výrazy hledány doslovně, takže slovo v jiném tvaru vyhledávač nenajde.

- Pokud hledáte maximálně den staré informace, nemusí se vám je podařit najít a to z výkonnostních důvodů.

Pro ulehčení práce s Wikipedií by určitě do budoucna jejímu vyhledávací prospělo alespoň přidání vyhledávání slov v jiném tvaru.

Dalším zlepšením, především pro nováčky, by byla možnost jednodušeji vytvořit novou stránku.

#### **5.2.4 Doplnění pahýlů**

Důvěryhodnost české Wikipedie závisí především na kvalitě článků. Proto by bylo velmi potřebné soustředit síly editorů na články, které jsou označené jako pahýly. Přes 68 tisíc nedostačujících informací v encyklopedii se 162 tisíci hesly nepůsobí na její uživatele příznivě. Wikipedie z tohoto pohledu působí jako encyklopedie, která uvádí obecné, neúplné, nebo nepravdivé informace o hledaném hesle. Proto se lze velmi často setkat s odepřením Wikipedie jako zdroje informací pro školní projekty a práce. Většinou jsou požadovány informace z důvěryhodnějších zdrojů sestavených odborníkem v oboru.

# Závěr

Prostudování celé Wikipedie a zhodnocení každého článku z více jak 150 tisíc neustále se měnících článků je pro jednotlivce v podstatě nemožné. Při posuzování tohoto projektu jsem čerpala především z vlastních zkušeností, ale nemalou pomocí mi byly i názory jiných uživatelů a také správců.

Použitelnost článků z Wikipedie se liší dle záměru uživatele. Pokud nepotřebujete podrobné vysvětlení hesel, nebo nepoužíváte tyto poznatky jako ověřený zdroj informací, např. k vytvoření odborné práce, články vám téma dostatečně vysvětlí. Pokud se vám zdá popis hesla nepravdivý, nebo nemožný, můžete se vždy obrátit na uživatele diskutující Pod Lípou a veřejně se o svůj názor podělit. Tímto krokem podporujete kvalitu informací na Wikipedii.

U takto otevřených projektů je vždy potřeba počítat s možností, že heslo je záměrně vysvětleno špatně, nebo není dostatek přispívajících odborníků, kteří by byly schopni dostatečného popisu. Raději si při psaní odborných textů informace z Wikipedie ověřte nejlépe v knižních publikacích.

Přestože se česká Wikipedie neustále rozšiřuje, vyvíjí a počet nových článků roste přibližně o 2 tisíce měsíčně, zdaleka svou kvalitou neodpovídá anglické verzi. Velmi vysoký a neklesající počet pahýlů českou verzi znehodnocuje.

K vylepšení tohoto projektu přispívá vytváření potřebných robotů. Ti jsou nemalým přínosem z důvodu opravování velkého počtu stránek. Tímto způsobem dostává encyklopedie jednotný tvar článků, čímž se zjednodušuje její čtení i editování.

Zápornou stránkou Wikipedie se pro mnohé, ne příliš zkušené uživatele, stali její správci. Mnoho wiki webů v základních podmínkách pro komunikaci s uživatelem říká, že se k němu má vždy přistupovat vstřícně a se snahou pomoci. Nikdy se

nemá označovat za vandala, před vysvětlením jeho záměru. Pro wiki weby je každý uživatel přínosem. Ve Wikipedii tento názor zastává nejspíše jen omezené množství správců. Většina z nich vaše problémy neřeší a málokdy odpoví na vysvětlení vašeho počínání. Co je v menších wiki samozřejmostí ve Wikipedii chybí.

Aby se Wikipedie zdárně rozrůstala, je potřeba vážit si uživatelů a práce, kterou pro ni dělají. Bez obyčejných lidí by Wikipedie nikdy nedosáhla takového obsahu a popularity.

# Seznam použitých zdrojů

- [1] *Wikipedie*[online]. 2003 [cit. 1.1.2010].  
Dostupné na: <<http://cs.wikipedia.org/wiki/Wikipedie>>
- [2] *Wikipedia:Wikipedia articles per population*[online]. 2005 [cit. 2.1.2010]  
Dostupné na: <[http://en.wikipedia.org/wiki/Wikipedia:Wikipedia_articles_per_population](http://en.wikipedia.org/wiki/Wikipedia:Wikipedia_articles_per_population)>
- [3] *Wikipedie:Historie české Wikipedie*[online]. 2005 [cit. 1.1.2010]  
Dostupné na: <[http://cs.wikipedia.org/wiki/Wikipedie:Historie_Wikipedie#Historie_spr.C3.A1vc.C5.AF_Wikipedie](http://cs.wikipedia.org/wiki/Wikipedie:Historie_Wikipedie#Historie_spr.C3.A1vc.C5.AF_Wikipedie)>
- [4] *MediaWiki* [online]. 2004 [cit. 1.1.2010].  
Dostupné na: <<http://cs.wikipedia.org/wiki/MediaWiki>>.
- [5] *Wikipedie:Jak vytvořit článek* [online]. 2002 [cit. 3.1.2010].  
Dostupné na: <[http://cs.wikipedia.org/wiki/Wikipedie:Jak_vytvo%C5%99it_%C4%8D1%C3%A1nek](http://cs.wikipedia.org/wiki/Wikipedie:Jak_vytvo%C5%99it_%C4%8D1%C3%A1nek)>.
- [6] *Wiki* [online]. 2004 [cit. 1.1.2010].  
Dostupné na: <<http://cs.wikipedia.org/wiki/Wiki>>.
- [7] *Knol* [online]. 2009 [cit. 3.1.2010].  
Dostupné na: <<http://knol.google.com/k>>.
- [8] Jan Kapoun *Jimmy Wales: Zakladatel Wikipedie* [online]. 2006 [cit. 3.1.2010].  
Dostupné na: <<http://scienceworld.cz/technologie/jimmy-wales-zakladatel-wikipedia-1401>>.
- [9] *Benutzer:Stefan Kühn* [online]. 2003 [cit. 4.1.2010].  
Dostupné na: <[http://de.wikipedia.org/wiki/Benutzer:Stefan_K%C3%BChn](http://de.wikipedia.org/wiki/Benutzer:Stefan_K%C3%BChn)>.

- [10] *Wikipedie:WikiProjekt Kvalita/Kritéria* [online]. 2007 [cit. 15.1.2010].  
Dostupné na: <[http://cs.wikipedia.org/wiki/Wikipedie:WikiProjekt_Kvalita/Krit%C3%A9ria](http://cs.wikipedia.org/wiki/Wikipedie:WikiProjekt_Kvalita/Krit%C3%A9ria)>.
- [11] *Wikipedia Statistics Czech* [online]. 2010 [cit. 20.3.2010].  
Dostupné na: <<http://stats.wikimedia.org/EN/TablesWikipediaCS.htm>>.
- [12] *Wikipedie:Vandalismus* [online]. 2005 [cit. 20.3.2010].  
Dostupné na: <<http://cs.wikipedia.org/wiki/Wikipedie:Vandalismus>>.
- [13] *Standard ECMA-334* [online]. [cit. 18.4.2010].  
Dostupné na: <<http://www.ecma-international.org/publications/standards/Ecma-334.htm>>.



# Obsah CD

Přiložené CD obsahuje:

- Diplomova_prace.pdf – Text diplomové práce
- Zdrojové kódy – Adresář se zdrojovými kódy k programům použitých v práci
  - RobotWikipedie – opravuje reference, špatně zadané externí odkazy a vnitřní odkazy
  - DPL1 – vyhledává poměr slov a vnitřních odkazů, počítá počet pahýlů na anglické a české Wikipedii
  - DPL-temata – vyhledává pod jaké portály patří dané stránky
  - Informace_o_programech.txt – informace k prvnímu použití programů
- DotNetWikiBot_2.91 – DotNetWikibot framework