

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE

PROVOZNĚ EKONOMICKÁ FAKULTA

Katedra informačního inženýrství
Obor Informatika



DIPLOMOVÁ PRÁCE

**POČÍTAČOVÝ PŘEKLAD
MEZI ŽIVÝMI JAZYKY**

Vedoucí diplomové práce:
Doc. Ing. Vojtěch Merunka, Ph.D.

Autor:
Bc. Martin Bartoš © 2012

Česká zemědělská univerzita v Praze

Provozně ekonomická fakulta

Katedra informačního inženýrství

Akademický rok 2009/2010

ZADÁNÍ DIPLOMOVÉ PRÁCE

Martin Bartoš

obor Informatika

Vedoucí katedry Vám ve smyslu Studijního a zkušebního řádu ČZU v Praze
čl. 17 odst. 2 určuje tuto diplomovou práci.

Název práce: **Počítačový překlad mezi živými jazyky.**

Osnova diplomové práce:

1. Úvod
2. Cíl práce a metodika
3. Literární rešerše
4. Výběr varianty řešení
5. Projekt
6. Závěr
7. Seznam použitých zdrojů
8. Přílohy

Rozsah hlavní textové části: 60 - 80 stran

Doporučené zdroje:

PHP5, MySQL, Apache : Vytváříme webové aplikace. [s.l.] : Computer Press, 2006. 816 s. ISBN 80-251-1073-7.

GUTMANS, Andi, SAETHER BAKKEN, Stig , RETHANS, Derick. Mistrovství v PHP 5 . Computer Press : [s.n.], 2007. 656 s. ISBN 978-80-251-1519-0.

DARWIN, Ian F. . Java : Kuchařka programátora. [s.l.] : Computer Press, 2006. 800 s. ISBN 80-251-0944-5.

HOPKINS, Trevor, HORAN, Bernard. Smalltalk : An Introduction to Application Development Using VisualWorks (Textbook Binding). [s.l.] : Prentice Hall, 1995. 400 s. ISBN 0133183874.

Vedoucí diplomové práce: **doc. Ing. Vojtěch Merunka, Ph.D.**

Termín odevzdání diplomové práce: duben 2011


.....
Vedoucí katedry




.....
Děkan

Čestné prohlášení

Prohlašuji, že svou diplomovou práci „Počítačový překlad mezi živými jazyky“ jsem vypracoval samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou citovány v práci a uvedeny v seznamu literatury na konci práce. Jako autor uvedené diplomové práce prohlašuji, že jsem v souvislosti s jejím vytvořením neporušil autorská práva třetích osob.

V Praze, dne 04. 04. 2012

Poděkování

Rád bych touto cestou poděkoval panu doc. Ing. Vojtěchu Merunkovi, PhD. za vedení, konzultace a pomoci při tvorbě této diplomové práce.

Počítačový překlad mezi živými jazyky

Souhrn

Tato diplomová práce se zaměřuje na vývoj webu a webových technologií jako prostředku pro vývoj webové služby. Dále se zaměřuje na lingvistiku a zpracování přirozeného jazyka počítačem. Pro tento účel byla vytvořena aplikace, která demonstruje na příkladu jednoduché zpracování jazyka v podobě transliterace znaků z jedné abecedy do druhé.

Klíčová slova

Web, HTML, transliterace, lingvistika, jazyk, překlad

Computer translation between spoken languages

Summary

There thesis are focused on the development of new web applications while also describing the progress of the web technologies and the web itself. Consequently, it focuses on linguistics and a spoken language processing by a computer. To support this, a web application was created, which demonstrates a simple spoken language processing in the form of a character transliteration from one alphabet to another on a specific example.

Keywords

Web, HTML, transliteration, linguistics, language, translation

Obsah

1	Úvod.....	6
2	Cíl diplomové práce	8
3	Metodika a postup řešení	9
	3.1 Popis obsahu rešerše.....	9
4	Seznam použitých zkratek	11

Část I Literární rešerše

5	World Wide Web	14
	5.1 Historie World Wide Web	15
	5.2 Funkčnost WWW	16
	5.3 Linkování.....	17
	5.4 Prefix www	18
	5.5 Bezpečnost.....	19
	5.6 Webové standardy	19
	5.7 DOM.....	20
	5.7.1 Historie	20

5.7.2	Legacy DOM.....	20
5.7.3	Intermediate DOM.....	21
5.7.4	Standardizace.....	21
5.7.5	Využití DOM u webových prohlížečů	22
5.7.6	Layout engines.....	23
6	Web 2.0	24
6.1	Web jako platforma.....	26
6.2	Kolektivní spolupráce	28
6.2.1	Blogging.....	29
6.3	Datová základna.....	30
6.4	Konec release fází u vývoje softwaru.....	31
6.4.1	Služby na webu nahrazují desktopové aplikace	32
6.4.2	Příklady aplikací Web 2.0	34
6.5	Jednoduché programovací modely	39
6.6	Bohaté uživatelské schopnosti	40
7	Lingvistika a přirozený jazyk.....	42
7.1	Rozdělení lingvistiky	42
7.2	Přirozený jazyk.....	44
8	Počítačové zpracování živých jazyků.....	45
8.1	Morfologie.....	45

8.2 Korpus	47
8.2.1 Vznik korpusu	48
8.3 Korpusové nástroje	51
8.3.1 Nutnost standardizace.....	52
8.4 Strojový překlad	52
8.4.1 Historie překladu	53
8.4.2 Princip překladu.....	54
8.5 Transliterace.....	56
8.5.1 Užití transliterace	57
8.5.2 Varianty transliterace	57
8.5.3 Pravidla pro transliteraci.....	57
8.5.4 ČSN ISO 9.....	58
8.5.5 GOST 16876-71.....	59
9 Písma použítá v praktické části.....	60
9.1 Latinka	62
9.2 Alfabet.....	63
9.3 Cyrilice.....	64
9.4 Mezinárodní fonetická abeceda	66
9.5 Užití IPA	67
9.6 Znaký IPA	68

Část II Praktická část

10 Praktické využití transliterace	70
10.1 Popis řešení	70
10.1.1 Hlavní soubor - index.php.....	71
10.1.2 Soubor s formulářem - head.php.....	72
10.1.3 Soubor pro manipulaci s dokumentem a transliteraci - main.php.....	74
10.1.4 Soubor s transliterační tabulkou – transtab.sql.....	81
10.1.5 Soubor se skriptem pro připojení k databázi - db.php.....	82
10.1.6 Soubor s formátováním stránky – theme.css	83
10.1.7 Ukázka výsledné podoby stránek po transliteraci.	84
11 Závěr	86
12 Použitá literatura	88
13 Seznam obrázků	94

Část III Přílohy

Úvod

Tématem této práce je počítačové zpracování živých, neboli, mluvených jazyků s využitím webových technologií. S webovými službami a zpracováním jazyka se prakticky setkáváme neustále, například při vyhledávání informací na Internetu, slovnících, encyklopediích, strojovém překladu v aplikaci Google Translator, opravách chyb v textu apod.

Přirozený jazyk se dá chápat jako prostředek ke komunikaci. Od umělého jazyka se liší v mnoha směrech. Živý jazyk jde s dobou, neustále se vyvíjí, obsahuje jisté nejednoznačnosti a další a proto oproti umělému jazyku, který je v zásadě neměnný, dál se nevyvíjí, je těžké přirozený jazyk formalizovat, aplikovat na něj algoritmy pro zpracování. Proto je počítačový překlad nesmírně složitá a komplexní záležitost. Aby mohlo docházet ke zpracování jazyka, je nejprve nutné čerpat poznatky z řad mnoha oborů zabývajících se zkoumáním přirozeného jazyka neboli lingvistikou. Samotný obor lingvistika se dělí na desítky dalších podoborů. Při analýze živého jazyka se nejčastěji využívají poznatky z morfologie, syntaktiky nebo sémantiky. Cílem zkoumání počítačového zpracování jazyka je zautomatizování jednotlivých procesů.

Původně Web vznikl za účelem propojení statických dokumentů neboli webových stránek. Časem se stránky začaly generovat dynamicky s využitím

jazyka PHP, což dalo základ vzniku interaktivních webových aplikací. S využitím webových technologií je možné komunikovat mezi různými aplikacemi, které se si mezi sebou přenášejí data. Aplikace komunikují mezi sebou pomocí webových služeb. Čili hovoří se zde o aplikacích, které nabízejí nějaké služby.

Webová služba představuje technologii, která dokáže integrovat aplikace, které se vyskytují na různých platformách a lze je ovládat pomocí webového rozhraní, typicky pomocí internetového prohlížeče. Služby dále představují nástroj pro zefektivnění business procesů v rámci organizace. Základem úspěchu je jazyk HTML, který je nezávislý na platformě. Tudíž představuje nejvhodnější základ pro řešení webových služeb, pomocí kterého je možné propojovat mezi sebou zdánlivě nepropojitelné aplikace.

Cíl diplomové práce

Cílem předkládané diplomové práce je seznámení se s problematikou automatického překladač webových stránek a tvorbou webových aplikací.

V rámci toho cíle bude vytvořena webová služba, která bude umět transliterovat obsah webové stránky mezi třemi abecedami: latinkou, cyrilicí a řeckou abecedou podobným způsobem, jako pracuje Google Translate. Výsledná aplikace bude dostupná na Internetu.

Metodika a postup řešení

K vypracování diplomové práce byl použit sběr informací z dostupných informačních zdrojů, jejich následná analýza a zobecnění.

Pro realizaci cíle bylo seznámení s technologií Google, programováním webových stránek, problematikou překladu a transliterace.

S použitím aplikací PHPStorm, PHPMyAdmin, databázového prostředí MySQL, skriptovacího jazyka PHP, HTML, CSS a SQL byla vytvořena webová aplikace, která demonstruje praktické využití transliterace na webu.

3.1 Popis obsahu diplomové práce

První část literární rešerše je zaměřena na World Wide Web jehož první návrh vytvořil v roce 1989 Tim Berns Lee v rámci projektu ENQUIRE, dále na vývoj a použití webových technologií pro tvorbu webu.

Další kapitola se orientuje na etapu vývoje Web 2.0 a na nový způsob vývoje webových aplikací a také na rozdíly ohledně způsobu využití webu uživateli. V závěru kapitoly jsou uvedeny příklady jednotlivých Web 2.0 aplikací.

Následující kapitola pojednává o oboru lingvistika, jejím rozdělení do jednotlivých podoborů a dále se zaměřuje na přirozený jazyk.

Další část práce se zaměřuje na počítačové zpracování živého jazyka a oblast morfologie, která je pro počítačové zpracování důležitá. Tato část práce se také zabývá vznikem a využitím korpusů u elektronických slovníků. Dále se zabývá strojovým překladem, jeho vznikem a rozdíly v použitých technikách strojového překladu. Závěr kapitoly je věnován transliteraci, jejími normami, které se využívají při konverzi znaků z jedné abecedy do druhé.

Závěr literární rešerše je orientován na vznik a vývoj abeced, které jsou využity v praktické části.

Druhá část diplomové práce se zabývá vývojem webové služby pro transliteraci obsahu webové stránky.

4

Seznam použitých zkratk

WWW	<i>(World Wide Web)</i>
URL	<i>(Uniform Resource Locator)</i> . Slouží pro jednoznačnou identifikaci zdroje na Internetu.
DNS	<i>(Domain Name System)</i> . Jedná se o hierarchický systém doménových jmen na Internetu.
HTTP	<i>(Hyper Text Transfer Protocol)</i>
IP	<i>(Internet Protocol)</i>
CSS	<i>(Cascading Style Sheets)</i> . Jde o jazyk, který se používá pro popis prezentační sémantiky u dokumentu psaného ve značkovacím jazyku.
CNAME	<i>(Canonical NAME record)</i> . Slouží pro označení, že daná doména je aliasem domény jiné.
POP3	<i>(Post Office Protocol)</i>
RSS	<i>(RDF Site Summary, Really Simple Syndication)</i> . Jedná se formát nebo skupinu XML formátů, které se na Internetu využívají k publikaci často aktualizovaných článků.
IMAP	<i>(Internet Message Access Protocol)</i> . Protokol k přístupu k mailové schránce, který umožňuje práci v off-line i online režimu.

TLS	<i>(Transport Layer Security)</i> . Zabezpečený protokol pro komunikaci na Internetu. Předchůdce SSL.
AJAX	<i>(Asynchronous JavaScript and XML)</i> . Skupina navzájem propojených technik na straně client-side pro vytvoření asynchronních webových aplikací.
SOAP	<i>(Simple Object Access Protocol)</i> . Základní protokol webových služeb. Využívá se k posílání XML zpráv.
XML	<i>(eXtensible Markup Language)</i> . Značkovací jazyk, který definuje sadu pravidel pro kódování dokumentů ve formátu, který je strojově i lidsky čitelný.
B2B	<i>(Business to Business)</i>
REST	<i>(Representation State Transfer)</i> . Architektura rozhraní pro distribuovaná prostředí.

Část I

Literární rešerše

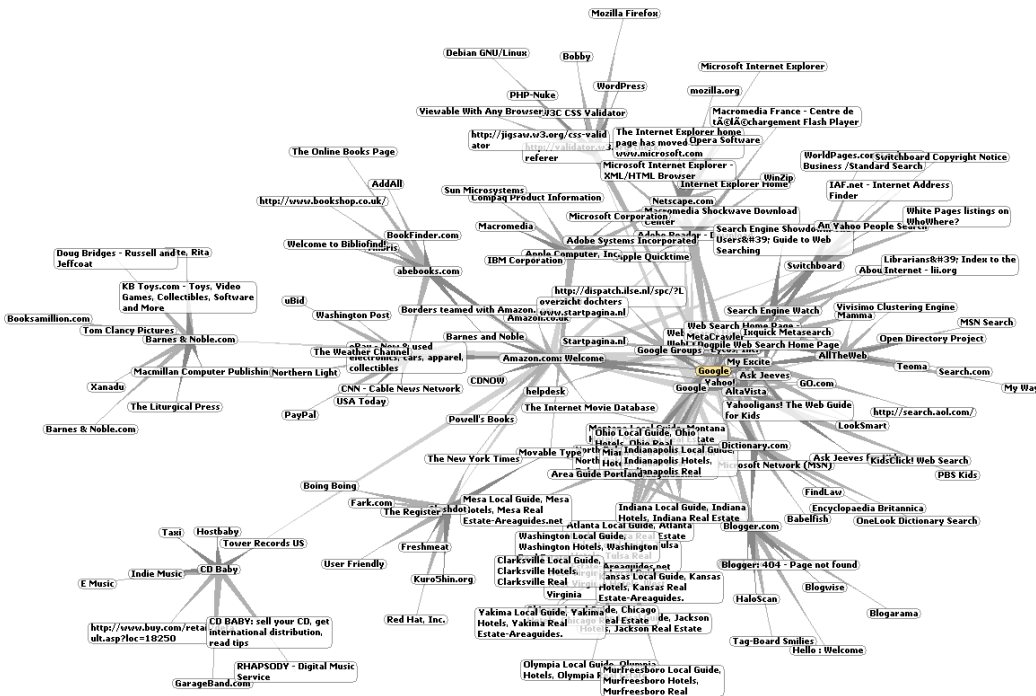
World Wide Web

WWW ve volném překladu znamená „celosvětovou pavučinu“. V podstatě jde o systém navzájem prolinkovaných dokumentů zpřístupněných na Internetu. Webem lze označit i soustavu dokumentů dostupných na jednom webovém serveru, které jsou adresovány pomocí URL.



Obrázek 1: Historické logo

WWW [Robert Cailiau]



Obrázek 2: Grafické znázornění webu podle [Examiner.com]

5.1 Historie World Wide Web

První návrh WWW vznikl už v roce 1989, který vycházel z projektu ENQUIRE¹ na němž se podílel v CERNu Tim Berns Lee. ^[1] S pomocí Roberta Cailliau publikoval řadu návrhů týkající se „Hypertext project“ neboli „World Wide Web“ nebo také „W3“, kdy s využitím webového prohlížeče a client-server architektury mohly být zobrazeny hypertextem provázané dokumenty. ^[2] Návrh odhadoval, že tvorba read-only webu bude trvat tři měsíce a dalších šest měsíců, až uživatelé budou moci na web přidávat odkazy a dokumenty. Pro čtenáře se uvažovalo vyvinout notifikaci při přidání článků s tématikou, o kterou se zajímali. Zatímco vyvinout read-only web se podařil, další fáze nikoliv. Objevila se až v konceptech wiki, Web 2.0 a RSS.

První člověk na světě, který použil webserver byl Berners-Lee na stroji NeXT Computer² v roce 1990. V témže roce byl také vyvinut první browser, WorldWideWeb³. ^[3]

¹ Softwarový projekt, který vytvořil Tim Berners-Lee v roce 1980 v CERNu. Tento projekt byl považován za předchůdce World Wide Web. Jednalo se o jednoduchý hypertextový program, který sloužil tamním vědcům pro snadnější přístup k informacím.

² Jednalo se o high-end workstation, který vyráběla a prodávala firma NeXT Stevena Jobse v letech 1988 – 1990.

³ Prohlížeč byl schopný zobrazovat základní styly, stahovat a otevírat soubory podporující NeXT systém, dále procházet různé diskuzní skupiny a kontrolovat pravopis. První obrázky se zobrazovaly v oddělených oknech. Prohlížeč fungoval i jako WYSIWYG editor.



Obrázek 3: NeXT Computer [Cern.ch]

Zásadní zvrát přinesl 6. srpen roku 1991, kdy se Web dostal na veřejnost jako volně dostupná služba na Internetu. O rok později se na celosvětové síti poprvé objevila fotografie. Fotografie zachycovala budovy CERNU, Les Cernettes, kterou zveřejnil Berners-Lee. ^[4]

World Wide Web se odlišoval od již existujících hypertextových systémů. Web vyžadoval pouze jednosměrný link, což umožnilo odkázat se na zdroj bez nutnosti zásahu vlastníka konkrétního zdroje. ^[5]

Po opuštění CERNU v roce 1994, Tim Berners-Lee založil World Wide Web Consortium (W3C). Podnět pro vznik W3C byly odlišné verze HTML, které nabízeli různí dodavatelé, což způsobovalo nekonzistenci mezi webovými stránkami. Cílem této organizace bylo zajištění kompatibility a jednotnosti mezi členy při přijímání nových standardů.

5.2 Funkčnost WWW

Aby se webová stránka zobrazila, je třeba napsat její URL do browseru nebo kliknout na link na stránce. Webový prohlížeč poté iniciuje řadu zpráv pro

zobrazení webové stránky. Nejprve prohlížeč přeloží URL adresu za pomoci DNS do IP adresy. Browser poté vyšle požadavek na zdroj odesláním HTTP žádosti na konkrétní adresu počítače. Počítač, který přijme HTTP požadavek, doručí žádost na web server, který naslouchá na portu 80. Ten požadavek vyřídí a poté pošle odpověď počítači, který vznesl požadavek, dále následuje obsah požadované stránky.

Prohlížeč parsuje HTML, interpretuje značky, jako např. `<title>`, `` Webové stránky často obsahují mnoho dalších HTML tagů, obrázků, multimediální obsah, skripty, které ovlivňují chování stránky nebo CSS styly ovlivňující layout stránky.^[6]

5.3 Linkování

Struktura webu je popsána pomocí webgrafu, kde uzly grafu představují jednotlivé webové stránky, orientované hrany zase hypertextové odkazy.

S postupem času zdroje, na které existují hyperlinky, mizí, mění svojí lokaci nebo jsou nahrazeny jiným obsahem. Takovéto linky zastarávají a stávají se z nich „dead links“.

V roce 1996 vznikl Internetový archiv, který se snaží předejít neustálým změnám na webu tím, že část obsahu webu archivuje.^[6]

5.4 Prefix www

Z dlouhodobé praxe se u doménových jmen pro World Wide Web používá prefix *www*, podle serverů, které tyto služby poskytují. Hostname pro web server je často *www*, *ftp* zase pro FTP server atd. Použití *www* jako jméno subdomény nevyžaduje žádné technické či politické normy, např. název prvního webserveru byl *nxoc01.cern.ch*, díky tomu řada webových stránek existovala bez prefixu *www*. Velká řada websites stále používá prefixu *www* nebo si prefixy vymýšlí vlastní, jako např. *www2* atd.^[7]

Pro vyvažování zátěže příchozího webového provozu se používá názvu subdomény, kdy se vytvoří CNAME záznam⁴, který odkazuje na cluster webového serveru.

HTTP nebo HTTPS odkazuje na Hypertext Transfer Protokol nebo HTTP Secure, pro příjem žádostí nebo odesílání odpovědí. Základem fungování World Wide Web je protokol HTTP, přidaná šifrovací vrstva v HTTPS je nutná u důvěryhodných informací, jako jsou hesla nebo bankovní informace, které mají být vyměněny v prostředí veřejného Internetu.

U aplikací typu Web 2.0 nebo webových aplikací pro mobilní zařízení se často již prefixu *www* nevyužívá, jedná se o aplikace např. *Gmail.com*, *Facebook.com* nebo *Twitter.com*.^[8]

⁴ Je typ záznamu v DNS, který specifikuje, že doménové jméno je alias jiného. Pomáhá to při běhu více služeb, které běží na jiných portech u jedné IP adresy. Každá služba může pak mít svůj vlastní záznam v DNS.

5.5 Bezpečnost

Web se stává prostředím i pro šíření malware a je také prostředkem pro šíření počítačové kriminality, která zahrnuje např. krádeže identity, podvody, špionáž, shromažďování zpravodajských informací. Nejběžnější hrozbou je SQL injection⁵ mířená proti websites. HTML a URL postihuje útok zvaný cross-site scripting⁶ (XSS), který přišel s nástupem JavaScriptu. Dle odhadu může být až 70% internetových stránek vystaveno XSS útokům.^[9]

5.6 Webové standardy

Velké množství technických specifikací a formálních norem týkající se WWW pochází z dílny World Wide Web Consorcia (W3C), některé jsou počínem Internet Engineering Task Force (IETF) a dalších organizací. Technické specifikace a normy se zejména týkají:

- doporučení pro značkovací jazyky, zejména HTML a XHTML (W3C),
- doporučení pro styly, zejména CSS (W3C),
- standardů pro ECMAScript, obvykle pro JavaScript (ECMA International),
- doporučení pro Document Object Model (ECMA).

⁵ Technika, která zneužívá chybu v zabezpečení na databázové vrstvě

⁶ Útočníci pomocí toho útoku obcházejí kontrolu přístupu

Další publikace definují základní technologie pro World Wide Web:

- Uniform Resource Identifier (URI) – univerzální systém pro odkazování se na zdroje na internetu, jako na hypertext documents a obrázky. URIs, často zvané URLs jsou definovány dle organizace v IETF RFC 3986/STD 66.
- HyperText Transfer Protocol (HTTP), zvláště jako definice dle RFC 2616.

[10] [11]

5.7 DOM

Document Object Model, neboli objektový model dokumentu, je multiplatformní a jazykově nezávislá konvence pro reprezentaci a interakci objektů v HTML, XHTML a XML dokumentů. DOM je v zásadě API, které umožňuje modifikaci obsahu, úpravu stylu dokumentu. [12]

5.7.1 Historie

Historie DOM je spojena s „válkou prohlížečů“, která probíhala na konci roku 1990 mezi Netscape Navigator a Microsoft Internet Explorerem, obdobně jako mezi JavaScriptem a JScriptem, skriptovacími jazyky webových prohlížečů v layout engine. [13]

5.7.2 Legacy DOM

Javascript vyvinula firma Netscape Communications v roce 1996. Jako první prohlížeč, kde jádro prohlížeče bylo založeno na Javascriptu byl Netscape Navigator 2.0. Microsoft, konkurent Netscapu, vydal ve stejný rok Internet Explorer 3.0, kde jádro prohlížeče bylo také založeno na skriptovacím jazyku

s názvem JScript. S tímto rozšířením umožnili webový vývojáři uživatelům interaktivní práci s webovými stránkami.

U Legacy DOMu byly přístupné pouze určité druhy elementů. Na formuláře, odkazy, obrazové prvky se odkazovalo prostřednictvím hierarchického názvu, který začínal stejně jako kořenový objekt dokumentu. ^[14]

5.7.3 Intermediate DOM

V roce 1997 vyšly nové verze prohlížečů Netscape Navigator a Internet Exploreru 4.0, které navíc podporovaly dynamické HTML, funkce umožňující změnu načteného HTML. Ačkoliv implementace Legacy DOM byla do značné míry kompatibilní, protože JScript⁷ byl založen na JavaScriptu, rozšíření DHTML DOM byly vyvíjeny paralelně u každého výrobce zvlášť, v důsledku pak byl DHTML DOM navzájem nekompatibilní. ^[13]

Intermediate DOM umožňoval manipulaci s CSS. Také umožňoval přístup k novým doplňkům, zvaným „layers“ přes vlastnost „document.layers“ u Netscape Navigator a „document.all“ u Microsoft Internet Exploreru. Zásadním nedostatkem u Intermediate DOM byla nekompatibilita mezi prohlížeči, proto při vývoji prohlížečů bylo požadováno speciální handlování pro každý podporovaný prohlížeč. ^[13]

5.7.4 Standardizace

Netscape Communications společně s Microsoftem a dalšími společnostmi přinesly základ pro vytvoření standardu pro prohlížeč skriptovacích jazyků, tzv.

⁷ Skriptovací jazyk založen na ECMAScript standardu, který je využit v Internet Exploreru. JScript je implementován jako Windows Engine.

ECMAScript⁸. První verze toho standardu byla publikována v roce 1997. Následující verze JavaScriptu a JScriptu v rámci standardu ECMAScript přinesla větší kompatibilitu mezi prohlížeči. ^[13] ^[15]

Po vydání ECMAScriptu W3C začala pracovat na standardizaci DOM. Počáteční standard, známý pod označením „DOM Level 1“ vyšel v roce 1998. Zhruba ve stejné době vyšel Internet Explorer ve verzi 5, který částečně podporoval tento první standard. ^[15]

Na začátku roku 2000 vyšel další standard jako DOM Level 2, který přinesl nové funkce, jako getElementById, nebo podporu XML namespaces a CSS. ^[16] DOM Level 3, aktuální standard, byl zveřejněn v dubnu 2004, kde byla přidána například podpora pro XPath nebo zachytávání událostí z klávesnice ^[17].

Koncem roku 2005 velkou část W3C DOMu podporovaly běžné prohlížeče podporující ECMAScript, jednalo se například o Microsoft Internet Explorer ve verzi 6, Operu, Safari a prohlížeče založené na Gecku (Mozilla, Firefox, SeaMonkey a Camino).

5.7.5 Využití DOM u webových prohlížečů

Webový prohlížeč není povinen používat DOM pro zobrazení HTML dokumentu. DOM slouží hlavně k tomu, aby bylo možné vytvářet na webových stránkách dynamický obsah. Obecně lze říct, že DOM určuje způsob práce webového prohlížeče s JavaScriptem.

⁸ skriptovací jazyk standardizovaný ECMA International v ECMA-262 specifikaci a ISO/IEC 16262. Jazyk je široce využíván pro skriptování na straně klienta. Nejvíce je znám pod různými dialekty, jako JavaScript, JScript nebo ActionScript.

5.7.6 Layout engines

Pro parsování HTML na DOM webové prohlížeče využívají layout engines. Například layout engines Trident/MSHTML a Presto využívá Internet Explorer nebo Opera. Prohlížeče jako Safari, Google Chrome, Firefox jsou založeny na Webkitu a Gecku.^[18]

Web 2.0

Termín Web 2.0 je často spojován s webovými aplikacemi, které umožňují participativně sdílet informace, změnit design dle představ uživatele. ^[19] Web 2.0 umožňuje uživatelům interakci a spolupráci s ostatními lidmi v rámci sociální sítě na rozdíl od webových stránek, kde si uživatelé, spotřebitelé, mohou pouze prohlížet obsah webu. Příkladem využití Web 2.0 mohou být sociální sítě, blogy, wiki, sdílení videí, hostované služby, webové aplikace nebo mashups.

Termín Web 2.0 je spojen s Tim O'Reillyem a jím pořádanou konferencí v roce 2004, kde tento termín byl představen. ^[20] Nová verze World Wide Web neznamena změnu v technické specifikaci, ale spíše změnu ve způsobu vývoje webových aplikací a jejich následném použití konečnými uživateli. Jak ve svém článku Tim O'Reilly píše, Web 2.0 je hlavně o změně přístupu: „*Web 2.0 is the business revolution in the computer industry caused by the move to the internet as platform, and an attempt to understand the rules for success on that new platform. Chief among those rules is this: Build applications that harness network effects to get better the more people use them. (This is what I've elsewhere called "harnessing collective intelligence.")*“ .^[21]

Pro určení projektu, zda splňuje pravidla a zákonitosti Web 2.0 lze využít níže uvedený tagcloud, neboli množinu vzájemně provázaných klíčových slov. Z níže uvedeného obrázku je patrné, že málokterý web plně odpovídá všem těmto požadavkům.



Obrázek 4: Web 2.0 Mincloud podle [Markus Angermeier]

V běžné praxi ne všechno, co se tváří jako Web 2.0 jím skutečně je, často tento pojem je spíše použit jako nálepka, která se dobře prodává. Vidina může být třeba v odkoupení u začínajících projektů velkými společnostmi jako Google nebo Yahoo. Tyto projekty mají s Web 2.0 společné pouze chytlavé fráze, které ve skutečnosti nic neznamenají. [22]

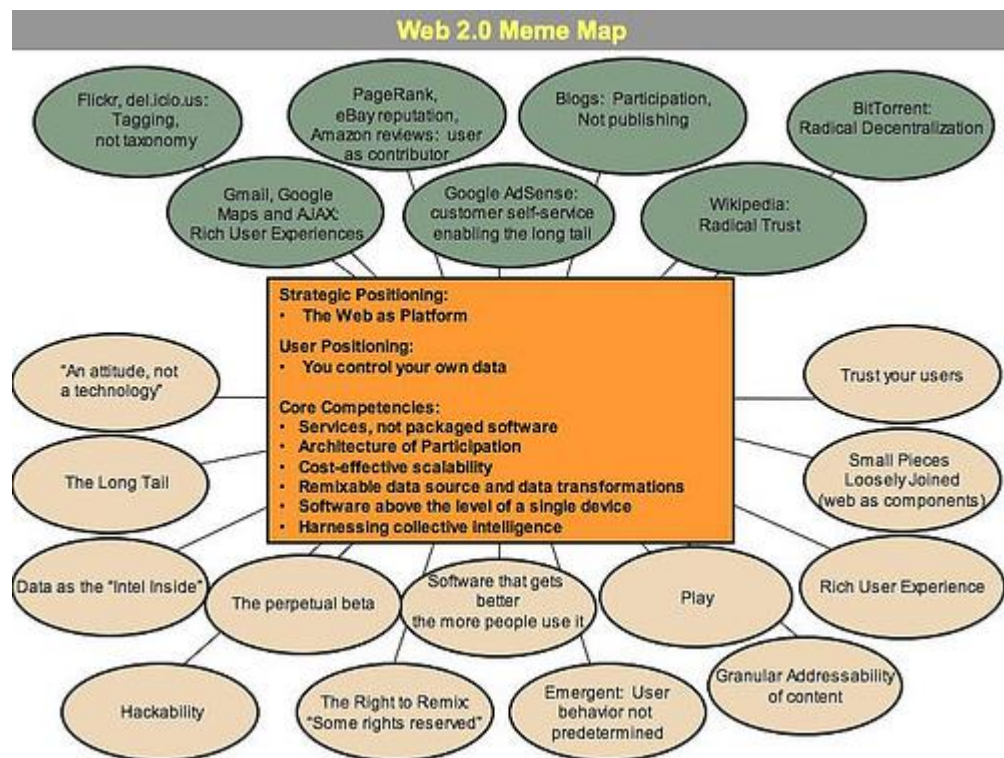
6.1 Web jako platforma

Nejčastější tvrzení v souvislosti s Web 2.0 je heslo „web jako platforma“. Bohužel se jedná o neurčité označení, což ve svém článku uznává i Richard MacManus⁹: *„The Web as Platform”, because I can then fill in the blanks depending on who I’m talking to. For corporate people, the Web is a platform for business. For marketers, the Web is a platform for communications. For journalists, the Web is a platform for new media. For geeks, the Web is a platform for software development. And so on.“* ^[23]

Web jako platforma byl původně obchodním modelem společnosti Netscape, kdy se snažili pomocí svého prohlížeče ovlivnit uživatele, jaké bude používat aplikace nebo jaký obsah bude prohlížet na Internetu.

Web 2.0 si lze představit jako soubor zásad a postupů, které s sebou úzce souvisí a mají společenský dopad pro obchod, kulturu, politiku atd.

⁹ Autor prestižního blogu Read/Write, který se zabývá tématem Web jako platforma



Obrázek 5: Meme Map podle [O'Reilly Media]

Výše uvedený obrázek zobrazuje „meme map“ Webu 2.0, která vznikla v průběhu brainstormingu FOO Camp na konferenci O'Reilly Media.

Pojem Web jako platforma poprvé vznikl v říjnu roku 2004 na konferenci Web 2.0, kdy John Battelle a O'Reilly navrhli předběžný soubor principů, první představoval právě pojem Web jako platforma.

Web 2.0 umožňuje uvést řadu příkladů do praxe, tzn. média naprogramovat. Vznikla řada aplikací, např. od společnosti Google, Bittorrent, Akamai, Last.fm, Facebook atd. s využitím různých technologií, jako Ajax, Adobe Flash, Adobe Flex Framework, JavaScript, PHP, MYSQL, jQuery apod.

Základem pro Web 2.0 je modularita a variabilita systému, kdy si uživatelé mohou poskládat stránku dle vlastního uvážení, z toho ale vyplývá, že musí být striktně oddělen vzhled od obsahu. ^[24]

6.2 Kolektivní spolupráce

Éra Webu 1.0 zaznamenala velký úspěch, který přetrval i do éry Webu 2.0, kdy lidé našli velký potenciál ve sdílení znalostí:

- Základ webu tvoří *hyperlinky*. Uživatelé přidávají obsah, vytváří nové stránky, zasahují do struktury webu jiných uživatelů. Podobně jako u mozku synapse, kvůli které se mozek stává silnější, tak i síla webu roste díky spolupráci mezi uživateli.
- První úspěch na webu zaznamenal *Yahoo* s katalogem linků na díla tisíců později milionů uživatelů webu. Zatímco se Yahoo posunul do další fáze tvorby obsahu, jeho role jako portálu pro kolektivní spolupráci práci stále zůstává velmi významná.
- Nepochybným leaderem ve vyhledávání na webu se stal Google, jehož vyhledávač využívá algoritmu PageRank, kterým hodnotí Google důležitost zaindexovaných stránek. Díky němu se uživatelům zobrazují co možná nejrelevantnější výsledky.
- Ebay je dalším produktem, ve kterém mají možnost lidé spolupracovat. Ebay organicky roste v reakci na aktivitě uživatelů a společnosti jako prostředku, ve kterém se tato aktivita odehrává.

- Další významnou aplikací je Amazon.com¹⁰. Uživatelé mají možnost vkládat komentáře k jednotlivým produktům nebo hodnotit recenze ostatních uživatelů. Amazon sleduje aktivitu svých uživatelů a díky ní zobrazuje uživatelům lepší výsledky ve vyhledávání.
- Wikipedia je online encyklopedie, kam každý uživatel může přidávat vlastní články a tyto články pak můžou editovat jiní uživatelé. Tento revoluční způsob ve tvorbě obsahu použil ve výroku Eric Raymond v souvislosti s open source software: „*with enough eyeballs, all bugs are shallow*“. Wikipedia je již mezi webovými stránkami v TOP 100, zanedlouho může být i mezi TOP 10.
- SourceForge.net je největší projekt na světě, kam si kdokoliv může umístit svůj open-source software. V současné době je zde umístěno více než 260 tis. projektů. Kdokoliv si může software stáhnout nebo použít jeho kód.^[24]

6.2.1 Bloging

Pojem *blog* je odvozenina od *web log*¹¹ [25], což lze volně přeložit jako webový zápisník. Weblogy představují diferenciované pole, takže pokusy o stručnou definici jsou takřka nemožné. Spektrum blogů je velmi široké, sahá od osobních deníčků až po zpravodajství firem, politických stran apod. Příspěvkatelem může být

¹⁰ Amazon.com byl spuštěn v roce 1995, jehož zakladatelem byl Jeff Bezos. Při průzkumu podnikatelských činností narazil na článek, který popisoval obrovský nárůst uživatelů World Wide Webu. Hovořilo se o tom, že o pouhý rok vzrostl nárůst uživatelů o 2300 procent. Po finanční analýze trhu si sestavil seznam nejdůležitějších produktů, které by šlo nejsnáze nabízet na internetu. Nejefektivnější se ukázaly právě knihy.

¹¹ Tento termín pravděpodobně vznikl v roce 1997 Jornem Bargerem. O dva roky později vznikla jeho zkrácená verze blog Peterem Merholzem, kdy došlo k rozdělení slovo weblog na frázi web blog.

jedna osoba nebo celá komunita. U většiny článků je možné přidávat komentáře, kolem těchto příspěvků pak vzniká tzv. čtenářská komunita.

Na webu se vyskytují blogy, které mají odlišnou strukturu. Může se jednat o prostý seznam odkazů nebo o dlouhé původní texty. Jednotlivé příspěvky jsou označeny datem a časem, což jsou pro blogy velmi důležité informace. Odkazy jsou pro blogy relevantní informace, podle nich lze archivovat starší příspěvky, kdy se jim určí stabilní URL, tzv. permalink. ^[26] V současné době je již považováno za standart nabízet přehled nových příspěvků pomocí RSS.

Před blogy byly populární v mnoha formách digitální komunity. Takovým byl i Usenet, elektronický mailing list a Bulletin board systém (BBS). V roce 1990 se objevil software pro internetová fóra (WebX), který byl vytvořen pro související konverzace.

6.3 Datová základna

Každá význačná internetová aplikace je již v současné době postavena nad databází: databáze vyhledávání od Google, u Yahoo katalog odkazů, Amazon využívá databázi produktů a prodávajících, MapQues mapové databáze, Napster databázi songů. V roce 2004 Hal Varian¹² poznamenal, že: „*SQL is the new HTML.*“ Správa databází je proto velmi důležitá u společností zabývajících se Web 2.0. ^[24]

V éře internetu byly u firem zaznamenány řady případů, kdy sledováním databáze bylo dosaženo kontroly trhu, díky tomu i nadměrným finančním

¹² Specialista v oboru mikroekonomie a informatice v ekonomii. V současné době je ekonomickým ředitelem Google a také emeritním profesorem University of California Berkely.

výnosům. Monopolem se například stala firma Network Solution, která se zabývala registrací doménových jmen.

Velké společnosti jako Google, Yahoo, MapQuest do svých databází uskutečnily významné investice, např. Navteq údajně investoval 750 mil. dolarů, aby mohl vytvořit databázi ulic, adres. Digital Globe, utratil 500 mil. za vypuštění vlastního satelitu pro zlepšení vládou dodávaných snímků. ^[24]

6.4 Konec release fází u vývoje softwaru

Jednou z určujících charakteristik éry Internetu je skutečnost, že software je dodáván jako služba, nikoliv jako produkt. Tato skutečnost vede k řadě zásadních změn v obchodním modelu společností: ^[24]

- *Operace se stávají základními nástroji.* Vývoj aplikací od Google nebo Yahoo musí reflektovat každodenní zpracování operací. Google neustále prochází web a aktualizuje indexy, filtruje odkazy na spam, průběžně reaguje na stovky milionů asynchronních uživatelských dotazů. Není náhodou, že správa systému, sítí a technik zabývajících se vyrovnáváním zátěže je možná střeženějším tajemstvím než vyhledávací algoritmus. Úspěch Google v automatizaci těchto procesů je velký a drží tak významný náskok před konkurencí. ^[24]

V současné době skriptovací jazyky jako Perl, Python, PHP a nyní i Ruby hrají důležitou roli při vývoji aplikací Web 2.0. Tímto popisem charakterizoval jazyk Perl Hassan Schroeder, první webmaster Sunu: „*the duct tape of the internet*“. Řada aplikací pro správu systému nebo sítě jsou

psané v různých skriptovacích jazycích a tak si správci mohou vybrat ty, které budou používat. [24]

- S uživatelem musí být zacházeno jako s tvůrcem, v odkazu na způsob vývoje open source aplikací. Výrok týkající se open source „*vydávej brzo, vydávej často*“ se radikálněji změnil na „*věčná beta*“. Produkt se ve fázi beta dále volně vyvíjí, je rozšiřován novými doplňky v týdenních či měsíčních intervalech. Není proto zvláštní, že produkty Google, např. Gmail, Google Maps byly ve fázi beta i několik let. Cal Henderson, hlavní vývojář Flickr prozradil, že každou půlhodinu dochází k nasazování nových buildů. [24] Toto je již ale extrémní způsob vývoje. Zcela jistě ne všechny webové aplikace jsou vyvíjeny v podobném duchu jako Flickr, přesto je snaha dodržovat nějaký vývojový cyklus, který je ale zcela odlišný od klasického vývoje desktopových aplikací. Redaktoři ZDnet nedávno dospěli k závěru, proč Microsoft nikdy nemůže porazit Google: „*Microsoft's business model depends on everyone upgrading their computing environment every two to three years. Google's depends on everyone exploring what's new in their computing environment every day.*“ [27]

6.4.1 Služby na webu nahrazují desktopové aplikace

Mezi aplikacemi Web 2.0 existuje řada těch, které nahrazují klasické desktopové aplikace. Podmínkou jejich využití bývá vysokorychlostní připojení.¹³ V roce 2004 došlo k zlomu, kdy Google představil službu Gmail s ajaxovým rozhraním

¹³ Dle metodiky ČSÚ lze označit za vysokorychlostní připojení takové, jehož přenosová rychlost dosahuje 144 kb/s a více.

(umožňující full-textové vyhledávání, kapacitu schránky převyšující 1 GB). Dále například služba del.icio.us, která slouží pro ukládání oblíbených záložek z prohlížeče. Tyto záložky se navíc dají sdílet mezi uživateli.

V oblasti tvorby textových či tabulkových dokumentů v roce 2005 Google přišel se službou Google Docs. O dva roky později Microsoft představil svoji online podobou Office. U těchto online podob kancelářských aplikací je velkou výhodou možnost simultánní úpravy dokumentů, tj. uživatelé pracující na jednom dokumentu mohou v reálném čase navzájem vidět úpravy, které jsou zde prováděny.

Mezi další aplikace, které se přesouvají na Internet, lze zařadit i grafické editory (Rsizr) nebo instant messaging (Meebo), RSS čtečky (Bloglines) a spoustu dalších typů aplikací. ^[22]

6.4.2 Příklady aplikací Web 2.0

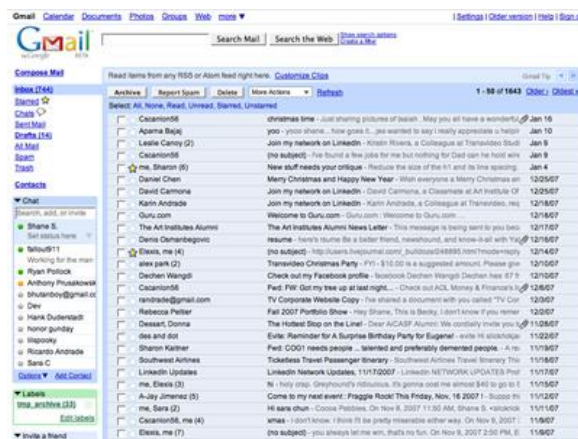
Gmail

Gmail je bezplatná emailová služba dosažitelná přes webové rozhraní, která umožňuje stahovat poštu přes protokoly POP3 nebo SMTP. [28] Od dubna 2004 byl umožněn přístup k této aplikaci pouze přes pozvánky, od února 2007 byla zpřístupněna verze Beta. V té době měla kapacitu schránky dosahující pouze 1 GB. [29] V dnešní době se kapacita pohybuje přes 7 GB a průměrně narůstá o 3,3 MB za den. Uživatelé mají možnost si za příplatek rozšířit datový prostor až na 16 TB (datový prostor je rozložen mezi Gmail, Picasa Web Albums a Google Docs). [30]

Do prostředí Gmailu je integrována služba Google Talk určená pro instant messaging, Google Calendar, Google Docs.

Webové rozhraní využívá technologie AJAX [31], která umožňuje například kontrolu překlepů, klávesové zkratky nebo RSS.

K zabezpečení komunikace mezi klientem a serverem je využito šifrovaného spojení, k přístupu na POP3 a IMAP se využívá TLS. [32] Přílohy v odchozí a příchozí poště jsou automaticky testovány na přítomnost virů.



Obrázek 6: Gmail podle [TechNewsWorld]

Google+

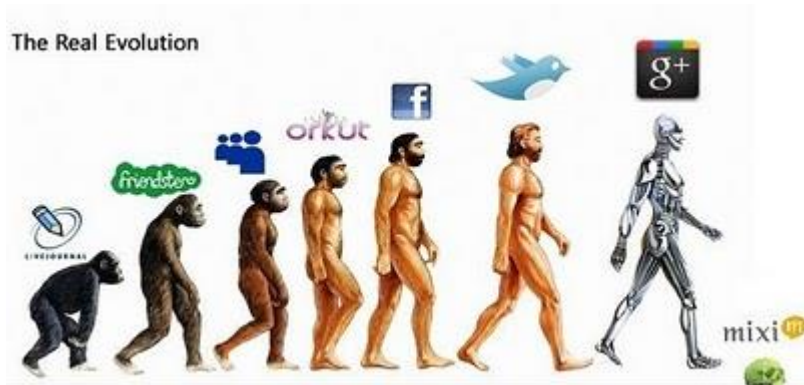
Jedná se sociální službu od společnosti Google. ^[33] Služba byla spuštěna 28. června 2011 do testovací fáze, kdy byla přístupná pouze přes pozvánky. Následujícího dne měli existující uživatelé možnost pozvat přátele, kterým bylo 18 let a výše.

Google integruje sociální služby jako Google Profiles nebo Google Buzz, dále byla představena nová služba Circles, Hangouts, Sparks a Huddles. ^[34]

Google+ je dostupná přes webové rozhraní, ale bude dostupná i jako desktopová aplikace. V současné době je přístupná i jako mobilní aplikace pouze na Androidu a iOS.

14. června 2011, tj. dva týdny po spuštění již bylo na Google+ registrováno více než 10 mil. uživatelů. Po 4 týdnech od uvedení do provozu službu využívalo již 25 mil. uživatelů. ^[35]

Kruhy umožňují třídit kontakty do skupin pro třídění. Ostatní uživatelé mohou vidět seznam lidí v kruhu, ale nemohou vidět jeho název. Nastavení ochrany osobních údajů umožňuje skrytí uživatelů v kruhu. Zařazování přátel do kruhů se provádí pomocí drag-and-drop.

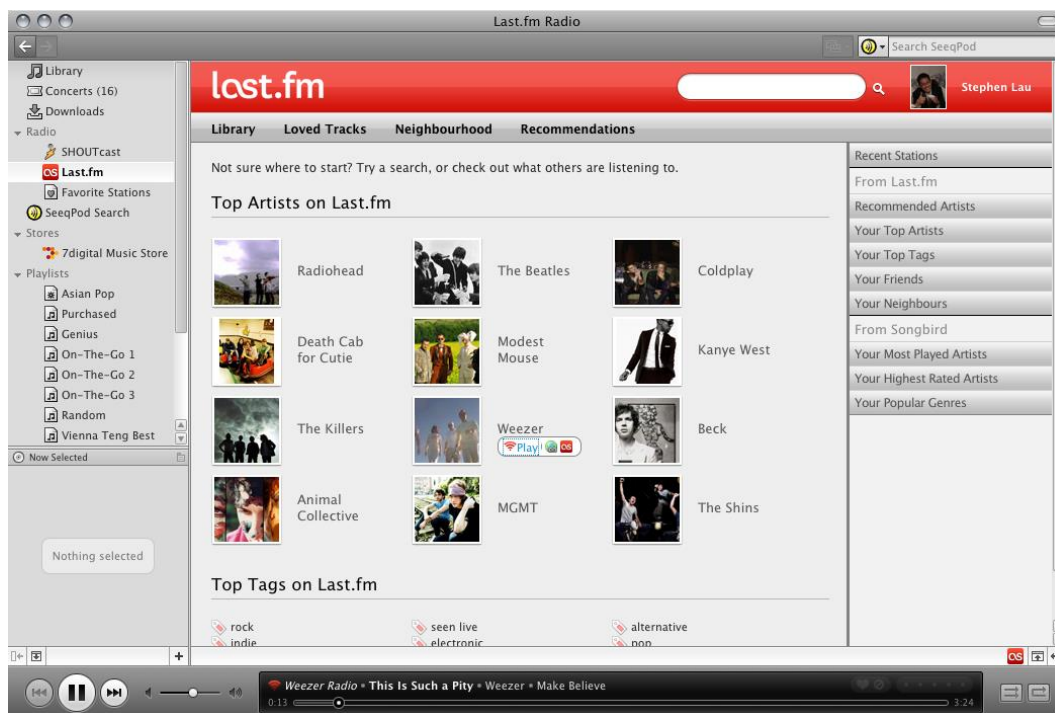


Obrázek 7: Google + podle [Rightyleft.com]

Last.fm

Tato služba může být využita jako internetové rádio, hudební encyklopedie nebo systém pro doporučení hudby. Systém tvoří profil toho, co uživatel poslouchá. Eviduje se například název umělce, název skladby. Tyto informace jsou získávány ze skladeb za pomoci pluginu Audioscrobler, který má uživatel nainstalován ve svém multimediálním přehrávači.

Uživatelé této služby si mohou tvořit svůj profil dvěma způsoby, jedním za použití výše uvedeného pluginu nebo poslechem Last.fm rádia. Názvy skladeb jsou ukládány do databáze, ze které se tvoří žebříčky a doporučení. Profilové stránky zobrazují také nedávno hrané skladby. [36]



Obrázek 8: Last.fm podle [Songbirdnest.com]

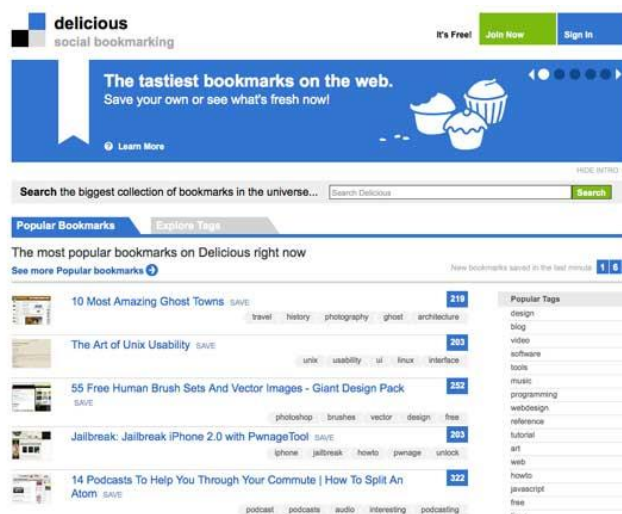
Del.icio.us

Tato služba je využívána pro sdílení, ukládání a objevování webových záložek. V roce 2003 byla založena Joshuou Shcachterem, o dva roky později ji získal Yahoo a nakonec v 27. 4. 2011 tuto službou koupil AVOS Systems. Na konci roku 2008 zde bylo zaregistrováno 5,3 mil. uživatelů a čítala 180 mil. unikátních záložek. [37]

Del.icio.us je založena na nehierarchickém kategorizačním systému, kde uživatelé mohou tagovat každou ze svých záložek libovolným počtem znaků. Uživatelé mají možnost se podívat na záložky ostatních uživatelů s definovaným tagem, tj. <http://del.icio.us/tag/idnes>, znamená, že se zobrazí všechny záložky, které jsou tagované „idnes“.

Delicious má také hotlist, kde se zobrazují populární a nedávno navštívené stránky.

Standardně jsou všechny záložky veřejné, tj. kdokoliv se na ně může podívat, ale existuje možnost záložky označit jako soukromé. [38]



Obrázek 9: Delicious podle [Crowdshifter.com]

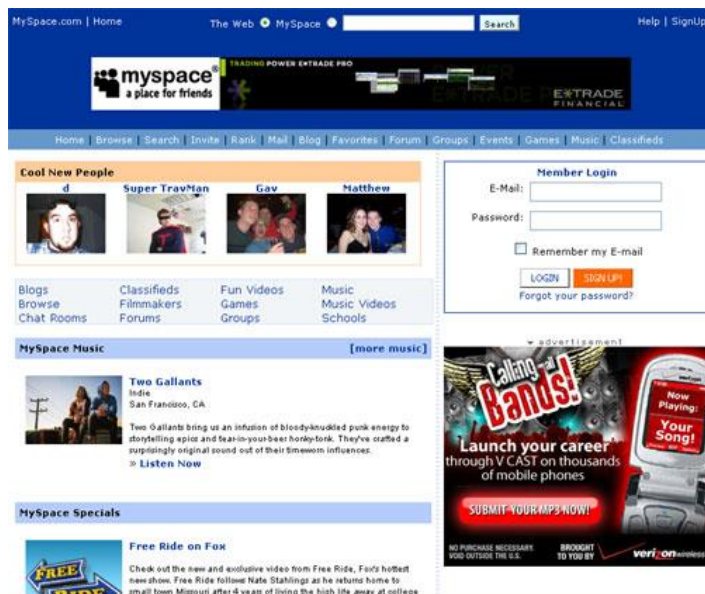
MySpace

MySpace se stal nejpoblárnější síí ve Spojených státech v roce 2006 a v této pozici setrval až do roku 2008, kdy ho v počtu unikátních návšívěvníků předběhl jeho hlavní konkurent Facebook.

Jedná se o komunitní server, kde se kromě obyčejných profilů lidí nacházejí i profily muzikantů, herců z celého světa. Kvůli ochraně dětí je služba přístupná pouze lidem starších 18 let. [39]

Je zde integrován vlastní instant messenger MySpaceIM, kterého je možné využít jako samostatnou aplikaci pro operační systém Windows.

Pro zajímavost je vhodné uvést, že MySpace byl koupen v roce 2005 společností News Corporation magnáta Ruperta Murdocha za 580 mil. dolarů a díky poklesu zájmu na úkor Facebooku bylo 95% akcií prodáno v roce 2011 společnosti Specific Media za pouhých 35 mil. dolarů. [40]



Obrázek 10: MySpace podle [Wired.com]

6.5 Jednoduché programovací modely

V rámci webových služeb velké společnosti začali poskytovat uživatelům komplexní nástroje s cílem tvorby spolehlivého vývojového prostředí pro distribuované aplikace.

Pro svou jednoduchost se právě služba RSS stála nerozšířenější webovou službou na Webu.

Amazon.com nabízí své webové služby ve dvou variantách, jedna je založena na SOAP, další na výměně XML dat přes HTTP, v odlehčeném významu se hovoří jako o REST. Webová služba založená na SOAP má velkou hodnotu v rámci B2B, jako mezi firmami Amazon.com a ToysRUs. Nicméně Amazon uvádí, že 95% webových služeb je založeno na REST.

Jednoduché business modely jsou přirozeným jevem jednoduchého programování a spojení. Myšlenkou Web 2.0 je tyto modely opět využít. Příkladem jednoduchého propojením dvou existujících služeb může být služba Housingmaps.com.

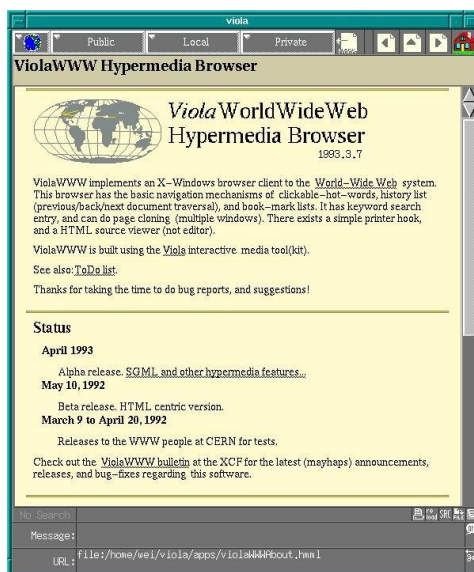
Výše uvedené příklady ukazují, že na Web 2.0 lze nahlížet jako na *"innovation in assembly"*. Web 2.0 dává firmám příležitost porazit konkurenci tím, želepší využití a integraci svých služeb oproti konkurentům.^[21]

6.6 Bohaté uživatelské schopnosti

Již prohlížeč Pei Wei's Viola v roce 1992 podporoval applety a další formy zobrazení aktivního obsahu. ^[41] V roce 1995 vznikl JavaScript ^[42] a od té doby se datuje vznik DHMTL, což umožnilo uživatelům vytvářet dynamický obsah na Webu. Před několika lety se objevil termín „Rich Internet Application“, který prosazovala hlavně firma Macromedia kvůli svému open source systému Flash.

Potencionál Webu se v této době ještě naplno neprojevil dokud Google nepředstavil Gmail a Google Maps, aplikace s uživatelským rozhraním podobné interfacům desktopových aplikací. ^[24] Ve své práci se Jesse James Garret z firmy Design Web zaměřil na využití AJAXu ve společnosti Google: *„Ajax isn't a technology. It's really several technologies, each flourishing in its own right, coming together in powerful new ways. Ajax incorporates:*

- *standards-based presentation using XHTML and CSS;*
- *dynamic display and interaction using the Document Object Model;*
- *data interchange and manipulation using XML and XSLT;*
- *asynchronous data retrieval using XMLHttpRequest;*
- *and JavaScript binding everything together.”* ^[43]



Obrázek 11: Viola podle [Viola.org]

Lingvistika a přirozený jazyk

Klasický obor lingvistika v sobě zahrnuje několik významných podoborů jako gramatika¹⁴ nebo sémantika¹⁵.^[44]

Lingvistika je definována jako vědecký přístup pro studium jazyka, na který lze nahlížet z mnoha směrů. Řada dalších oborů, které jsou s lingvistikou spřízněny, mohou tento obor ovlivňovat. Příkladem může být sémiotika, která se zabývá studiem obecných znaků a symbolů v jazyce i mimo něj.^[44]

7.1 Rozdělení lingvistiky

Lingvistika se zabývá popisem a vysvětlení podstaty lidského jazyka. Za relevantní lze pokládat tyto otázky, co je to univerzální jazyk, jak se jazyk může odlišovat, nebo jak se lidé učí rozumět jazyku. Každý člověk má možnost se naučit mluvit jazykem z prostředí, ve kterém vyrůstá a nepotřebuje k tomu zvláštní

¹⁴ Gramatika představuje soubor pravidel, kterými si řídí struktura složení věty, frází a slov v přirozeném jazyce. Tato věda v sobě zahrnuje morfologii (tvorba a skládání slov), skladbu (soubor pravidel, jak slova spojit do frází a vět) a fonologii (zvuková stránka přirozených jazyků).

¹⁵ Sémantika se zabývá studiem významu jednotlivých slov, morfémů nebo také vztahem slov k realitě, kterou popisují.

potřebu výuky. Lingvisté proto předpokládají vrozenou schopnost používání jazyka podobně jako schopnost chůze. ^[45]

Vedle konstrukčně motivovaných oborů jako lexikologie, fonetika, pragmatika apod. existují obory v oblasti lingvistiky, které nezkoumají jazykové podsystémy, jedná se o: ^[45]

- **Deskriptivní lingvistika**

Od počátku vzniku oboru lingvistika se lingvisté zabývali popisem a dokumentováním jazykového systému, ale i užíváním jazyka v různých komunikačních situacích. Výstupem práce jsou pak různé lexikony, mluvnice popisující slovní zásobu a gramatiku určitého jazyka.

- **Teoretická lingvistika**

Tento podobor lingvistiky zkoumá principy fungování přirozeného jazyka, často využívá produkty deskriptivní lingvistiky. Výsledky tohoto zkoumání jsou různé teorie a hypotézy, jejichž důvěryhodnost je testována na datech z konkrétních jazyků.

- **Aplikovaná lingvistika**

Termín aplikované lingvistiky se často odkazuje na lingvistický výzkum v oblasti výuky jazyků. Výsledky zkoumání jsou pak dále použity v mnoha dalších oblastech, například v lexikografii nebo v překladech. V dnešní době se široce využívají v oblastech aplikované lingvistiky počítače. Syntéza a rozpoznávání řeči využívají fonetickou znalost k poskytnutí zvukového interface k počítači. Do popředí se dostávají aplikace týkající se počítačové lingvistiky jako například zpracování nebo překlad přirozeného jazyka. Jejich vliv měl efekt na teorie syntaxe

a sémantiky stejně tak jako modelování syntaktických a sémantických teorií na omezené možnosti počítačů.

- **Preskriptivní lingvistika**

Cílem tohoto oboru je popsat tzv. spisovný jazyk, který by měl být použit ve školách nebo veřejných sdělovacích prostředcích. Spisovný jazyk je sjednocen pro češtinu v Pravidlech českého pravopisu. Mnoho přirozených jazyků takto oficiálně ustanoveno podobu nemá. ^[45]

7.2 Přirozený jazyk

Přirozený jazyk je takový jazyk, který slouží jako dorozumívací prostředek mezi lidmi. Používá se v mluvené nebo psané formě. Přirozený jazyk na jedné straně vymezuje konstruované jazyky, označované jako programovací jazyky (Java) a uměle vytvořené jazyky (Esperanto) a na druhé straně systémy komunikace v přírodě jako například tanec včel.

Pochopení přirozeného jazyka odhaluje, jakým způsobem funguje jazyk z hlediska syntaxe, sémantiky, fonetiky, ale také i jak funguje lidská mysl.

Formalizovaný jazyk není vlastně jazykem přirozeným, jelikož se podřizuje určitým pravidlům a výzkumem a ne tomu, co lidé ve skutečnosti chtějí říci. ^[46]

Počítačové zpracování živých jazyků

Počítačovému zpracování živých jazyků se věnuje počítačová lingvistika neboli též komputační lingvistika, která se na rozdíl od klasické lingvistiky opírá o termíny jako algoritmus, datové struktury apod. V této vědní disciplíně je často využíváno i poznatků z dalších oblastí výzkumu, mimo jiné z oblasti umělé inteligence.

Trh s programy zabývající se počítačovému zpracování přirozených jazyků lze rozdělit do dvou větších skupin. První představují programy pro zpracování textu v přirozeném jazyce. Do této skupiny lze zařadit různé korektory překlepů, korektory gramatiky, fulltextové programy, aplikace umožňující strojový překlad z jednoho jazyka do druhého apod. Do druhé skupiny patří například aplikace založené na dotazovacích jazycích, které přistupují k databázím, různé systémy založené na automatizaci umožňující komunikaci s klienty v různých institucích.

[47]

8.1 Morfologie

Důležitou oblastí počítačového zpracování živých jazyků je morfologie. Tento obor se zabývá tvorbou tvarů a jejich významu. Nejmenší jednotku jazyka tvoří tzv. morfém, který je nositelem významu. Morfémy lze rozdělit na *lexikální*, čili nesoucí

význam slova nebo *gramatické*, které určují gramatickou roli příslušného slovního tvaru. Například u slova *cestou* je lexikální morfém *cest-* a gramatický morfém *-ou*.

O češtině a všeobecně o slovanských jazycích je známo, že mají složitou morfolologii, což v počítačové lingvistice činí jisté problémy. Příkladem může být vyhledávání slov v textu v jazyku FLEX¹⁶. Pokud budeme chtít v angličtině vyhledat všechny výskyty jednotného a množného čísla slova *table*, stačí pouze napsat řetězec „table(s)?“, což znamená, že hledaný vzorek je ukončen maximálně písmenem „s“. Výsledkem je, že se vrátí dva tvary slova a to *table* nebo *tables* a to je přesně to, co uživatel potřebuje. U jazyků se složitějším skloňováním nastává problém, je totiž potřeba vypsat všechny tvary jednotného a množného čísla: *tabul(ka|ky|ce|ku|kou|ek|kám|kách|kami)*.^[44]

Morfolologii lze rozdělit do dvou hlavních oblastí:

- formální morfologie, neboli vlastní tvarosloví – tato oblast se týká tvorby slovních tvarů ohýbáním nebo dalšími procesy kam patří skloňování (deklinace) a časování sloves (konjugace).
- derivační morfologie, tvorba slov – patří sem odvozování (derivování) jednoslovných výrazů z jiných, obvykle považovaných za základní na úrovni morfémů (*škola* – *školník*, *učitel* – *učitelka*, *kniha* - *knižní*).^[44]

Tvorba slov zatím není dotčena metodologickými postupy, které se v poslední době začínají objevovat v jazykovém inženýrství a oblastech umělé inteligence s počítačovým zpracováním přirozeného jazyka. Rozumí se tím zachytit

¹⁶ jedná se o imperativní objektově orientovaný jazyk, který vznikl v pražské firmě A && L soft

slovotvorné vztahy pomocí lexikálních databází využívající různých typů ontologií – dobrým příkladem může být EuroWordNet¹⁷ a další lexikální zdroje.^[47]

V rámci morfologie tvorba slov znamená subdisciplínu a k algoritmickému popisu se zlehka začíná přistupovat. Pro češtinu je nejprve vhodné vytvořit tzv. derivační strom, čili slovník, ve kterém jsou kořeny českých slov a morfémy, které se s kořeny budou kombinovat.^[44]

Algoritmický popis českého tvarosloví zahrnují deklinaci, konjugaci, stupňování a další slovotvorné procesy. K vytvoření je potřeba naleznout pravidla, dle kterých se budou ohýbat slova a jak pravidla implementovat. U algoritmického přepisu se tento problém dá vyřešit zavedením počtu podvzorců, které zachycují dané hláskové změny, např. lékař – lékaři, pes – psi (výčet všech obměn lze nalézt v práci Osolsobě, 1994). Na tomto algoritmu je třeba založen program AJKA, kde se rozpoznávají slovní tvary nebo jsou programem přímo generovány, vstupnímu slovu přiřazuje slovní druh, pád, osobu, číslo, vid apod.^[47]

8.2 Korpus

Korpus je soubor počítačově uložených textů (v případě mluveného jazyka - přepisů záznamu mluvy), který slouží k jazykovému výzkumu.^[48]

¹⁷ EWN je vícejazyčná databáze sítí slov několika evropských jazyků (holandština, italština, španělština, němčina, francouzština, čeština a estonština). Sítě jsou podobně strukturovány jako americký WordNet (angličtina) pro synonymické řady (synets). Každý wordnet reprezentuje unikátní jazyk, vnitřní systém lexikalizace. Od WordNetu se EWN liší, že je možné přejít od slova z jednoho jazyka k podobnému slovu v jiném jazyce.

K práci s korpusem se využívá vyhledávacího programu, který dokáže vyhledat slova a slovní spojení v kontextu a také dokáže zjistit jejich frekvenci výskytu v korpusem, ale i původní textový zdroj.

V současné době korpusová lingvistika představuje novou větev lingvistiky, kde se pracuje s korpusem uloženými v počítačích. V mnohých aspektech se prolíná s počítačovou lingvistikou. Díky tomuto spojení vznikají přesnější elektronické slovníky a robustní počítačové gramatiky.

V České republice se v rámci projektu na Filosofické fakultě Univerzity Karlovy v Praze od roku 1994 buduje Český národní korpus (ČNK). V současné době ČNK čítá cca 200 mil. českých slovních tvarů.^[48]

V dnešní době se korpusem organizují dle potřeb lexikografů, sociologů a vycházejí z těchto teoretických předpokladů:

- jazyková data v korpusem jsou uložena ve své přirozené textové podobě,
- velký rozsah dat v korpusem minimalizuje nebezpečí, že by mohlo dojít k převaze okrajových jevů nad základními a typickými,
- velký rozsah dat v korpusem je podmínkou dostatečné reprezentativnosti, což při budování slovníků nemusí být jednoduchá záležitost.^[47]

8.2.1 Vznik korpusem

Zdroj pro korpusem je psaný jazyk, ale i mluvený, který se převádí počítačem do čitelné podoby. Z psaných textů se data získávají následujícími způsoby:

- konverzí ze sázecích disket a pásek, které se získávají od většiny nakladatelství vydávajících noviny, časopisy, knihy,

- pomocí technik OCR, kde úspěšnost je závislá na kvalitě použitého scanneru nebo softwaru,
- manuálním opisováním psaného textu. ^[47]

Ve všech způsobech je nutná kontrola chyb. Výsledek je kódován do ASCII kódu, ale v poslední době je preferován UTF-8. Při tvorbě korpusů je také často nutné se vyrovnat s autorskými právy. ^[47]

Na příkladu British National Corpus lze pozorovat, z čeho je například korpus tvořen:

- je to výběrový korpus, tj. skládá se z vzorků ne delších než 40 000 slov, které jsou vybrány v následujících proporcích:

1. přírodní vědy a čistá věda5%
2. aplikované vědy5%
3. sociální vědy15%
4. politická15%
5. publicistika obchodní a finanční10%
6. publicistika umělecká (rock & pop, divadlo, . . .)10%
7. publicistika náboženská a filosofická5%
8. publicistika zábavná (sport, zahrádkáři, . . .)15%

Podíl těchto textů se pohybuje v rozmezí 70-80%, podíl uměleckých textů činí 20 - 30%.

Další rozlišení se týká toho, zda vzorky pocházejí z knih, deníků, časopisů, dopisů apod.:

1. knihy 55-65%
2. periodika 20-30%

3. brožury, letáčky, příručky, reklamy 5-10%
4. dopisy, memoranda, zprávy, eseje..... 5-10%
5. mluvené texty..... 7-10%

- je synchronní, tedy obsahuje výhradně texty ne starší než např. od r. 1987, a vždy se uvádí datum, kdy byl text publikován poprvé,
- je obecný čili není specificky orientován na nějakou konkrétní oblast nebo žánr a zahrnuje vzorky od všech věkových skupin, viz výše,
- je jednojazyčný – obsahuje jen vzorky pocházející od anglických (českých,...) mluvčích.
- jsou zavedeny klasifikační rysy, které nesledují vyhraněné proporce a jsou orientovány na pozdější využití korpusu (lze podle nich třídit a vyhledávat v celém korpusu):
 1. identifikátor vzorku,
 2. rozsah vzorku (počet slov), začátek a konec vzorku,
 3. rozsah textu příslušného typu (počet slov),
 4. kompozice textu (hladký, složený, sbírka),
 5. standardní bibliografický odkaz,
 6. datum vzniku,
 7. předmětná oblast,
 8. úroveň složitosti textu,
 9. autorství (individuální, společné, institucionální, neznámé),
 10. pohlaví autora,
 11. věková skupina autora,

12. etnická skupina autora,
13. autorovo bydliště,
14. věk cílové skupiny (na kterou je text orientován).

[47]

8.3 Korpusové nástroje

Konkordanční programy (MicroOP) jsou základem korpusových nástrojů, které třídí a počítají objekty nalezené v korpusu. Objekt může být například slovní tvar, interpunkce, případně jiný znak (vyznačující hranici vět, odstavců). Tyto objekty jsou součástí SGML. Do korpusových nástrojů se počítají i různé gramatické analyzátoři, které se orientují na morfologii nebo syntax ale také i na sémantiku. V současnosti se do těchto nástrojů vmísil pojem taggování, tzn. snaha každému slovu či slovnímu spojení přiřadit gramatickou značku, čili slovní druh včetně relevantních gramatických teorií. [47]

8.3.1 Nutnost standardizace

Volně uložené textové soubory nelze nazývat korpusem. Korpusem se dají označit data, která jsou uložena v:

- elektronické archivy – kolekce celkově nesourodých textů, příkladem může být OTA (Oxford Text Archive), který skýtá rozsáhlou sbírku literárních textů v různých formátech a různých jazycích,
- vlastní korpusy tvořící relativně úplné celky lišící se řadě parametrů,
- dle jazyků – v dnešní době v Evropě není mnoho jazyků, které nemají vlastní korpus.

S rostoucím počtem typů korpusů bylo zapotřebí korpusy standardizovat, proto vznikla Text Encoding Initiative (TEI), která vydala doporučení pro společný formát, znakové sady a značkovací metajazyk, čímž je Standard Generalized Markup Language (SGML), který je od roku 1986 uznávaný mezinárodní standard (ISO 8879).^[47]

8.4 Strojový překlad

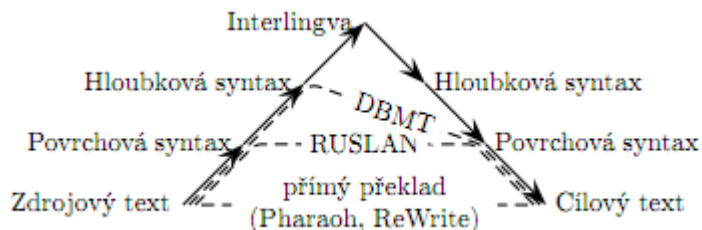
Strojový překlad se někdy označuje zkratkou MT (Machine Translation). Jedná se o automatický překlad z jednoho živého jazyka do druhého s využitím počítače. Tento typ překladu je součástí počítačové lingvistiky.

V základním pojetí strojový překlad poskytuje jednoduchý překlad slova v jednom přirozeném jazyce do druhého. Problémy s translací vznikají u textů, kdy strojově přeložená věta je kostrbatě přeložená. Řešením toho nedostatku spočívá

v korpusech a technikách statistického překladu, kdy radikální rozšíření pole povede k lepším překladům. [47]

Současný překladový software umožňuje nastavit profesi, která je zohledněna při překladu textů, tím se zlepší kvalita překladu. Tento způsob je zvláště účinný u překladu formálních jazyků. Překlad právní dokumentace má ve výsledku lepší výstup než překlad u běžné konverzace. Lidskými zásahy je možnost regulovat kvalitu překladu tak, že uživatel má možnost identifikovat, jaká slova v textu jsou názvy.

Proces strojového překladu se obvykle zakresluje do trojúhelníku, kde je vyobrazena cesta od zdrojového textu přes interlingvu do cílového textu. Cílem tohoto znázornění je překlad s co nejmenší chybou. [47]



Obrázek 12: Trojúhelník strojového překladu s ilustracemi existujících systémů podle [Ondřej Bojar, CUNI]

8.4.1 Historie překladu

První systém pro strojový překlad byl představen v roce 1954 v rámci GeorgeTown experimentu výzkumníky IBM. Experiment spočíval v plně automatickém překladu šedesáti ruských vět do angličtiny. Tento projekt zaznamenal velký úspěch a také předpoklad pro velké finanční zisky. [49]

8.4.2 Princip překladu

Strojový překlad využívá metody lingvistických pravidel, která znamenají, že slova budou přeložena lingvistickou cestou – cílová slova nahradí ty, která jsou ve zdrojovém jazyce. Úspěch strojového překladu závisí hlavně na pochopení přirozeného jazyka.^[50]

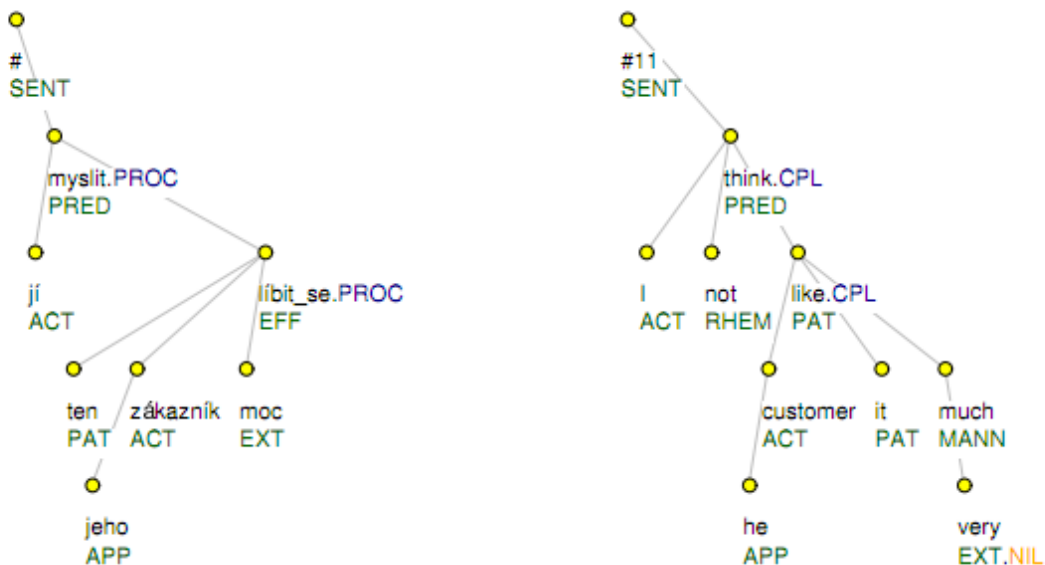
Obecně u strojového překladu platí, že při analýze textu a využití pravidel dojde k vytvoření spojky, symbolické reprezentace, pomocí které se vygeneruje text v požadovaném jazyce. Dle této povahy se strojový překlad rozděluje na interlingvu a statistický strojový překlad. Obě tyto metody využívají rozsáhlé lexikony s morfologickými a syntaktickými pravidly.^[50]

Interlingva překlad

Tento způsob je jedním z klasických přístupů strojového překladu. Zdrojový text je napřed transformován do tzv. interlingvi, neboli abstraktního jazyka, odkud je pak přeložen do požadovaného jazyka. Tento přístup je alternativou k přímému překladu.

V této metodě interlingva umožňuje zanalyzovat text zdrojového jazyka tak, aby bylo možné změnit jeho morfologické, syntaktické i dokonce pragmatické vlastnosti u významu v cílovém jazyce. Při překladu je možné použít dva druhy interlingvy, kdy jeden druh se více zaměřuje na analýzu zdrojového jazyka a další na vlastnosti cílového jazyka, čili se ve skutečnosti překládá ve dvou etapách. Výsledkem je pak konkrétnější překlad, protože je kladen důraz na použití přesnější slovní zásoby u cílového jazyka.

U překladu z cizích jazyků do češtiny se používá metoda funkčního generativního popisu FGP. Na této metodě jsou založeny systémy RUSLAN pro překlad z češtiny do ruštiny a DBMT pro překlad z češtiny do angličtiny. [50]



Obrázek 13: Ukázka hloubkové korespondence mezi českou a anglickou větou podle [Ondřej Bojar, CUNI]

Statistický strojový překlad

Tento způsob se od logistického přístupu liší hlavně v jednoduchosti procesu překladu, tzn. nejprostší reprezentace vstupu a výstupu. Dříve byly systémy založeny na překladu jednotlivých slov s možností prohazování. V současné době se využívá překladu po *frázích*, nedělitelných víceslovných celcích. [50]

V roce 1949 Warren Weaver představil svůj návrh na statistický strojový překlad, kde uplatnil Claude Shannonovu teorii informace. V dalších letech tento způsob překladu nebyl moc využíván až do roku 1991, kdy se o metodu začali

zajímat vědci v IBM a přispěli tak k výraznému oživení zájmu o tento typ překladu. ^[50]

Prvním krokem k překladu je zavedení algoritmu zarovnávání trénovacích vět po slovech, jak je tomu využito v implementaci GIZA++. Při samotném překladu systém hledá, jakým způsobem je možné rozdělit větu na fráze, které je pak možno dle tabulky přeložit. Fráze na cílové straně jsou na v případě mírného přeuspořádání spojeny a výsledná hypotéza věty je prověřena jazykovým modelem, který má za úkol kontrolovat hladkost výstupní formulace. ^[50]

8.5 Transliterace

Transliterace je proces konverze znaků z jedné abecedy do druhé, jehož výsledkem je přibližný fonetický přepis v cílovém jazyce. Příkladem můžou být v řečtině psaná slova „Φόβος“ a „Δεῖμος“, jež po provedení transliterace v latince označují „Phobos“ a „Deimos“. Transliterace tak umožňuje zpětnou rekonstrukci slov do původního jazyka. Termín transliterace může znamenat jak reverzibilní nebo nevratnou transformaci textu. ^[51]

Transliterace není to samé jako transkripce. Transkripce se speciálně zaměřuje na mapování zvuků jednoho jazyka do zvuků jiného jazyka. Stále však existují systémy transliterace, které mapují znaky zdrojového jazyka do znaků výslovnosti cílového jazyka. V případě, že vztahy mezi znaky a zvuky jsou téměř stejné, transliterace může být to samé jako transkripce. ^[51]

8.5.1 Užití transliterace

Řecká abeceda používá 24 písmen, která se kryjí s 26 znakovou verzí římské abecedy (latinky), ve které je psaná angličtina. Etymologové v anglických slovnících jsou často schopni identifikovat slova, která původně pocházejí z řečtiny. U přepisu slov z řecké abecedy do římské je právě využito transliterace.

[51]

8.5.2 Varianty transliterace

Pro potřeby transliterace se využívá mnoho systémů, tzn., každý systém může stejný text transliterovat jiným způsobem. Klasická transliterace řeckého písmene vypadá takto, $\varphi \rightarrow ph$, zatímco s využitím alternativního systému UNGEGN vypadá takto, $\varphi \rightarrow f$. [51]

Projekt CLDR¹⁸ stanovuje obecné mapování znaku do znaku, například mezi Cyrilicí a Latinkou, nebo také specifické mezi Ruštinou – Francouzštinou a Srbštinou – Němčinou. V tomto projektu je řešena i částečně obecná transliterace mezi rusko-latinskými znaky a mezi znaky v cyrilici a francouzštině. Převod znaků z ostatních jazyků do Latinky se nazývá *Romanizace*¹⁹. [51]

8.5.3 Pravidla pro transliteraci

Pro provedení transliterace existuje celá řada obecně platných postupů, které však nejsou vždy dodržovány současně. Proto tak dochází ke kombinaci pravidel na

¹⁸ Unicode Common Locale Data Repository je nejrozsáhlejší úložiště dat, která jsou využita celé řadě programů při jazykové lokalizaci (formátování data, časových pásem, formátování čísel, výběr jazyků a zemí dle jména)

¹⁹ V lingvistice je pod tímto pojmem označována reprezentace psaného slova nebo mluvené řeči v Latince. Metody romanizace můžou být například transliterace nebo transkripce.

základě požadavků na transliteraci. Tyto požadavky jsou především důležité pro osoby, které transliteraci provádí.

Níže lze vidět obecné zásady pro Unicode CLDR transliteraci:

- *Standardní*: následuje používané systémy (standarty, předlohy) kdykoliv je to možné a pozměňuje je příležitostně, pokud je to nezbytné pro reversibilní přepis. V CLDR se systémy obecně popisují v komentářích v datových souborech XML.
- *Kompletní*: každá dobře zformovaná sekvence znaků ve zdrojovém písmu by měla být transliterována do sekvence znaků v cílovém písmu a naopak
- *Předvídatelná*: samotné znaky (bez znalostí zápisu znaku v jazyce) by měly být dostatečné pro transliteraci, která je založena na relativně malém počtu pravidel. Tento způsob umožňuje mechanický přepis.
- *Vyslovitelná*: výsledné znaky udávají výslovnost cílových znaků. Transliterace není tak přesná, pokud proces mapování znaků nehledí na výslovnost.
- *Reverzibilní*: je možné zobrazit text v původních znacích po provedení transliterace z cílových znaků. ^[51]

8.5.4 ČSN ISO 9

Tato norma zavádí systém transliterace znaků cyrilice do latinky v souladu se zásadami řetězové konverze, která umožní mezinárodní výměnu informací, zejména elektronickými prostředky. Cílem normy je poskytnutí prostředků pro mezinárodní komunikaci psaných zpráv ve formě, která umožní automatický přenos a obnovu zpráv mezi lidmi a stroji. Systém konverzí musí být v tomto

případě jednoznačný a zcela reverzibilní. Nebere ohled na fonetickou stránku, která se ignoruje u strojové konverze. Přijetím této normy má každá země možnost přijmout pro vlastní potřeby normu, která se může lišit, pod podmínkou, že bude kompatibilní s mezinárodní normou. Takto navržený systém musí toto umožňovat a je přijatelný pro mezinárodní použití, jestliže v něm obsažené grafické znaky jsou takové povahy, že mohou být automaticky konvertovány na grafické znaky používané v kterémkoliv přesně vymezeném národním systému.

[52]

8.5.5 GOST 16876-71

GOST je systém romanizace (pro transliteraci textů v Cyrilici do Latinky), který vzniknul v Národní správě pro geodézii a kartografii v bývalém Sovětském svazu. Je založen na podobných principech jako ostatní systémy používané při transliteraci ve vědě. GOST byl navržený tak, aby reflektoval všechny jazyky, které se nacházely na území tehdejšího Sovětského svazu. Tato norma byla třikrát revidována a to v roce 1973, 1980 a 1983, kdy byly prováděny menší úpravy.

V roce 1978 s drobnými úpravami tento standart přijala Rada vzájemné a hospodářské pomoci pod oficiálním názvem SEV 1362-78. Norma byla také přijata na páté konferenci OSN v Montrealu roku 1987, kdy se v OSN začal používat systém romanizace zeměpisných názvů.

V roce 2002 však Ruská federace spolu s ostatními zeměmi tento standard opustila ve prospěch ISO 9, který přijala jako standard GOST 7.79-2000. [53] [54]

Písma použitá v praktické části

Většina vědců považuje za počátek písma okamžik, kdy konkrétní znak označoval jazykovou jednotku, tj. celé slovo, slabiku nebo hlásku. Jiní vědci naopak argumentují tím, že za počátek písma se dají již považovat obrázková sdělení o nějaké skutečnosti bez návaznosti na konkrétní jazyk. Aby se tímto způsobem lidé dorozumívali, muselo se použít nejzákladnějších obrázků. ^[55]

Ve vývoji písma došlo nejvíce k posunu například v Mezopotámii, ve starém Egyptě, Číně a ve střední Americe. Došlo ale ke ztrátě jazykové univerzálnosti, kvůli zobrazení složitých pojmů, místních jmen. Mnoho současných písem nemá spojitost se smyslem sdělení, ale předává pouze jeho znění symboly.

Další krok vývoje písma představoval rozbití slova na menší jednotky, na slabiky, což u mnoha jazyků způsobovalo jisté problémy, protože pojem slabiky byl pro mluvčího neznámý nebo nepochopitelný. Snáze o slabikách se mluví při pravidelném střídání souhlásek a samohlásek, což v japonštině nebyl problém než například u němčiny, češtiny. Objev slabiky vedl k jejímu zápisu systémem akrofonie²⁰. Ze symbolu pro slovo se tak stal symbol pro počáteční slabiku. Tímto způsobem vznikla řada slabičných písem. ^[55]

²⁰ Jedná se o systém, kdy se „klikyháku“ označující věc nebo zvíře se přiřadila první slabika slova, která danou věc či zvíře měla charakterizovat.

Neméně obtížným krokem bylo rozbití slabiky na jednotlivé hlásky. Příčiny těchto problémů jsou hluboké a souvisí s funkcí mozkových hemisfér a neuronových sítí. Dle názorů fyziologů jsou samostatné hlásky zpracovány hlavně pravou mozkovou hemisférou, slova a slabiky zpracovává levá hemisféra, která se označuje za „řečovou“. Čtení a psaní ve slabičném písmu probíhá většinou v jedné mozkové kůře a to řečové, oproti hláskovému písmu, kde je slabičné písmo dále štěpeno na hlásky a dochází tak k zapojení obou mozkových částí.

V dalším kroku vývoje písma vzniklo tzv. souhláskové písmo, kdy západosemitské souhláskové písmo vedlo k vývoji téměř všech zvukových hláskových písem světa. ^[55]

Dle názorů řady vědců, souhláskové písmo vzniklo jakýmsi nedopatřením daným strukturou semitských a hamitských jazyků včetně staré egyptštiny. V těchto jazycích je významová složka soustředěna v souhláskách, zatímco samohláskami je význam modifikován. Tuto vnitřní reflexi má do jisté míry každý jazyk, např. v angličtině sing – sang – sung – song.

Za nejstarší souhláskové písmo je považováno foinické písmo, které se vyvinulo souběžně s písmem protosinajským. Z foinického písma vznikla naprostá většina i v minulosti používaných písem. Vzniklo například řecké písmo a z něj pravděpodobně prostřednictvím etruské abecedy písmo latinské, slovanské (cyrilické) nebo germánské. Z foinického písma vychází přes písmo aramejské většina písem indických, písem jihovýchodní Asie, arabské. ^[55]

Sami Řekové ve svých bájích uvádějí, že řecké písmo, alfabet, které je mimochodem základem pro všechna evropská písma, se vyvinulo z písma foinického. Z toho písma Řekové převzali všechny potřebné souhlásky, nepotřebné

vypustili nebo přeměnili na znaky pro samohlásky. Ve skutečnosti určili foinickým slabikám funkci souhlásek, což z hlediska mozkových funkcí bylo činem vskutku revolučním. Další potřebná písmena pro jazyk časem vytvořili, jiná opustili. Těmito úpravami řecké písmo dosáhlo svého současného tvaru. ^[55]

9.1 Latinka

Latinské (římské) písmo patří bezesporu mezi písmena italická. Díky příznivému vývoji v dějinách latinské písmo (latinka) vyniklo tak, že přežilo pád Římské říše a stalo se základem většiny novodobých písemných soustav.

Latinové, kteří dali jméno střeidoitalskému Latiu (dnešní Lazio), byli italskou národností a patřili k nim i Římané. Za mýtické datum patří rok 753 př. n. l., kdy vznikl Řím. Následný společenský vývoj vyvolal potřebu široké vzdělanosti, která nemohla existovat bez vhodného písemného systému. ^[55]

Největší vliv na vznik písmo měli sousedé Latinů, neindoevropští Etruskové. Řím jako takový byl původně etruskou osadou. Důležité pro vznik písmo byl fakt, že jak Řekové, tak Etruskové v 8. Století již vokalizované písmo znali a používali.

Nejstarší latinské nápisy byly ještě psány stylem zprava doleva. Záhy ale převládl směr zleva doprava a příslušně byla ustálena i orientace písmen. ^[55]

Vzhled latinského písmo navazuje na západní varianty řecké alfabety nebo na variantu euboickou, což je potvrzeno způsobem užití X, ale také dalšími písmeny, jako např. C, D, P, V, jejichž tvary jsou blízké západořeckým písmenům. Pouze písmena /Θ/ – théta, /Φ/ – phi a /Ψ/ – psi nebyly převzaty do rané latinské abecedy.

Latina je základem většiny novodobých písemných systémů na celém světě. Latinské písmo má omezený počet písmen, cca 26 – 30 znaků (dle úhlu pohledu), ale je doplňováno v národních soustavách znaky diakritickými, spřežkami nebo dalšími způsoby.

Málo známá je historická výslovnost latinského písma, protože od počátku středověku se klasická latina stala mrtvým jazykem a byla vyslovována různými způsoby. Jazyky, které se vyvinuly z hovorové latiny, procházely rychlým vývojem nejen v gramatice ale také po zvukové stránce. Novodobé románské jazyky se výslovností značně odlišily od klasické latiny. Nejblíže původnímu jazyku je italština. Nejvíce se vzdálila francouzština, která byla romantizována Galy a Germány. ^[55]

9.2 Alfabet

Řekové jsou právem označováni za vynálezce plně hláskového písma, které se oprostilo od předchozích slabičných či souhláskových soustav.

Řecké písmo je potomkem písma foinického (neboli kanaánského, jelikož Foinikie je jen řeckým překladem slova Kanaán, „země purpuru“), což potvrzuje řecká báje²¹. ^[55]

²¹ V této pověsti vyplul Kadmos, syn foinického krále Agénora, z přístavu Tyros, aby našel svou sestru Európu, kterou unesl Zeus v podobě býka na Krétu. Přišel do Boiótie, založil město Théby a uvedl do Řecka foinické písmo. Také Hérodotos uvádí, že „od Foiničanů se Řekové naučili mimo jiné i písmu, které jak soudím, dříve neznali“, a nazývá řecká písmena foinickými.

Tuto skutečnost potvrzují původní tvary řeckých písem, které se nelišily od své foinické předlohy, ale také názvy, jež ve foiničtině dávají smysl, ale v řečtině nikoliv. Pořadí písmen je téměř stejné a také směr psaní „zprava doleva“. ^[55]

Skutečná výslovnost antické řečtiny není známa. Ze způsobů převzetí znaků od Foiničanů a předání jiných znaků Etruskům a Římanům (později Slovanům) plynou nepřesné závěry. Například písmeno Y, vzniklé jako vokalizační [u] bylo původně vyslovováno jako [u], tedy *upsilon*. Později se výslovnost posunula k [ü] a varianta Y byla převzata do latiny jako [y] pro řecká slova s řeckým ypsilon. Obdobně probíhala úprava výslovnosti u dalších znaků.

Řecká abeceda ovlivnila písma dalších národů, zejména písma staré Itálie, etruské, latinské, venetské nebo slovanská. ^[55]

Jednotná abecední norma byla přijatá roku 403 př. n. l. Velkou zásluhou na přijetí spadala na rétory (Ísokratés) nebo filosofy (Platón, Aristoteles).

Počátkem 4. století př. n. l. vznikla klasická řecká abeceda, západní verze byla převzata Etrusky a stala se zárodkem pozdější latinky a dalších abeced starověké Itálie. Tato abeceda čítala 24 písem, z toho 7 samohláskových, 17 souhláskových a nadbytečná písmena *diagamma (vau)*, *koppa* a *san (sampi)* zůstala v systému jen pro označení číslic. ^[55]

9.3 Cyrilice

Stejně jako hlaholice byla cyrilice vytvořena v 9. století za účelem evangelizace slovanských zemích, zpočátku Velké Moravy. Již před nástupem hlaholice nebo cyrilice na slovanských územích probíhaly pokusy o zápis slov pomocí řeckého či

latinského písma bez úpravy. Jednalo se však o primitivní pravopis, který nebyl schopen pochytit bohatství slovanského hláskosloví. ^[55]

Záhadou ale zůstává, proč vznikla téměř současně dvě písma, která byla navzájem odlišná vzhledem, ale téměř shodná principy, pořadím hlásek v modifikované řecké řadě i názvy jednotlivých písem. Není také jisté, které písmo vlastně vzniklo nejdříve a jestli lze přičítat věrozvěstu Konstantinovy – Cyrilovi autorství obou soustav. Název „cyrilice“ je poněkud matoucí, jelikož hlaholice byla nejspíše počinem sv. Cyrila, skutečným misionářským písmem, zatímco cyrilice byla víceméně modifikací existující řecké abecedy pro jazyky slovanské.

Vzorem pro uspořádání jednotlivých písem cyrilice byla zřejmě hlaholice. Cyrilice je téměř přesným převodem stanovených pravidel do odlišné podoby znaků. Pořadí písmen u hlaholice i cyrilice odpovídá až na určité výjimky řecké abecedě. ^[55]

Cyrilice je známá spíše jako abeceda pravoslaví díky návaznosti na řeckou abecedu. Činnost Cyrila a Metoděje přispěla ke stabilizaci formy písma a k masivnímu rozšíření na území Balkánu a ruských knížectví.

Vývoj ruské cyrilice probíhal zprvu živelně. V letech 1707 – 1710 byly reformou Petra Velikého stanovena závazná pravidla, kdy z cyrilice byla vyloučena nepotřebná písmena. Jednalo se většinou o ta, která byla užívána ve slovech řeckého původu. : *psi*, *fita* (théta), *yžica*, *otʙ* (ómega) a některá písmena změnila svou výslovnost. Reforma byla ale zmírněna a byly vyloučeny pouze nosovky *psi* a *ómega*.

V průběhu několika následujících století došlo opět k vyloučení několika dalších písmen. Změnil se také i vzhled jednotlivých písmen, ornamentální formy iniciál,

vzhled kurzivy apod. V 18. – 20. století vznikla forma psacích písem, které se s menšími obměnami používají dodnes. ^[55]

9.4 Mezinárodní fonetická abeceda

Mezinárodní fonetická abeceda (IPA) je abecední systém fonetické notace, která primárně vychází z Latinské abecedy. Tento systém jako standardizovanou reprezentaci mluveného jazyka navrhla Mezinárodní Fonetická Asociace. Fonetickou abecedu využívají studenti, kteří studují zahraniční jazyk, učitelé, lingvisti, zpěváci, překladatelé apod. Cílem IPA je zachycení mluveného projevu lidí různých jazykových skupin tak, aby byl foneticky správný. ^[56]

IPA je navržena tak, aby obsahovala pouze ty vlastnosti, které charakterizují způsob řeči v mluveném jazyce, jedná se o intonaci, oddělení slova od slabiky, hlásku ale také i zubní skřípání, šišlání, tyto vlastnosti jsou pak součástí rozšíření IPA. Obecným principem IPA je poskytnout jedno písmeno pro každý typický zvuk, tzn., nepoužívají se kombinace písmen pro jeden zvuk. ^[56]

Symbole IPA jsou složeny z jednoho nebo více elementů dvou základních typů, písmena nebo diakritiky. Pro příklad, zvuk anglického písmena „t“ může být přepsáno ve fonetické abecedě jako [t] pomocí jednoho písmena nebo písmena a diakritiky [t^h]. V roce 2008 IPA obsahovala 107 písmen, 52 symbolů pro diakritiku a 4 prozodické značky. ^[56]

Písmena, která se používají ve fonetické abecedě, vycházejí většinou z latinky nebo řeckého písma. Nicméně existují i případy, kdy Mezinárodní fonetická asociace připustí písmena, která nevycházejí z těchto písem.

IPA používá dva druhy závorek pro transkripci: ^[56]

- [hranaté závorky] – používají se pro fonetické znázornění výslovnosti
- /lomítka/ - pro označení fonémů, které jsou charakteristické v jazyce

9.5 Užití IPA

IPA nabízí více než 160 symbolů pro transkripci řeči, přesto se používá malá část. Řeč je možné přepsat v různých úrovních přesnosti. Přesný fonetický přepis, ve kterém je velké množství detailu, se označuje jako „úzký přepis“. Hrubší přepis, který ignoruje některé detaily, se nazývá „hrubý přepis“. Obě tyto formy přepisů se vyskytují v hranatých závorkách. ^[56]

Velké množství Britských slovníků, zejména Oxford Advanced Learner's Dictionary nebo Cambridge Advanced Learner's Dictionary používá mezinárodní fonetickou abecedu k vyjádření výslovnosti slov. Ve většině Amerických i některých Britských děl se používá systém přepisu výslovnosti, který je pro čtenáře pohodlnější, tzn. znak <y> v IPA je vyjádřen [j] nebo <sh> jako [ʃ]. Snahou je reflektovat společnou reprezentaci zvuků v psané angličtině pouze pomocí písmen anglicko-římské abecedy. ^[56]

9.6 Znaký IPA

Písmena v mezinárodní fonetické abecedě jsou rozdělena do tří kategorií:

- plicní souhlásky,
- souhlásky,
- samohlásky.

Každému znaku je přiděleno číslo, aby nedošlo k záměně u podobných písmen.

Kategoriím zvuků jsou přiřazeny různé rozsahy čísel. Tyto kategorie souhlásek se liší svou výslovností. ^[56]

Část II

Praktická část

Praktické využití transliterace

Tato část práce bude zaměřená na prezentaci transliterace na konkrétním příkladu. Pro tento účel byla vytvořena jednoduchá webová aplikace, která je dostupná na <http://transliteration.tode.cz/>. Zde si uživatel může zadat URL stránku do připraveného formuláře a zvolit si z / do jaké abecedy se má transliterace provést. Stránka se následně v nezměněném layoutu vypíše. Transliterace v tomto případě je prováděna mezi znaky latinky, cyrilice a řecké abecedy.

Pro vytvoření aplikace bylo využito HTML, CSS , PHP, SQL, webového serveru Apache 2.2, databáze MySQL 5.5.19, nástroje pro správu databáze PHPMyAdmin 3.4.9 a vývojového prostředí PHP JetBrains PhpStorm 3.0.2.

10.1 Popis řešení

Aby se transliterace mohla provést, je nejprve nutné obsah webové stránky načíst do paměti. Pomocí nativní podpory DOM parseru s využitím DOM API v PHP 5 dochází k manipulaci s HTML souborem. Cílem je nahrazení jednotlivých znaků textového obsahu mezi elementy. Po nahrazení znaků dochází k vypsání stránky z paměti s modifikovaným obsahem.

10.1.1 Hlavní soubor - index.php

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN">
<html>
  <head>
    <title>Transliteration</title>
  </head>
```

Tato stránka je stránkou úvodní, přes kterou má možnost uživatel webovou aplikaci obsluhovat.

Stránka je složena z hlavičky, ve které je deklarace `DOCTYPE` a elementu pro název stránky `<title>Transliteration</title>`.

```
<frameset rows="13%,87%" border="1">
  <frame name="top" src="pages/head.php" class="top" />
  <frame name="bottom"/>
</frameset>
</html>
```

Dále je vidět, že stránka je ve skutečnosti složena ze dvou stránek pomocí rámců. 13% velikosti stránky zabírá `head.php` a 87% velikosti, do které se vypíše obsah stránky, u které došlo k transliteraci.

Transliterate:

From: To:

Obrázek 14: Hlavní stránka, index.php

10.1.2 Soubor s formulářem - head.php

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN">
<html>
<head>
    <meta http-equiv="content-type" content="text/html;
charset=charset=utf-8">
    <link rel="stylesheet" href="../../css/theme.css" type="text/css"
media="screen">
</head>
```

Stránka slouží pro zobrazení formuláře, kam si uživatel může vložit URL webové stránky a zvolit si z / do jaké abecedy se má transliterace provést.

V hlavičce stránky je uvedena deklarace `DOCTYPE`, kódování stránky `charset=charset=utf-8` a odkaz na CSS soubor `link rel="stylesheet" href="../../css/theme.css"`.

```
<body>
<form class="forms" action="main.php" method="POST" target="bottom">
    <label class = "label" for="url">Transliterate:</label>
    <input type="text" name="url" value="" size = "80" id="url"
class="input"/><br /><br />
    <label for="transFrom">From:</label>
```

V sekci `<body>` se nalézá samotný formulář. Data z formuláře jsou odeslána pomocí metody `POST` `method="post"` a následně jsou zpracována souborem `main.php` `action="main.php"`.

10.1.3 Soubor pro manipulaci s dokumentem a transliteraci - main.php

Pro práci s dokumentem je nejprve nutné ošetřit správnost vyplnění formuláře.

```
<?php
$url = $_POST['url'];
$z=$_POST['transFrom'];
$do=$_POST['transTo'];
```

Nejprve dochází k definici proměnných `$url`, `$z`, `$do`.

```
if ($_POST['send']=='true') {
    $errors=array();
```

Pro výpis chybových hlášek se připraví pole, do kterého se budou chyby zapisovat.

```
function url_exists($url)
{
    $headers = @get_headers($url);
    if(preg_match("|200|",$headers[0]))
        return(true);
    else return(false);
}
```

Funkce `url_exists` má za cíl ověřit platnost zadané URL tím, že testuje, zda daná stránka obsahuje v hlavičce http status kód 200, což je odpověď pro úspěšný http požadavek, neboli značí existenci stránky.

```
if (empty($url)){
    $errors[] = 'URL address is not fill';
}
```

Ověření, zda `$url` obsahuje nějakou hodnotu. Pokud ne, vypíše se chybová hláška, že URL adresa není vyplněná.

```
if (($z==$do) && (url_exists($url)==true)) {  
    $errors[] = "Resource and target alphabets are the same";  
}
```

Podmínka testuje, zda cílová abeceda se shoduje se zdrojovou. Za podmínky, že funkce `url_exists` vrátí `true`, neboli, zadaná stránka existuje, vrátí chybovou hlášku s informací, že jsou abecedy shodné.

```
if ((url_exists($url)==false) && (!empty($url))){  
    $errors[] = "URL address is not exist";  
}
```

Ověření platnosti URL. Za podmínky, že proměnná URL obsahuje nějakou hodnotu, zobrazí se chybová hláška, že URL adresa neexistuje.

```
$error = join('<br />', $errors);
```

Po zkontrolování pole ve formuláři se chybové hlášky vypíší do jedné proměnné.

```
if(!empty($error)){  
  
    echo '<div class="error">' . $error . '</div>';  
    exit();  
}  
}
```

Pokud se vyskytnou při zpracování formuláře nějaké chyby, dochází k jejich vypísání a ukončení zpracování dokumentu.


```
require "../scripts/db.php";
```

Do souboru `main.php` je vložen soubor `db.php`, který obsahuje skript pro připojení k databázi.

```
$dom = new DOMDocument();  
$dom->preserveWhiteSpace = false;  
$file = file_get_contents($_POST['insert']);  
$dom->loadHTML($file);
```

Aby se mohlo s dokumentem pracovat, je nejprve nutné v paměti vytvořit DOM reprezentaci celého dokumentu `$dom = new DOMDocument().preserveWhiteSpace = false` slouží pro vypuštění uzlů s bílými znaky.

Do proměnně `$file` se uloží obsah webové stránky, jejíž URL uživatel vložil do formuláře. Následně se v paměti vytvoří DOM reprezentace dané stránky.

Na průchod celého dokumentu byla vytvořena rekurzivní funkce, která jako parametr očekává uzel, který má zpracovat. Pokud uzel má další děti, funkce je také zpracuje.

```
Walk($dom->documentElement, true);  
  
function Walk($node, $replace)  
{  
    $ignoredTags = array("style", "script");  
  
    if ($node->nodeType == XML_ELEMENT_NODE)  
    {  
        echo "<" . $node->nodeName;  
        foreach ($node->attributes as $attr)  
        {
```

Nejdříve se spustí průchod stromem `Walk($dom->documentElement, true)`. Funkce `function Walk($node, $replace)` s parametry `$node`, `$replace` vypíše aktuální uzel.

Do proměnné `$ignoredTags = array("style", "script")` se načtou uzly, u kterých je nežádoucí nahrazování znaků z jedné abecedy do druhé.

Pokud je typ uzlu `element XML_ELEMENT_NODE`, vypíše se "<" a připojí název elementu `.$node->nodeName`. Všechny elementy jsou postupně zpracovány příkazem `foreach`. V dalším kroku se pak budou do elementu doplňovat atributy a jejich obsah.

```
$isAbsolutePath = substr($attr->value, 0, 4) == "http" ? true : false;
$parsed_url = parse_url($_POST['url']);
    $domain_url = $parsed_url['scheme'].'://'.$parsed_url['host'] .
    '/';
    if ($attr->name == "src" && !$isAbsolutePath) $attr->value =
    $domain_url . $attr->value;
    if ($attr->name == "href" && !$isAbsolutePath) $attr->value =
    $domain_url . $attr->value;
echo ' '. $attr->name . '="' . $attr->value . '"';
}
echo ">";
```

Jelikož dochází k uložení webové stránky mimo server, na kterém se zdrojová stránka nachází, je třeba relativní adresy, které se nacházejí na stránce (například odkaz na CSS soubory) převést na absolutní. Pro identifikaci absolutní adresy slouží proměnná `$isAbsolutePath`, kdy se do této proměnné načte obsah atributu a zároveň se zjišťuje, zda na prvních čtyřech pozicích řetězce se vyskytuje `http`.

V dalším kroku dochází k parsování URL. Cílem je z adresy, která má podobu <http://www.xy.cz/test/hra.php> dostat pouze URL domény, tzn. <http://www.xy.cz/>. Dále je třeba relativní adresy převést na absolutní. Převod adres je relevantní pro atributy `src` a `href`. Úprava probíhá tak, že do hodnoty atributu se uloží URL domény `$attr->value = $domain_url` a následně se k této hodnotě připojí zbytek obsahu pomocí `. $attr->value`.

```
if ($node->nodeType == XML_TEXT_NODE || $node->nodeType ==
XML_CDATA_SECTION_NODE
    ) {
    $text = $node->nodeValue;
```

Nyní se přistoupí k výpisu obsahu textových uzlů. Nejdříve se do proměnné `$node` vypíší uzly s konstantami `XML_TEXT_NODE` a `XML_CDATA_SECTION_NODE`, tzn. uzly, u kterých bude docházet k výpisu textového obsahu. Pro výpis textu se využívá vlastnosti uzlu `nodeValue`, obdoba `textContent`.

```
if (!$replace)
{
    $newText = $text;
}
```

Pro atributy s parametrem `!replace` (`style`, `script`) nedochází k nahrazování znaků a do proměnné `$newText` se tak vypíše obsah atributu beze změny. Do této proměnné se jinak ukládá modifikovaný text po nahrazení znaků.

```
else
{
    $z=$_POST['transFrom'];
    $do=$_POST['transTo'];
```

Do proměnných `$z` a `$do` se přiřadí hodnoty z pole formuláře pomocí proměnných `$_POST['transFrom']` a `$_POST['transTo']`, které obsahují uživatelem zvolenou abecedu pro provedení transliterace.

```
$query = "SELECT $z, $do FROM transtab";
```

Do proměnné `$query` se vypíše dotaz pro výběr záznamů z tabulky `transtab`.

```
$result = mysql_query($query);
```

Provedení dotazu na aktuální spojení v databázi a vypsání výsledků do proměnné `$result`.

```
$vyh = array();  
$nah = array();
```

Inicializace polí `$vyh` a `$nah`.

```
while ($row = MySQL_Fetch_array($result)) {  
    $vyh[] = $row[$z];  
    $nah[] = $row[$do];  
}
```

Pro vypsání všech záznamů z databáze do polí se využije cyklu `while ($row = MySQL_Fetch_array($result))`.

```
$newText = str_replace($vyh, $nah, $text);  
}
```

Pro samotné nahrazování znaků je využito funkce `str_replace()`, která nahrazuje všechny výskyty jednoho řetězce dalším řetězcem. Funkce pracuje se třemi parametry. První parametr slouží pro vyhledání hodnot v poli, pomocí druhého parametru se nahradí znaky, které se vyhledají na základě prvního parametru. Ve

třetím parametru je specifikována proměnná, ve které se nalézá textový obsah stránky, na který se tato funkce aplikuje. Modifikovaný text se přiřadí do proměnné `$newText`.

```
        echo $newText . "";  
    }
```

Nyní dojde k vypsání modifikovaného textu na obrazovku.

```
if ($node->hasChildNodes()){  
    foreach($node->childNodes as $subNode){
```

Pokud uzel má nějaké děti, dochází k jejich zpracování. Výstupem metody `hasChildNodes()` je `true`.

```
        if (!in_array($node->nodeName, $ignoredTags)){  
            Walk($subNode, true);  
        }  
        else {  
            Walk($subNode, false);  
        }  
    }  
}
```

Děti jsou zpracovány za předpokladu, že nejsou v pozici rodičů a neobsahují nežádoucí tagy pro nahrazování znaků.

```
if ($node->nodeType == XML_ELEMENT_NODE)  
{  
    echo "</" . $node->nodeName . ">\n";  
}  
}  
?>
```

Pro element se nakonec vypíše koncový tag.

10.1.4 Soubor s transliterační tabulkou – transtab.sql

Pro sestavení transliterační tabulky, která je v databázi, bylo využito systémů pro transliteraci cyrilice do latinky ISO 9:1995 (GOST 7.79 System A, B), pro transliteraci alfabety do latinky ISO 843 a transliterační tabulky doc. Vojtěcha Merunky.

```
CREATE TABLE IF NOT EXISTS `transtab` (  
  `id` int(10) unsigned NOT NULL AUTO_INCREMENT,  
  `lat` varchar(10) CHARACTER SET utf8 COLLATE utf8_bin NOT NULL,  
  `cyr` varchar(10) CHARACTER SET utf8 COLLATE utf8_unicode_ci NOT NULL,  
  `alfb` varchar(10) CHARACTER SET utf8 DEFAULT NULL,  
  PRIMARY KEY (`id`),  
  UNIQUE KEY `id` (`id`)  
) ENGINE=MyISAM DEFAULT CHARSET=latin1 AUTO_INCREMENT=126 ;
```

SQL skript pro vytvoření tabulky.

```
INSERT INTO `latcyr` (`id`, `lat`, `cyr`, `alfb`) VALUES  
(1, 'A', 'A', 'A'),  
(2, 'Á', 'A', 'A'),  
(3, 'B', 'B', 'Мп'),  
(4, 'C', 'Ц', 'тσ'),  
(5, 'Č', 'Ч', 'Tζ'),  
(6, 'Ch', 'Ч', 'X'),  
(7, 'D', 'Д', 'Δ'),  
(8, 'Ď', 'дб', 'Δ'),  
(9, 'E', 'E', 'E'),  
(10, 'E', 'Э', 'E'),  
(11, 'É', 'E', 'E'),
```

Ukázka vložení části znaků do transliterační tabulky. Celou tabulku je možno vidět v příloze 3.

10.1.5 Soubor se skriptem pro připojení k databázi - db.php

```
<?php
$dbUser = 'root';
$dbPwd = 'heBgUb9D';
```

Definice přihlašovacích údajů k databázi.

```
spojeni a nazev db
global $dbConn;
$dbUse = 'transliteration';
```

Definice proměnné `$dbConn` pro spojení s databází a používané databáze `$dbUse`.

```
    if (!($dbConn = mysql_connect('localhost', $dbUser, $dbPwd))) {
        die ('Nepodařilo se připojit k databázovému serveru');
    }
```

Připojení k databázovému serveru.

```
    if(!(mysql_select_db($dbUse, $dbConn))) {
        die ('Nepodařilo se připojit k databázi');
    }
```

`mysql_select_db ()` nastaví aktuální databázi na severu a provede asociaci databáze s aktuálním identifikátorem spojení.

```
mysql_set_charset("utf8", $dbConn);
?>
```

Pro aktuální spojení se nastaví kódování znaků `utf8`.

10.1.6 Soubor s formátováním stránky – theme.css

```
.forms {
  margin-top: 20px;
  /*nastavení horního vnějšího okraje*/
  margin-left: 20px;
  /*nastavení levého vnějšího okraje*/
  font-family: Verdana;
  /*nastavení stylu písma*/
  font-size: 14px;
  /*nastavení velikosti písma*/
}
```

Třída `forms` pro formátování formuláře.

```
.input {
  height: 30px;
  /*nastavení výšky pole pro zadání URL*/
  font-size: 14px;
  padding-left: 10px;
  /*nastavení vnitřního okraje pole*/
}
```

Třída `input` pro formátování pole formuláře.

```
.error {
  margin-left: 20px;
  font-family: Verdana;
  font-size: 11px;
  font-weight: bold;
  /*nastavení tučného písma*/
}
```

Třída `error` pro formátování oddílu pro zobrazování chyb.

10.1.7 Ukázka výsledné podoby stránek po transliteraci

Novinky.cz

Hledej

[Главній stránка](#) [Стано се](#) [Домáці](#) [Ваше зпрáвы з региону](#) [Захраничні](#) [Крими](#) [Культура](#) [Економика](#) [Финанце](#) [Спорт](#) [Жена](#)

[Коктейл](#)

[Интернет а ПЦ](#) [AutoMoto](#) [Въда а школы](#) [Быдлені](#) [Карьера](#) [Цестовáні](#) [Тема](#) [Специáлы](#) [Почасі](#) [Хороскоп](#) [ТВ програм](#) [Денній тиск](#) [Емайлем](#) [Киндлем](#)
[Видео](#)

Зпрáвы дне



ФОТО: Иван Хандрейцк, [Право](#)

[Вѣтшина лиди́ цхе предчасне волбы, укázалы прузкумы](#)

4. 4. [Актуализовáно](#) Вѣтшина обывател се в рьцхлNULLцх прузкумех высловила про конáни́ предчаснNULLцх волб. Подле агентуры Медиан, ктерá прузкум выпрацовала про ТВ Прима, ѝе про toto решени́ стáвайи́ ци́ ситуацие 80 процент лиди́. [ЦелNULL члáнек »](#)



[Їохн оточил: Министри демиси зати́ м неподайи́, ОДС ѝе прNULL встри́ пná](#)

4. 4. [Актуализовáно](#) Прeдседа Вѣлї верейнNULLцх Радек Їохн пред



[До утерNULL цхци ми́ т ѝасно, рекл Нечас. Цхе зрушит три министерства](#)

4. 4. [Актуализовáно](#) Прeмїер а шеф ОДС Петр Нечас по стредечни́ м ѝеднáни



[УпрїхлNULL подникател Пїтр уж ѝе ве вѣзници в Брнї](#) [► Видео](#)

4. 4. [Актуализовáно](#) ОдсоузенNULL подникател Томáш Пїтр, ктерNULL од свехо затчени́ в червенци предлонскего

Obrázek 15: Novinky.cz v Cyrilici

Ζπράβυ δνε



ΦΟΤΟ: Iav χανδρείσυχ, Πράβο

Βιετσζινα λιδί τοχτισε πρεδτζασνέ βολμπυ, ουκάζαλυ προυζκουμυ

4. 4. Ακτουαλιζοβάνο Βιετσζινα ομπυβατελ σε βρυτοχλύτοχ προυζκουμετοχ βυςλοβιλα προ κονάνι πρεδτζασνύτοχ βολεμπ. Ποδλε αγεντουρυ Μεδιαν, κτερά προυζκουμυ βυπρατσοβαλα προ ΤΒ Πριμα, ιε προ τοτο ρεσζενί ζτάβαϊϊτοί ριτουατσε 8ο προτσεντ λιδί. [τοελύ τζλάνεκ »](#)



Ίοχν στοτζύλ: Μινιστρι δεμειζι ζατιμ νεποδαϊί. ΟΔΣ ιε πρύ βετριτονά

4. 4. Ακτουαλιζοβάνο Πρεδσεδα Βιετσι βερεϊνύτοχ Ραδεκ Ίοχν πρεδ ζτρεδετζνίμ ιεδνάνιμ βεθενί ζτρανυ Πράβου ρεκλ, ζχε ιεχο μινιστρι ζατιμ δεμειζε νεποδαϊί, στονα...



Δο ουτερύ τοχτσι μίτ ιασνο, ρεκλ Νετζας, τοχτισε ζρουσζιτ τρι μινιστερστβα

4. 4. Ακτουαλιζοβάνο Πρεμιέρ α σζέφ ΟΔΣ Πετρ Νετζας πο ζτρεδετζνίμ ιεδνάνι γρέμμα ζτρανυ ουβεδλ, ζχε ο λοκρατζοβάνι βλάνυ ταντσο μίτ ιασνο δε ουτερύ το...



Ουπρτοχλύ ποδνικατελ Πιτρο ουζυ ιε βε βιεζνιτσι β Μπρνιε ► Βιδεο

4. 4. Ακτουαλιζοβάνο Οδσοουζενύ ποδνικατελ Τομάσζ Πιτρ, κτερύ οδ εβέχο ζαττζενί β τζερβεντσι πρεδλονκchéχο ροκου ρεδιελ βε σζβύτσαρσκέ βαζμπιε, ιμπυλ βε στορεδου ζα ποϊενέγο...

Obrázek 16: Novinky.cz v Alfabetě

Závěr

Webové technologie a služby se stále více dostávají do popředí a představují jakýsi standard pro stále rozvíjející se generaci e-businessu. Pro tuto domněnku hovoří fakt, viz kapitola [6.4.1], že v současné době lze na Internetu nalézt množství aplikací, které se funkčně vyrovnají ryze desktopovým aplikacím. Velkou výhodou webových aplikací je možnost k aplikaci přistupovat odkudkoliv přes webový prohlížeč a tím pádem odpadá i nutnost jakékoliv instalace aplikace. Důkazem může být například společnost Microsoft, IBM a další, kdy svoje produkty uvedly na Internet v podobě webové služby.

Díky novým webovým technologiím je možné dnes už prakticky cokoli. Lidé mohou spolu udržovat sociální kontakt, publikovat, obchodovat, sdílet informace atd. Největší zásluhou na tom má bezpochyby vznik Web 2.0, který je určitým milníkem v novodobé historii Internetu. Web 2.0 také zjednodušuje možnost využívat moderní technologie, zvláště při tvorbě webových stránek.

Pro demonstraci využití moderních webových technologií byla autorem vytvořena webová aplikace, kterou bude zejména přínosná pro občany bývalého Sovětského svazu, kterým tato služba umožní porozumět textu slovanského jazyka na internetové stránce. Předpokládá se, že tito občané budou mít částečné nebo žádné znalosti latinky, která jim právě bude znesnadňovat přečíst si daný text. Příkladem může být Rus neznalý latinky, který přijede do České republiky a bude

si chtít přečíst události, které se tu udály. Díky této službě bude mít možnost porozumět textu. Dále je tato služba využitelná u široké veřejnosti, protože se jí dostává nástroje, který umožňuje jednoduše zjistit podoby znaků v uvedených abecedách.

Obecně má transliterace sloužit k usnadnění čtení textu v jiné abecedě bez znalosti potřebných abeced. Proto vznikla řada systémů pro transliteraci, např. GOST 16876, který, jak již bylo uvedeno v kapitole [8.8.5], byl použit při konverzi zeměpisných názvů. Zdaleka se ale nevyužívá právě jen u těchto záležitostí. Často je využita i u přepisu jmen osob. Dále transliterace poskytuje prostředky pro mezinárodní komunikaci psaných zpráv ve formě, která umožní automatický přenos zpráv mezi lidmi a stroji. U tohoto případu je důležité, aby systém byl jednoznačný a zcela reverzibilní. Právě reverzibilita u jednotlivých systémů a abeced může být problém. Při rekonstrukci původního textu tak není dosaženo shody. Typickým příkladem můžou být třeba čínské znaky, které výrazně převyšují znaky kteréhokoliv hláskového písma.

Použitá literatura

1. Berners-Lee, Tim. The original proposal of the WWW. *W3.org*. [Online] CERN, květen 1990. [Citace: 2. červenec 2011.]
<http://www.w3.org/History/1989/proposal.html>.
2. Berners-Lee, Tim a Cailiau, Robert. WorldWideWeb: Proposal for a HyperText Project. *W3.org*. [Online] 12. Listopad 1990. [Citace: 2. červenec 2011.]
<http://www.w3.org/Proposal.html>.
3. Tim Berners-Lee: WorldWide Web, the first client. *W3.org*. [Online] [Citace: 2. červenec 2011.]
4. Cowell, Jude. August 6 1991: first-ever webpage online. *Wordpress.com*. [Online] 26. březen 2010. [Citace: 3. červenec 2011.]
<http://judecowell.wordpress.com/2010/03/26/august-6-1991-first-ever-webpage-online/>.
5. Ten Years Public Domain for the Original Web Software. *Cern.ch*. [Online] [Citace: 10. červenec 2011.] <http://tenyears-www.web.cern.ch/tenyears-www/Declaration/Page1.html>.
6. Web Architecture. *W3.org*. [Online] [Citace: 10. 07 2011.]
<http://www.w3.org/standards/webarch/>.
7. Berners-Lee, Tim. Frequently asked questions. *W3.org*. [Online] [Citace: 12. červenec 2011.] <http://www.w3.org/standards/webarch/>.

8. Castellucio, Michael. It's not your grandfather's Internet. *Findarticles.com*. [Online] Říjen 2010. [Citace: 12. Červenec 2011.] http://findarticles.com/p/articles/mi_hb6421/is_4_92/ai_n56479358/.
9. Ben-Itzhak, Yuval. Infosecurity 2008 - New defence strategy in battle against e-crime. *Computerweekly.com*. [Online] 18. Duben 2008. [Citace: 15. Červenec 2011.] <http://www.computerweekly.com/Articles/2008/04/18/230345/Infosecurity-2008-New-defence-strategy-in-battle-against.htm>.
10. Chisholm, Wendy, Vanderheiden, Gregg a Jacobs, Ian. W3C Recommendation. W3. [Online] 5. Květen 1999. [Citace: 15. červenec 2011.] <http://www.w3.org/TR/WAI-WEBCONTENT/>.
11. Sikos, Leslie. Website standardization. *Lesliesikos*. [Online] [Citace: 15. červenec 2011.] <http://www.lesliesikos.com/services/standardization/>.
12. XML DOM Introduction. *W3school.com*. [Online] [Citace: 16. červenec 2011.] http://www.w3schools.com/dom/dom_intro.asp.
13. Koch, Peter-Paul. The Document Object Model: an Introduction. *Digital-web.com*. [Online] 2001. květen 2001. [Citace: 19. 07 2011.] http://www.digital-web.com/articles/the_document_object_model/.
14. Flanagan, David. *JavaScript Pocket Reference*. Second Edition.: O'Reilly Media, Inc., 2002. str. 122. 978-0-596-00411-8.
15. Bartlett, Kynn, Cohen, Gail T. a Edlund, Jens. Frequently Asked Question. *Webstandards.org*. [Online] 2002. [Citace: 20. srpen 2011.] <http://www.webstandards.org/learn/faq/>.
16. Pixley, Tom. Document Object Model (DOM) Level 2 Events Specification. *W3.org*. [Online] 13. Listopad 2000. [Citace: 23. srpen 2011.] <http://www.w3.org/TR/2000/REC-DOM-Level-2-Events-20001113/>.
17. Whitmer, Ray. Document Object Model (DOM) Level 3 XPath Specification. *W3.org*. [Online] 26. únor 2004. [Citace: 25. srpen 2011.] <http://www.w3.org/TR/DOM-Level-3-XPath/>.

18. Wolter, Jan. Javascript Madness: Layout Engines. *Unixpapa.com*. [Online] 30. březen 2011. [Citace: srpen. 25 2011.] <http://unixpapa.com/js/gecko.html>.
19. Sharma, Prashant. Core Characteristics of Web 2.0 Services. *Techpluto.com*. [Online] červen 2009. [Citace: 20. srpen 2011.] <http://www.techpluto.com/web-20-services/>.
20. Graham, Paul. Web 2.0. *Paulgraham.com*. [Online] listopad 2005. [Citace: 21. srpen 2011.] <http://www.paulgraham.com/web20.html>.
21. O'Reilly, Tim. Web 2.0 Compact Definition: Trying Again. *radar.oreilly.com*. [Online] 12. říjen 2006. [Citace: 25. srpen 2011.] <http://radar.oreilly.com/archives/2006/12/web-20-compact.html>.
22. Zbiejczuk, Adam. *Web 2.0-charakteristika a služby*. 2007.
23. MacManus, Richard. Web 2.0 Definition and Tagging. *ReadWriteWeb.com*. [Online] 1. únor 2009. [Citace: 1. srpen 2011.] http://www.readwriteweb.com/archives/web_20_definiti.php.
24. O'Reilly, Tim. What Is Web 2.0. *oreilly.com*. [Online] 30. září 2005. [Citace: 26. srpen 2011.] <http://oreilly.com/web2/archive/what-is-web-20.html>.
25. Blood, Rebecca. Weblogs: A History and Perspective. *Rebecablood.net*. [Online] 07. září 2000. [Citace: 26. srpen 2011.] http://www.rebeccablood.net/essays/weblog_history.html.
26. Marlow, Cameron. *Audience, structure and authority in the weblog*. MIT Media Laboratory, květen 2004.
27. Wainwright, Phil. Why Microsoft can't best Google. *Zdnet.com*. [Online] 24. srpen 2005. [Citace: 27. srpen 2011.] <http://www.zdnet.com/blog/saas/why-microsoft-cant-best-google/13>.
28. Jones“Mart, K. C. Gmail Now Has IMAP Support. *Informationweek.com*. [Online] 24. říjen 2007. [Citace: srpen. 28 2011.] Gmail Now Has IMAP Support.

29. Google Gets the Message, Launches Gmail. *Google.com*. [Online] 1. duben 2004. [Citace: 27. srpen 2011.] <http://www.google.com/press/pressrel/gmail.html>.
30. Picasa and Picasa Web Albums Help. *Google*. [Online] [Citace: 28. srpen 2011.] <http://picasa.google.com/support/bin/answer.py?hl=en&answer=39567>.
31. Wei, Coach K. AJAX: Asynchronous Java + XML? *Developer.com*. [Online] 11. srpen 2005. [Citace: 30. srpen 2011.] <http://www.developer.com/design/article.php/3526681>.
32. Configuring other mail clients. *Google*. [Online] [Citace: 29. srpen 2011.] <https://mail.google.com/support/bin/answer.py?hlrm=cs&answer=13287>.
33. *Excerpt from interview with Google CEO Eric Schmidt*. 30. srpen 2011.
34. Kaste, Martin. Facebook's Newest Challenger: Google Plus. *Npr.org*. [Online] 29. červen 2011. [Citace: 2. září 2011.] <http://www.npr.org/2011/06/29/137507567/facebooks-newest-challenger-google-plus>.
35. Google+ grows to 10 million. Google+ grows to 10 million users. *Money.cnn.com*. [Online] 14. červen 2011. [Citace: 2. září 2011.] http://money.cnn.com/2011/07/14/technology/google_q2_earnings/index.htm.
36. Frequently Asked Questions. *Last.fm*. [Online] [Citace: 3. září 2011.] <http://www.last.fm/help/faq>.
37. Arrington, Michael. Exclusive: Screen Shots And Feature Overview of Delicious 2.0 Preview. *Techcrunch.com*. [Online] 6. září 2007. [Citace: 29. srpen 2011.] <http://techcrunch.com/2007/09/06/exclusive-screen-shots-and-feature-overview-of-delicious-20-preview/>.
38. delicious blog. *delicious.com*. [Online] [Citace: 29. srpen 2011.] <http://blog.delicious.com>.
39. Help. *MySpace.com*. [Online] [Citace: 1. září 2011.] http://myspace2.custhelp.com/app/answers/detail/a_id/290/related/1.

40. eko. Murdochova News Corp. prodává MySpace. Na investici prodělala půl miliardy dolarů. *Ihned.cz*. [Online] 29. červen 2011. [Citace: 2. září 2011.] <http://byznys.ihned.cz/zpravodajstvi-svet/c1-52202920-murdochova-news-corp-prodava-myspace-na-investici-prodelala-pul-miliardy-dolaru>.
41. Viola Web Browser. *Viola.org*. [Online] [Citace: 2. září 2011.] Viola Web Browser.
42. Hamilton, Naomi. The A-Z of Programming Languages: JavaScript. *Computerworld.com*. [Online] 31. červen 2008. [Citace: 2. září 2011.] http://www.computerworld.com.au/article/255293/-z_programming_languages_javascript.
43. Garrett, Jesse James. Ajax: A New Approach to Web Applications. *Adaptivepath.com*. [Online] 18. únor 2005. [Citace: 2. září 2011.] <http://www.adaptivepath.com/ideas/ajax-new-approach-web-applications>.
44. Hajičová, Eva, Panevová, Jarmila a Sgall, Petr. *Úvod do teoretické a počítačové lingvistiky*. Praha : Nakladatelství Karolinum, 2003. 80-246-0470-1.
45. Černý, Jiří. *Malé dějiny lingvistiky*. Portál, 2005. str. 240. 80-7178-908-9.
46. Allen, James. *Natural Language Understanding*. 2nd. Addison Wesley, 1994. str. 654. 978-0805303346.
47. Pala, Karel. *Počítačové zpracování přirozeného jazyka*. Brno : Fakulta informatiky Masarykovy university, 2000.
48. Čermák, František a Koček, Jan. Co je korpus? *Korpus.cz*. [Online] [Citace: 10. září 2011.] http://www.korpus.cz/co_je_korpus.php.
49. Hutchins, John. *MT News International no.8*. 8. květen 1994. 0965-5476.
50. John Hutchins, et. all. *Encyclopedia of Language & Linguistics*. [editor] Keith Brown. 2nd.: Elsevier Science, 2006. str. 9000. 978-0080442990.

51. Unicode Transliteration Guidelines. *CLDR - Unicode Common Locale Data Repository*. [Online] [Citace: 15. září 2011.] <http://cldr.unicode.org/index/cldr-spec/transliteration-guidelines>.
52. *Informace a dokumentace - Transliterace cyrilice do latinky - slovanské a neslovanské jazyky*. 010185 1. září 2002.
53. Rules of Cyrillic alphabet letters. *Hostmaster.net.ua*. [Online] [Citace: 2. říjen 2011.] <http://www.hostmaster.net.ua/docs/?gost16876-71>.
54. HOST 7.79-2000. *Ruscable.ru*. [Online] 12. únor 2002. [Citace: 2. říjen 2011.] <http://gost.ruscable.ru/cgi-bin/catalog/catalog.cgi?i=6464>.
55. Mojdl, Lubor. *Encyklopedie písem světa*. První. Praha : Libri, 2005. str. 191. 80-7277-288-0.
56. Association, International Phonetic. *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet*. Cambridge University Press, 1999. str. 214. 978-0521637510.

Seznam obrázků

Historické logo WWW [Robert Cailiau].....	14
Grafické znázornění webu podle [Examiner.com]	14
NeXT Computer [Cern.ch].....	16
Web 2.0 Mincloud podle [Markus Angermeier]	25
Meme Map podle [O'Reilly Media]	27
Gmail podle [TechNewsWorld]	34
Google + podle [Rightyleft.com]	35
Last.fm podle [Songbirdnest.com].....	36
Delicious podle [Crowdshifter.com]	37
MySpace podle [Wired.com]	38
Viola podle [Viola.org]	41
Trojúhelník strojového překladu s ilustracemi existujících systémů podle [Ondřej Bojar, CUNI].....	53
Ukázka hloubkové korespondence mezi českou a anglickou větou podle [Ondřej Bojar, CUNI]	55
Hlavní stránka, index.php	71
Novinky.cz v Cyrilici.....	84
Novinky.cz v Alfabetě	85

Část III

Přílohy

I	II	Transkripce		I	II	Transkripce	
Д	Д А	a	azъ	Ф	Ф ф	f	fertъ
Б	Б в	b	buky	Ѡ	Ѡ Ѡ	th/f	(fita)
В	В в	v	vědě	Х	Х х	ch	chěrbъ
Г	Г г	g	glagoľъ	Ѡ	Ѡ w	o	otъ
Д	Д А	d	dobro	Ѡ	Ѡ Ѡ	št/šč	šta
Е,Е	Е е	e/je	jestъ	Ц	Ц ц	c	ci
Ж	Ж ж	ž	živěte	Ч	Ч ч Ѹ ѹ	č	ča
С	С с Ѻ ѻ	dz	dzělo	Ш	Ш ш	š	ša
З	З з Ѻ ѻ	z	zemlja	Ъ	Ъ ъ	tvrdý znak	jetъ
І	І і ĭ	i	i	Ы, Ы	Ы ы	y	jery
Н	Н н	i	iže	Ь	Ь ь	měkky znak	jetъ, jerokъ
Н	Н н	ġ/dʹ/j	děrvъ/jot	Ѣ	Ѣ ѣ	je/ja	jať
К	К к	k	kako	Ю	Ю ю	ju	(ju)
Л	Л л	l	ljudъje	Ѡ	Ѡ Ѡ Ѡ Ѡ	ja	(ja)
М	М м	m	myslite	К	К к	je	(je)
Н	Н н	n	našъ	Ѡ	Ѡ Ѡ	ę	ęsъ
О	О Ѡ О Ѡ	o	otъ	Ж	Ж ж	o	osъ
П	П п	p	rokojъ	Ѡ	Ѡ Ѡ	ję	jęsъ
Р	Р р	r	rъci	Ѡ	Ѡ Ѡ	jo	josъ
С	С с	s	slovo	Ѡ	Ѡ Ѡ	locké ě	ksí
Т	Т т	t	tvrdо	Ѡ	Ѡ Ѡ	locké ѡ	psí
У	У Ѹ ѹ (8)	u	ukъ	Ѡ	Ѡ Ѹ ѹ	locké o	yžica

Пříloha 1: Několik typů staré cyrilice podle [Lubor Mojdl]

Ruština			Ukrajinština			Běloruština			Bulharština			Makedonština			Srbština		
А	а	a	А	а		А	а		А	а		А	а		А	а	
Б	б	b	Б	б		Б	б		Б	б		Б	б		Б	б	
В	в	v	В	в	v, (v̆)*	В	в		В	в		В	в		В	в	
Г	г	g	Г	г	h	Г	г	h	Г	г		Г	г		Г	г	
			Ґ	ґ	g							Ґ	ґ	g/d'	Ђ	ђ	dž
Д	д	d	Д	д		Д	д		Д	д		Д	д		Д	д	
Е	е	je/e	Е	е	e _{neměkl}	Е	е	je/e	Е	е	e _{neměkl}	Е	е	e _{neměkl}	Е	е	e _{neměkl}
Ё	ё	jo'o	Є	є	je/e	Ё	ё	jo'o		ьо	'o						
Ж	ж	ž	Ж	ж		Ж	ж		Ж	ж		Ж	ж		Ж	ж	
												С	с	dz			
												Ц	ц	dž	Ц	ц	dž
З	з	z	З	з		З	з		З	з		З	з		З	з	
И	и	i/i	И	и	yivrdé				И	и	i _{neměkl}	И	и	i _{neměkl}	И	и	i _{neměkl}
			І	і	i/i	І	і	i/i									
			Ї	ї	ji												
Й	й	j	Й	й		Й	й		Й	й		Ј	ј	j	Ј	ј	j
К	к	k	К	к		К	к		К	к		К	к		К	к	
												К	к	k/t	Ђ	ђ	ć
Л	л	l	Л	л		Л	л		Л	л		Л	л		Л	л	
												Љ	љ	l	Љ	љ	l
М	м	m	М	м		М	м		М	м		М	м		М	м	
Н	н	n	Н	н		Н	н		Н	н		Н	н		Н	н	
												Њ	њ	ñ	Њ	њ	ñ
О	о	o	О	о		О	о		О	о		О	о		О	о	
П	п	p	П	п		П	п		П	п		П	п		П	п	
Р	р	r	Р	р		Р	р		Р	р		Р	р		Р	р	
С	с	s	С	с		С	с		С	с		С	с		С	с	
Т	т	t	Т	т		Т	т		Т	т		Т	т		Т	т	
У	у	u	У	у		У	у		У	у		У	у		У	у	
						Ў	ў	u									
Ф	ф	f	Ф	ф		Ф	ф		Ф	ф		Ф	ф		Ф	ф	
Х	х	ch	Х	х		Х	х		Х	х		Х	х		Х	х	
Ц	ц	c	Ц	ц		Ц	ц		Ц	ц		Ц	ц		Ц	ц	
Ч	ч	č	Ч	ч		Ч	ч		Ч	ч		Ч	ч		Ч	ч	
Ш	ш	š	Ш	ш		Ш	ш		Ш	ш		Ш	ш		Ш	ш	
Щ	щ	šč	Щ	щ		Щч	щч		Щ	щ	št						
Ъ	ъ	ivrdý zn.	()	()	misto ь	()	()	misto ь	Ъ	ъ	ə (ä)						
Ы	ы	y				Ы	ы										
Ь	ь	měkký zn	Ь	ь		Ь	ь		Ь	ь							
Э	э	e _{neměkl}															
Ю	ю	ju'u	Ю	ю		Ю	ю		Ю	ю							
Я	я	ja'a	Я	я		Я	я		Я	я							

Příloha 2: Současná cyrilice – azbuka slovanských národů podle [Lubor Mojdl]

Foinická písmena	Transkripce a název	Řecká písmena
Ⲁ Ⲁ	ʾ <i>ʾaleph</i>	Α
ⲁ	b <i>beth</i>	Β
Ⲃ Ⲃ	g <i>gimmel</i>	Γ
ⲃ ⲃ	d <i>daleth</i>	Δ
Ⲅ	h <i>he</i>	Ε
ⲅ ⲅ	w <i>waw</i>	(F <i>vau/digamma</i>) Υ [u/w/ü]
Ⲇ Ⲇ	z <i>zajin</i>	Ζ
ⲇ	ḥ <i>heth</i>	Η
Ⲉ Ⲉ	tʿ <i>teth</i>	Θ
ⲉ ⲉ	j <i>jod</i>	Ι
Ⲇ Ⲇ Ⲇ	k <i>kaph</i>	Κ
Ⲍ Ⲍ Ⲍ	l <i>lamed</i>	Λ
ⲍ ⲍ ⲍ ⲍ	m <i>mem</i>	Μ
Ⲏ	n <i>nun</i>	Ν
ⲏ ⲏ	s <i>samekh</i>	Ξ
Ⲑ Ⲑ Ⲑ	ʿ <i>ʿajin</i>	Ο, Ω
ⲑ	p <i>pe</i>	Π
Ⲓ Ⲓ Ⲓ	s <i>sade</i>	(Μ <i>san</i>)
ⲓ ⲓ	q(kʿ) <i>qoph</i>	(Ϝ <i>koppa</i>)
Ⲕ	r <i>reš</i>	Ρ
ⲕ ⲕ	š/ś <i>šin</i>	Σ
Ⲗ Ⲗ Ⲗ	t <i>taw</i>	Τ

Příloha 3: Srovnání foinického písma řecké abecedy podle [Lubora Mojdl]

†	Ɱ	Ɱ	Ɱ	Ɱ	Ɱ	Ɱ	Ɱ
А	Б	В	Г	Д	Е	Ж	З
a	b	v	g	d	e,je	ž	dz
Ɱ	Ɱ	Ɱ	Ɱ	Ɱ	Ɱ	Ɱ	Ɱ
З	И	И	Ɱ	К	Л	М	Н
z	i	i	ǵ,d'	k	l	m	n
Ɱ	Ɱ	Ɱ	Ɱ	Ɱ	Ɱ	Ɱ	Ɱ
О	П	Р	С	Т	У	Ф	Х
o	p	r	s	t	u	f	ch
Ɱ	Ɱ	Ɱ	Ɱ	Ɱ	Ɱ	Ɱ	Ɱ
Ѡ	Ѣ	Ѥ	Ѧ	Ѩ	Ѭ	Ѯ	Ѱ
o	c	f	č	š	y	ь	y
Ɱ	Ɱ	Ɱ	Ɱ	Ɱ	Ɱ	Ɱ	Ɱ
Ь	Ѣ	Ю	Ѡ	Ѣ	Ѥ	Ѧ	Ѩ
b	ě,ja	ju	č,k,št	ř	ř	ě	je

Jen cyrilské znaky : Ɱ Ɱ
ksi psi

Hlaholice a cyrilice. Abecedu se učili žáci těmito názvy:
a - az; b - buky; v - vědi; g - glagolъ; d - dobro; e, je - jestъ; ž - živěte; dz - dzělo;
z - zemlja; i - iže; i; ǵ, d' - děrv; k - kako; l - ljudie; m - myslite; n - našъ; o - onъ;
p - pokojъ; r - rьci; s - slovo; t - tvrdo; u - ukъ; f - frьtъ; ch - chěrъ; o - otъ; c - ci;
f, t - fita (theta); č - črvъ; š - ša; y - yžica; ь - jer tvrdé; y - jery; ь - jer měkké;
ě, ja - jať; ju - jusъ; č, k' št - šta; ř - řsъ; ř - řsъ; ę - ęsъ; je - ješъ

Příloha 4: Abecedy staroslověnského písemnictví - první řádek hlaholice, druhý řádek cyrilice, třetí řádek běžný přepis výslovnosti podle [M. Beranová]

Příloha 3: Transliterační tabulka použitá při konverzi znaků

INSERT INTO `latcyr` (`id`, `lat`, `cyr`, `alfb`) VALUES			
(1, 'A', 'A', 'A'),	(21, 'Í', 'И', 'I'),	(41, 'Shh', 'ИИ', 'Σxx'),	(62, 'a', 'a', 'α'),
(2, 'A', 'Б', NULL),	(22, 'Ī', 'Ī', 'Ī'),	(42, 'T', 'T', 'T'),	(63, 'a', 'б', NULL),
(3, 'Á', 'A', 'A'),	(23, 'J', 'Й', 'Ī'),	(44, 'Th', 'Tr', 'Θ'),	(64, 'á', 'á', 'α'),
(4, 'B', 'B', 'Μπ'),	(24, 'K', 'K', 'K'),	(45, 'U', 'Y', 'Ου'),	(65, 'b', 'б', 'μπ'),
(5, 'C', 'И', 'τσ'),	(25, 'L', 'Л', 'Λ'),	(46, 'Ú', 'Y', 'Ου'),	(66, 'c', 'и', 'τσ'),
(6, 'Č', 'Ч', 'τς'),	(26, 'M', 'M', 'M'),	(47, 'Ů', 'Y', 'Ου'),	(67, 'č', 'ч', 'τς'),
(7, 'Ch', 'Ч', 'X'),	(27, 'N', 'H', 'N'),	(48, 'V', 'B', 'B'),	(68, 'ch', 'ч', 'χ'),
(8, 'D', 'Д', 'Δ'),	(28, 'Ň', 'H', 'N'),	(49, 'W', 'W', NULL),	(69, 'd', 'д', 'δ'),
(9, 'Ď', 'Дб', 'Δ'),	(29, 'O', 'O', 'O'),	(50, 'X', 'X', 'Ξ'),	(70, 'd', 'Дб', 'δ'),
(10, 'E', 'E', 'E'),	(30, 'Ó', 'O', 'O'),	(51, 'Y', 'B', 'Y'),	(71, 'e', 'e', 'ε'),
(11, 'E', 'Э', 'E'),	(31, 'Ō', 'O', 'Ω'),	(52, 'ÿ', 'V', NULL),	(72, 'è', 'э', NULL),
(12, 'É', 'E', 'E'),	(32, 'Ó', ' ', 'Ω'),	(53, 'Ÿ', 'Б', 'Y'),	(73, 'é', 'e', 'ε'),
(13, 'Ě', 'Б', 'IE'),	(33, 'P', 'П', 'Π'),	(54, 'Yu', 'Ю', 'υου'),	(74, 'ě', 'Б', 'ιε'),
(14, 'F', 'Ф', 'Φ'),	(34, 'Ps', 'Пс', 'Ψ'),	(55, 'Ya', 'Я', 'υα'),	(75, 'f', 'Ф', 'φ'),
(15, 'Fh', 'Θ', 'Φχ'),	(35, 'Q', 'Q', NULL),	(56, 'Ye', 'Б', 'υε'),	(76, 'fh', 'θ', 'φχ'),
(16, 'G', 'Г', 'Γ'),	(36, 'R', 'P', 'P'),	(57, 'Yh', 'V', 'υx'),	(77, 'g', 'г', 'γ'),
(17, 'H', 'X', 'χ'),	(37, 'Ř', 'P', 'P'),	(58, 'Yo', 'Ë', 'υο'),	(78, 'h', 'x', 'χ'),
(18, 'I', 'И', 'I'),	(38, 'S', 'C', 'Σ'),	(59, 'Z', 'З', 'Z'),	(79, 'i', 'и', 'ι'),
(19, 'Ī', 'Й', 'H'),	(39, 'Š', 'ИИ', 'Σ'),	(60, 'Ž', 'Ж', 'Z'),	(80, 'ī', 'й', 'η'),
(20, 'Ī', ' ', 'H'),	(40, 'Sh', 'ИИ', 'Σx'),	(61, ' ', 'Б', NULL),	(81, 'ī', ' ', 'H'),

(89, 'ň', 'h', 'v'),	(99, 's', 'c', 'ç'),	(113, 'ÿ', 'ÿ', 'ü'),	
(90, 'o', 'o', 'o'),	(100, 'š', 'm', 'σç'),	(114, 'yu', 'io', 'uou'),	
(91, 'ó', 'ó', 'ó'),	(103, 't', 't', 'τ'),	(115, 'ya', 'я', 'υα'),	
(92, 'ō', 'ō', 'ω'),	(104, 'ť', 'ř', 'τ'),	(116, 'ye', 'ѣ', 'υε'),	
(93, 'ó', 'ó', 'ó'),	(105, 'th', 'tr', 'θ'),	(117, 'yh', 'yh', 'υχ'),	
(94, 'p', 'π', 'π'),	(106, 'u', 'y', 'ou'),	(118, 'yo', 'yo', 'υο'),	
(95, 'ps', 'πc', 'ψ'),	(107, 'ü', 'y', 'ou'),	(119, 'z', 'z', 'ζ'),	
(96, 'q', 'q', NULL),	(108, 'ú', 'y', 'ou'),	(120, 'ž', 'ж', 'ζχ'),	
(97, 'r', 'p', 'q'),	(109, 'v', 'b', 'β'),	(121, ' ', 'b', NULL),	
(98, 'ř', 'p', 'q'),	(110, 'w', 'w', 'w'),	(122, 'y', 'b', 'υ'),	
(101, 'sh', 'm', 'çχ'),	(111, 'x', 'x', 'ξ'),	(123, 's', 'c', 'σ'),	
(102, 'shh', 'm', 'çχχ'),	(112, 'y', 'b', NULL),	(125, 'ý', 'NULL', 'ú');	