

Masarykova univerzita

Filozofická fakulta

Ústav českého jazyka

Český jazyk se specializací počítačová lingvistika



**Urážlivé, vulgární a rasistické projevy na
internetových diskusních fórech**

Bakalářská diplomová práce

Irena Diatelová

jaro 2013

Vedoucí práce:
Mgr. Marek Grác

Prohlášení

Prohlašuji, že tato diplomová práce je mým původním autorským dílem, které jsem vypracovala samostatně. Všechny zdroje, prameny a literaturu, které jsem při vypracování používala nebo z nich čerpala, v práci řádně cituji s uvedením úplného odkazu na příslušný zdroj.

Irena Diatelová

Na tomto místě chci poděkovat
Mgr. Marku Grácovi,
vedoucímu mé bakalářské diplomové práce, za trpělivost, cenné rady, odbornou pomoc a připomínky.
Dále bych ráda poděkovala
doc. PhDr. Zdeňce Hladké, Dr.,
doc. PhDr. Kláře Osolsobě, Dr.
a Mgr. Haně Žižkové
za cenné rady a připomínky při vylepšování práce.

Anotace

Bakalářská práce se zabývá urážlivými, vulgárními a rasistickými komentáři na internetových diskusních fórech. Rozřazuje je z hlediska témat, ale i vzniku – podrobněji se zaměřuje na sémantické a slovo-tvorné procesy.

Podstatnou část práce tvoří abecedně řazený slovník zachycených výrazů z diskusních fór. Hesla slovníku doplňuje význam, použití v kontextu a lexikografické zmapování pomocí čtyřech českých slovníků: Slovník spisovné češtiny, Slovník spisovného jazyka českého, Slovník nespisovné češtiny a Slovník české frazeologie a idiomatiky.

Závěr práce obsahuje teoreticky popsaná pravidla pro program, který by dokázal odhalit urážlivé, vulgární a rasistické výrazy, ale i výrazy nově utvořené. V návaznosti na tuto práci vznikne program fungující podle pravidel popsaných v závěrečné kapitole.

Klíčová slova

urážky, urážlivá slova, urážlivé výrazy, vulgarismy, vulgární slova, vulgární výrazy, sprostá slova, rasistické výrazy, xenofobní výrazy, seznam urážek, seznam vulgarismů, program vyhledávající nevhodné diskusní příspěvky, program vyhledávající vulgarismy v internetových diskusích

Annotation

This bachelor thesis deals with insulting, vulgar and racist comments on internet discussion forums. It classifies them according to theme of subject, but also according to their origins – with detailed focus on semantic and word-formation processes.

A significant part of the thesis represents alphabetically arranged dictionary of expressions captured on the discussion forums. Key words in dictionary are accompanied by their meaning, their usage in context and lexicography mapping using four czech dictionaries: The Dictionary of Standard Czech, Dictionary of the Standard Czech Language, Dictionary of the non-standard Czech and Dictionary of Czech Phraseology and Idiomatics.

The closing part contains theoretically described rules for program, which should be able to detect insulting, vulgar and racist expressions, but also expressions newly formed. Following this work there will be created a program operating under the rules described in the final chapter.

Key words

insults, insulting words, insulting expressions, vulgarisms, vulgar expressions, rude words, racist expressions, xenophobic expressions, list of insults, list of vulgarisms, program searching for inappropriate forum posts, program searching for vulgarisms on internet discussion forums

Obsah

Obsah	5
Seznam zkratek	7
1 Úvod	9
2 Internetová diskusní fóra	11
2.1 Internet a jeho historie	11
2.2 Online diskusní fórum	12
2.3 Netiketa	13
2.4 Moderované diskuse	13
2.5 Typy diskusí na zpravodajských serverech	14
2.5.1 Diskuse bez registrace	14
2.5.2 Diskuse s registrací	15
2.5.3 Diskuse využívající automaty	17
2.5.4 Diskuse, v nichž na závadné příspěvky dohlíží sami uživatelé a programy	17
2.5.5 Další řešení nevhodných diskusních příspěvků	18
2.6 Shrnutí	19
2.7 Závěr	20
3 Urážlivé, vulgární a rasistické projevy na internetových diskusních fórech	21
3.1 Kategorizace urážlivých, vulgárních a rasistických výrazů	21
3.1.1 Kontextově závislé	22
3.1.2 Sexuální tematika	22
3.1.3 Výrazy urážející inteligenci	23
3.1.4 Výrazy týkající se náboženství	23
3.1.5 Výrazy týkající se vzhledu	23
3.1.6 Výrazy urážející věk	23
3.1.7 Výrazy týkající se povahy	23
3.1.8 Výrazy označující osoby závislé na návykových látkách	24
3.1.9 Výrazy označující hloupost	24
3.1.10 Výrazy označující obličej	24
3.1.11 Výrazy označující podřadnou skupinu	24
3.1.12 Výrazy označující nemanželské nebo zlobivé dítě	24
3.1.13 Politické výrazy	24
3.1.14 Výrazy označující národnost	25
3.1.15 Xenofobní výrazy	25
3.1.16 Nejčastější a nejpoužívanější výrazy	25
3.1.17 Přezdívky známých osob	25
3.1.18 Jazyková hra se jmény známých osob	26
3.1.19 Grafické výrazy	28
3.1.20 Speciálně upravené výrazy kvůli blokovacím automatům	28

4	Vznik urážlivých, vulgárních a rasistických výrazů	29
4.1	Přejímání slov z cizích jazyků	29
4.2	Sémantické procesy	29
4.2.1	Významové přenosy (sémaziologicko-onomazeologická derivace)	29
4.2.2	Významové posuny	30
4.2.3	Změna konotačních složek	30
4.3	Slovotvorné procesy	31
4.3.1	Derivace	31
4.3.2	Kompozice	35
4.4	Víceslovná spojení	35
4.5	Závěr	36
5	Zpracování slovníku (komentář k příloze)	37
5.1	O slovníku	37
5.1.1	Získávání materiálu	37
5.1.2	Rozčlenění slovníku	37
5.1.3	Heslové slovo	38
5.1.4	Struktura hesla	38
5.2	Charakteristika užitých lexikografických děl	39
5.2.1	Slovník spisovného jazyka českého (SSJČ)	39
5.2.2	Slovník spisovné češtiny pro školu a veřejnost (SSČ)	39
5.2.3	Slovník české frazeologie a idiomatiky (SČFI)	40
5.2.4	Slovník nespisovné češtiny (SNČ)	41
5.2.5	Zhodnocení užitečnosti lexikografických příruček	41
5.3	Příznakovost lexikálních prostředků	44
5.3.1	Časové hledisko	44
5.3.2	Frekvenční hledisko	44
5.3.3	Teritoriální hledisko	44
5.3.4	Sociální hledisko	45
5.3.5	Funkční hledisko	45
5.3.6	Přenesení významu lexikálních jednotek	45
5.3.7	Hovorovost – knižnost	45
5.3.8	Nocionálnost – pragmatičnost	45
6	Návrh programu vyhledávající urážlivé, vulgární a rasistické výrazy	47
6.1	Databáze	47
6.1.1	Posouzení vhodnosti slov správcem	48
6.1.2	Ruční zjednoznačnění	48
6.1.3	Seznam problémových osobností	48
6.1.4	Seznam zakázaných slov	48
6.1.5	Zohlednění derivovaných slov	49
6.1.6	Zohlednění skloňování, časování	49
6.2	Práce s diskusními příspěvky	49
6.2.1	Sledování negativního hodnocení příspěvků uživateli	49
6.2.2	Sledování méně používaných slov, které se začaly objevovat ve větší míře	50
6.2.3	Sledování editační (Levenstheinovy) vzdálenosti	50
6.2.4	Rozpoznávání známých a neznámých slov	50
6.2.5	Práce se vstupním textem	50
7	Závěr	53
8	Seznam použité literatury	57
	Příloha	60

Seznam zkratek

<i>aj.</i>	a jiné, a jiní
<i>angl.</i>	anglický, angličtina
<i>apod.</i>	a podobně
<i>arg.</i>	argot, argotický výraz
<i>atd.</i>	a tak dále
<i>AV ČR</i>	Akademie věd České republiky
<i>brněn.</i>	brněnský výraz
<i>cca</i>	cirka, kolem, asi
<i>cm</i>	centimetr
<i>ČSAV</i>	Československá akademie věd
<i>ČSSD</i>	Česká strana sociálně demokratická
<i>ČR</i>	Česká republika
<i>dět.</i>	dětsky, dětský výraz
<i>drog.</i>	drogový slang
<i>expr.</i>	expresivně, expresivní výraz
<i>hanl.</i>	hanlivě, hanlivé slovo
<i>hovor.</i>	hovorově, hovorový výraz
<i>char.</i>	charakteristika
<i>kniž.</i>	knižně, knižní slovo
<i>KSČM</i>	Komunistická strana Čech a Moravy
<i>LJ</i>	lexikální jednotka
<i>mor.</i>	oblastní moravský výraz
<i>morav.</i>	moravský dialektický výraz
<i>např.</i>	například
<i>nář.</i>	nářečí, nářeční výraz
<i>něm.</i>	německý, němčina
<i>ob.</i>	obecně česky, výraz obecné češtiny
<i>obl.</i>	oblastní výraz
<i>ODS</i>	Občanská demokratická strana
<i>os.</i>	osoba
<i>oslab.</i>	oslabené
<i>ostrav.</i>	ostravský výraz
<i>pl.</i>	plurál, množné číslo
<i>pohrd.</i>	pohrdavě
<i>polic.</i>	policejní slang
<i>poněk.</i>	poněkud
<i>poněk. zast.</i>	poněkud zastarávající slovo
<i>popř.</i>	popřípadě, po případě
<i>posm.</i>	posměšně
<i>přen.</i>	přeneseně, v přeneseném užití
<i>publ.</i>	publicistický výraz
<i>r.</i>	roku
<i>resp.</i>	respektive, vlastně, popřípadě, lépe řečeno

<i>SČFI</i>	Slovník české frazeologie a idiomatiky
<i>sg.</i>	singulár, jednotné číslo
<i>SNČ</i>	Slovník nespisovné češtiny
<i>spol.</i>	společnost
<i>PSJČ</i>	Příruční slovník jazyka českého
<i>SPOZ</i>	Strana práv občanů – Zemanovci
<i>SSČ</i>	Slovník spisovné češtiny pro školu a veřejnost
<i>SSJČ</i>	Slovník spisovného jazyka českého
<i>str.</i>	strana, stránka
<i>škol.</i>	školní slang
<i>tj.</i>	to jest
<i>TOP 09</i>	politická strana (zkratka počátečních písmen slov tradice, odpovědnost, prosperita; číslovky 09 označují rok vzniku)
<i>tzv.</i>	tak zvaný
<i>tzn.</i>	to znamená
<i>vězeň.</i>	vězeňský slang
<i>vulg.</i>	vulgárně, vulgární výraz
<i>xenof.</i>	xenofobní výraz
<i>zast.</i>	zastarávající slovo
<i>zdrob.</i>	zdrobnělina
<i>zhrub.</i>	zhruběle, zhrubělý výraz
<i>zprav.</i>	zpravidla
<i>zvl.</i>	zvláště, zvláštního
<i>žert.</i>	žertovně, žertovný výraz

1. Úvod

Internetová diskusní fóra přinesla jazykovědcům velkou příležitost ke zkoumání jazyka. Zatímco dříve získávali poznatky o jazyku z různých literárních děl, odborných a publicistických textů nebo ze soukromých dopisů, dnes jim k tomu stačí jedině: spustit internet a na různých diskusních fórech zkoumat současnou slovní zásobu.

Předmětem zkoumání se stávají různá témata: jazyk internetových seznamek, slovní zásoba diskusních fór, porovnávání dvou a více jazyků, zkoumání teritoriálních příznakových prvků aj.

Anonymita uživatelů na internetových diskusích hraje velkou roli: nebojí se otevřeně říct svůj názor, mnohdy doprovázený urážkami, vulgarismy nebo rasistickými projevy. Moderátoři diskusí musí několikrát denně zasahovat proti nevhodným příspěvkům, mazat je a mnohdy blokovat problémovým uživatelům přístup do fór. Některým administrátorům diskusí pomáhají různé programy, které příspěvky se závadným obsahem odhalí, jiní musí pročítat všechny komentáře, aby objevili ty nevhodné.

Tato diplomová práce se zaměřuje na urážlivé, vulgární a rasistické projevy, které se objevují na diskusních fórech. Cílem práce je sestavit popis programu, který by byl schopen vyhledávat v diskusích takovéto výrazy.

Pro sestavení pravidel pro tento program nejdříve potřebujeme získat přehled o výrazech, které se v diskusích objevují. K tomu nám poslouží výrazy pocházející z diskusí pod českými zpravodajskými servery (*Aha.cz*, *Aktuálně.cz*, *Blesk.cz*, *Deník.cz*, *EuroZprávy.cz*, *iDnes.cz*, *iHned.cz*, *Lidovky.cz*, *Novinky.cz*, *ParlamentníListy.cz* aj.), které byly shromážděny po dobu osmi měsíců. Seznam těchto výrazů obsahuje abecedně řazený slovník v příloze, doprovázený významem, použitím výrazů v kontextu a lexikografickým zmapováním celkem ve čtyřech českých slovnících: *Slovníku spisovného jazyka českého* (SSJČ), *Slovníku spisovné češtiny* (SSČ), *Slovníku nespisovné češtiny* (SNČ) a v případě víceslovných spojení i *Slovník české frazeologie a idiomatiky* (SČFI).

S těmito zachycenými výrazy pracujeme v průběhu celé práce: pokusíme se zjistit, jestli jsou to slova nově vzniklá a nezachycená žádným ze slovníků, nebo jsou to výrazy, které se ve slovní zásobě vyskytují dlouhodobě. Tuto skutečnost zjistíme pomocí výše zmíněných lexikografických příruček.

Dále se tyto výrazy pokusíme rozřadit podle témat, ale i podle vzniku. Podrobněji se zaměříme hlavně na vznik těchto lexikálních jednotek, důraz klademe hlavně na sémantické a slovtvorné procesy, u nichž doufáme, že by nám mohly pomoci nevhodné komentáře najít. Pokusíme se zjistit, zda nejsou problémová slova tvořena typickými prefixy či sufixy a zda bude efektivní takové prefixy či sufixy ve vstupním textu vyhledávat – tedy jestli program nalezne více závadných výrazů než výrazů bezproblémových.

K sestavení pravidel programu nám však nebudou sloužit pouze jazykové poznatky, nýbrž i doprovodné jevy, které se u nevhodných komentářů vyskytují. Na tyto jevy se zaměříme v poslední kapitole a pokusíme se je zanést do pravidel, podle nichž bude program pracovat.

Na tuto práci bude navazovat program, který se bude řídit sepsanými pravidly. S jeho pomocí se bude zjišťovat úspěšnost sepsaných pravidel a taky to, jestli bude možné některá pravidla vůbec použít.

Diskusní servery projevují i jinou snahu, jak omezit urážlivé, vulgární a rasistické projevy. Nepokouší se o sestavení programů, které dokážou takové výrazy vyhledat, ale spíše o odrazení problémového uživatele. Způsobům, jaké servery používají k omezení uživatelů, se věnuje následující kapitola.

Rádi bychom zdůraznili, že tato diplomová práce si nečiní nárok na hloubkovou analýzu urážlivých, vulgárních a rasistických výrazů, ani na jejich podrobný proces vzniku. Snaží se pouze o shromáždění, rozřazení a elementární popis materiálu, který dále slouží k sepsání pravidel programu.

Vzhledem k tomu, že jsme nenašli práci podobného typu, ani literaturu zabývající se výhradně tímto tématem, opírali jsme se o několik zdrojů.

Prvním z nich byl internetový článek Jakuba Čížka (*Jak se diskutuje na českých zpravodajských webech*), podle nějž jsme strukturovali kapitolu shromažďující informace o typech diskusních fór na českých zpravodajských serverech. Jako doplňkový zdroj této kapitoly jsme používali i knihu Evy Jandové – *Čeština na www chatu*.

V kapitolách 4 (*Vznik urážlivých, vulgárních a rasistických výrazů*) a 5 (*Zpracování slovníku (komentář k příloze)*) jsme se opírali o diplomové práce Jany Mališové (*Slovotvorná expresivita*) a Marty Šabatové (*Příznaková slovní zásoba v soukromé korespondenci mládeže*).

Nápomocný nám byl i *Slovník nespisovné češtiny* sepsaný kolektivem Jana Huga. Náš slovník v příloze, který zachycuje řadu urážlivých, vulgárních a rasistických výrazů, převzal velké množství významů právě z Hugova slovníku.

2. Internetová diskusní fóra

2.1 Internet a jeho historie

Za největší vynález 21. století lze považovat internet, mnozí jeho objev dokonce připodobňují k vynálezu knihtisku.

Oxfordský online slovník definuje internet jako celosvětovou počítačovou síť poskytující různorodé informace a komunikační možnosti; skládá se z propojených sítí, které používají standardizované komunikační protokoly. (Oxford Dictionaries, 2010)

Přestože se dnes na světě najde jen malá část lidí, kteří se s internetem nikdy nesetkali, je tento objev poměrně mladý: jeho historie sahá k roku 1957, kdy Sovětský svaz vypustil Sputnik 1, první umělou družici Země. Americká vláda zareagovala na sovětský úspěch založením organizace ARPA (*Advanced Research Project Agency*) v roce 1958, jejím cílem byla podpora výzkumných projektů vedoucích k objevení nových technologií. (Leiner et al., 1997; Jandová, 2006, str. 13–14; Lapáček, 2000, str. 6–8)

Tato organizace vybudovala komunikační síť, která propojila nejdůležitější vojenská, vládní a univerzitní pracoviště. Síť byla spuštěna v roce 1969 pod názvem ARPANET (od roku 1990 známá pod označením internet). Zpočátku spojovala jen pár počítačů, s postupem času se však jejich počet zvyšoval. Síť byla využívána hlavně ke komunikaci pomocí elektronické pošty nebo k pořádání videokonferencí.

Časem se síť rozšiřovala mezi univerzity a další organizace ve Spojených státech, v roce 1973 se připojila i londýnská univerzita nebo norský seismologický ústav.

I když původně síť nebyla určena široké veřejnosti, začala se mezi ni postupně rozšiřovat. Po zavedení služby WWW (*World Wide Web*) dokonce ještě více. V roce 1992 byl k internetu připojen více než jeden milion počítačů. Připojení České republiky k internetové síti se oficiálně datuje k 13. únoru 1992.

V dnešní době pokrývá internet celý svět. Už k němu nejsou připojeny jen počítače, ale i mobilní telefony, ledničky nebo jiná zařízení.

Internet v dnešní době slouží k tolika věcem, o kterých neměli jeho tvůrci ani zdání. Bezplatné telefonáty, placení účtů či nakupování z pohodlí domova nebo snadné vyhledávání informací pomocí internetového vyhledávače. Internet lidem zjednodušil život ve spoustě věcí.

Mnoho uživatelů využívá tohoto vynálezu ke komunikaci – ať už obchodní, vědecké nebo „jen tak pro zábavu“.

Existuje mnoho počítačových programů, které uživatelé využívají ke komunikaci: *jabber*, *skype*, *ICQ*. Ovšem mnozí dávají přednost komunikaci na internetových stránkách: mezi nejoblíbenější patří *facebook*, *twitter*, různé chatovací místnosti nebo internetová diskusní fóra.

A právě online diskusními fóry (a nevhodné komunikaci na nich) se v této práci zabýváme.

2.2 Online diskusní fórum

Online diskusní fórum je internetová stránka (nebo její část v případě zpravodajských serverů), která slouží ke vkládání názorů, reakcí či dotazů internetových uživatelů. (Garažová, 2010)

Diskusní fóra mají svá konkrétní témata, o nichž uživatelé diskutují. Do diskusních fór může přispívat jakýkoliv uživatel, který se registruje (některé diskuse ani registraci nevyžadují). Počet uživatelů diskusních fór se může pohybovat v desítkách či stovkách i tisících.

Po přihlášení odesílají uživatelé své příspěvky, ty se zveřejňují na obrazovce spolu s příspěvky dalších uživatelů. Diskutéri mohou reagovat na všechny příspěvky navzájem. Nové příspěvky lze zobrazit buď časově (od nejstarších po nejnovější a naopak) nebo hierarchicky (větvení diskuse na vlákna). (Jandová, 2006, str. 17)

Šmahel hovoří o dvojím typu online komunikace: synchronní a asynchronní. Při synchronní komunikaci musí být k internetu připojeny komunikující strany ve stejném okamžiku, jde o komunikaci v reálném čase (online). Do této skupiny spadají například chaty (z angl. *chat* – mluvit). (Šmahel, 2003, str. 102)

Oproti synchronní komunikaci stojí komunikace asynchronní, kdy není přítomnost uživatele během komunikace vyžadována. Sem patří například e-maily, webové nástěnky nebo diskusní fóra.

Diskusní fóra mají na rozdíl od chatu velkou výhodu: jejich příspěvky jsou archivovány a uživatel se k nim může kdykoliv vrátit, zatímco u chatu příspěvky rychle mizí (zobrazují se příspěvky v délce jedné obrazovky). (Jandová, 2006, str. 17) Archiv chatu jde vytáhnout do té chvíle, kdy je uživatel v místnosti a chatuje. Jakmile z místnosti odejde, k historii se už nedostane. Administrátoři chatů jsou schopni historii „vytáhnout“, k těmto případům však nedochází často: většinou ze závažných důvodů (např. pedofilové v chatovacích místnostech pro děti).

Právě proto, že na fórech příspěvky „visí“ dlouho, se stávají stále oblíbenějšími: dokážou zastupovat funkci chatů a diskusních fór zároveň. Samotné chaty však nedokážou zastupovat funkci diskusních fór.

Existují fóra, která fungují spíše jako chaty. Jsou zahlcena stovkami příspěvků během pouhé hodiny. V těchto případech však málokdo pročítá diskuse do konce nebo reaguje na staré příspěvky.

Naopak existují fóra, která mají velice málo příspěvků, ale jsou vysoce navštěvovaná. Příspěvky totiž obsahují důležité informace: například různé rady (od rad pro počítačové uživatele po řešení složitých matematických případů) či užitečné odkazy (odkazy na filmy, seriály, e-knihy).

V této podkapitole se důkladně nezabýváme všemi rozdíly mezi chatem a diskusními fóry, proto shrneme základní rozdíly pomocí následující tabulky:

charakteristika chatové komunikace	charakteristika komunikace na diskusních fórech
synchronní komunikace (online) příspěvky po čase mizí dynamické a rychlé vkládání příspěvků jednodušší tematika větší soukromost nepřipravenost, spontánnost	asynchronní komunikace archivace příspěvků pomalejší vkládání příspěvků složitější tematika veřejnost větší připravenost projevu než u chatu ¹

Tabulka 2.1: Shrnutí základních rozdílů mezi chatem a diskusním fórem.

¹Projev na diskusních fórech je většinou připravenější než u chatových příspěvků. Je to zejména kvůli tomu, že uživatel má větší čas na zareagování, a uvědomuje si, že příspěvek bude veřejně zobrazován delší dobu (i několik let). Neplatí to však vždy (např. diskusní fóra, která fungují spíše jako chaty).

2.3 Netiketa

Internetové prostředí poskytuje uživateli tak velkou anonymitu, že ztrácí zábrany a chová se nevhodně. Záměrně ruší komunikaci ostatních, vybíjí si své negativní emoce, zveřejňuje závadné příspěvky nebo svými komentáři obtěžuje ostatní.

Právě kvůli takovým uživatelům postupem času vznikla pravidla, která většina uživatelů respektuje. Tato pravidla bývají označována jako *netiketa* (z angl. *net etiquette*). Většina velkých serverů, které umožňují internetovým uživatelům komunikovat, zveřejňuje pravidla na svých stránkách. Servery upozorňují uživatele na doporučené a dovolené chování, ale taky na chování, které je zakázáno, včetně sankcí. (Jan-dová, 2006, s. 24)

Každé stránky mají různá pravidla, všeobecně však platí, že příspěvky nesmí být v rozporu se zákony (rasismus, xenofobie, dětská pornografie, propagace drog). Dále pak nesmí obsahovat příspěvky, které by narušovaly komunikaci ostatních uživatelů: urážky, vulgarismy, osobní údaje uživatelů (telefonní čísla, ...) nebo reklamy na místech, které pro ně nejsou určeny.

2.4 Moderované diskuse

S narůstajícím počtem internetových uživatelů a oblibou diskusních fór však nastává problém, jak hlídat nezávadnost příspěvků. Zvláště pak zpravodajské servery, které zaznamenávají v diskusích pod články denně několik tisíc příspěvků, musí tyto problémy řešit kvůli inzerentům: těm se nelíbí, že se jejich reklama objevuje hned vedle rasistických, urážlivých a vulgárních příspěvků, které na ně vrhají negativní stín.

Pokud nejsou servery schopné uhlídat nevhodné příspěvky, jejich prestiž klesá. Tento problém často řeší tím, že se diskuse stávají moderované. Moderátor diskuse pročítá nové příspěvky, a pokud mezi nimi nalezne takové, jejichž obsah je nevhodný, okamžitě je odstraní. Pokud problémový uživatel soustavně pokračuje ve zveřejňování příspěvků se závadným obsahem a nereaguje na upozornění moderátorů diskuse, může být z databáze uživatelů smazán, nebo na nějakou dobu zablokován.

Moderátor diskusních fór bývá často mylně označován jako cenzor. Zatímco cenzori odstraňují jakékoli názory, které se neshodují s (politickou) ideologií, ať už země nebo zpravodajského serveru, moderátor odstraňuje pouze nevhodné příspěvky: nejčastěji urážky nebo rasistické projevy. Právě uživatelé, jejichž příspěvky byly smazány, si na fóru opětovně stěžují na chování moderátora a označují ho jako cenzora, což se dá považovat za určitý projev urážení. Výraz cenzor se nicméně vžil mnohem více než moderátor diskuse.

Role moderátora je podstatná, nicméně se dá lehce zneužít: najdou se i takoví, kteří neodstraňují pouze závadné a nevhodné příspěvky, ale i komentáře, které svým obsahem netiketě ničím neodporují – jsou však v rozporu s názorem moderátora, a proto jsou smazány. Právě takové chování je chybné – neshoduje se s náplní práce moderátora diskuse.

Moderátoři ovšem odstraňují i příspěvky, které se netýkají diskutovaného tématu: dbají na věcnost. S tímto jevem jsme se často setkávali během analýzy dat k této bakalářské práci. Řada příspěvků, které byly smazány, nebyly vůbec urážlivé, vulgární nebo rasistické či jinak závadné, ovšem byly svým způsobem zbytečné. Mezi nejčastěji mazanými komentáři se objevily příspěvky typu „*opět tu rádí cenzor*“ nebo „*venku prší*“. O mazání takových příspěvků rozhoduje osoba, která má diskuse na starosti, většinou je nadřazená samotnému moderátorovi. Mazány byly i takové příspěvky, které samy o sobě nevhodné nebyly, avšak po zavedení do kontextu diskuse se daly tímto způsobem chápat.

Další funkcí moderátorů diskuse je zvýraznit mezi tisíci příspěvky právě takové, které stojí za přečtení. Zvýraznění zajímavých příspěvků nezajišťují jen moderátoři, ale i sami uživatelé, kteří mohou příspěvky hodnotit podle jejich obsahu buď kladně, nebo záporně. Toto hodnocení se zobrazuje u každého příspěvku – ten s největším skóre (pozitivním i negativním) se zobrazuje v internetové diskusi na prvním místě.

2.5 Typy diskusí na zpravodajských serverech

Není naším cílem, abychom v této bakalářské práci dopodrobna zkoumali online internetové diskuse, ani abychom mapovali všechny druhy, které se vyskytují na celosvětové síti internetu.

K získání dat pro naši bakalářskou práci jsme využívali výhradně diskuse pod články českých zpravodajských serverů, protože právě tyto diskuse jsou neaktivnější. Uživatelé na tyto zpravodajské portály lákají hlavně informace o současném dění doma a ve světě. Dále pak možnost vyjádřit se k článkům a přečíst si názory ostatních uživatelů, v některých případech i možnost vést diskusi s lidmi opačného mínění. Svou roli hraje i široká tematika článků nebo velká interaktivnost diskusního prostředí, neboť zpětné reakce přicházejí téměř ihned. Právě z tohoto důvodu se zaměříme v této podkapitole na typy diskusí, které se vyskytují na českých zpravodajských portálech.

Hned na začátku je potřeba upozornit, že se nezabýváme grafickou podobou diskusí a podobnými tématy, ale spíše způsoby, které zpravodajské servery využívají k omezení počtu urážlivých, vulgárních a rasistických projevů, případně uživatelé, kteří se chovají nevhodně.

Už jsme se zmiňovali o tom, co je to moderovaná diskuse, proto se tomuto tématu dále nevěnujeme. Na tomto místě bychom rádi zdůraznili, že všechny typy diskusí, které uvádíme v této podkapitole, využívají moderátora. Využívají však i dalších způsobů, jak omezit nevhodné příspěvky na svých fórech. A právě o těchto způsobech hovoříme v následující podkapitole.

Řeč přijde zejména na anonymní diskuse a diskuse s registrací. Diskusi s registrací dělíme na diskuse, které jsou vedeny skrz *facebook*, diskuse vyžadující tzv. korespondenční registraci a diskuse pod „skutečnými jmény“. Třetí typ diskusí (vedle anonymních diskusí a diskusí s registrací) představují diskuse, které využívají automaty, a čtvrtý typ diskuse, v nichž na závadné příspěvky dohlíží sami uživatelé nebo programy.

2.5.1 Diskuse bez registrace

Anonymní diskuse

Pokud chce uživatel přidat příspěvek do anonymní diskuse, nepotřebuje k tomu žádnou přezdívku s heslem nebo potvrzení e-mailu, zkrátka žádnou registraci.

Uživatel si může vybrat, jakým způsobem se podepíše (uživatel zadává přezdívku, v některých případech chce server i e-mail, ten se ale nijak neověřuje a uživatel si jej může vymyslet). Tato skutečnost vede ovšem k tomu, že jeden uživatel může mít deset různých podpisů a vystupovat tak například pod deseti identitami.

Zatímco registrace uživatelů bývá zdlouhavá a často od přidání příspěvku odradí, u anonymních diskusí uživateli žádná registrace nepřekáží. Neexistuje nic, co by uživatele „zdrželo“ od přidání příspěvku, proto anonymní diskuse obsahují velké množství komentářů, ve většině případů nevhodných.

Anonymní diskuse většinou nemají žádnou hodnotu pro ostatní (slušné) uživatele. Používají je hlavně diskutéři, kteří si chtějí hrubě „postěžovat“. Velice často jsou nesrozumitelné nebo se příspěvky neustále opakují. V dalších případech obsahují velké množství urážek, vulgarismů, rasistických projevů, spamů nebo reklam.

Administrátoři musí kontrolovat velké množství příspěvků, mnohdy na ně ani nestačí. Z tohoto důvodu nebývají některé diskuse k problémovým tématům ani otevírány. Nejčastěji se tomu děje u zpravodajských serverů. Administrátoři mají předem vytipované problémové články. Takový případ nastal například v říjnu roku 2012, kdy zemřel král olšských Romů: diskusi k tématu uzavřely *iDnes.cz*² nebo *Blesk.cz*³.

V současné době využívá anonymních diskusí zpravodajský server *Deník.cz* a server *Hospodářských novin – iHned.cz*.

²Uzavřená diskuse k tématu smrti krále olšských Romů na *iDnes.cz*: http://zpravy.idnes.cz/pohreb-olasskeho-krale-v-ostrave-d4e-/domaci.aspx?c=A121018_1842588_ostrava-zpravy_jog [cit. 2013-02-25].

³Uzavřená diskuse k tématu smrti krále olšských Romů na *Blesk.cz*: <http://www.blesk.cz/clanek/zpravy-udalosti/183588/v-ostrave-zemrel-kral-olasskych-romu-selhalo-mu-srdce.html> [cit. 2013-02-25].

2.5.2 Diskuse s registrací

Diskuse vyžadující registraci

K tomu, aby mohli uživatelé přispívat svými názory do některých diskusí, potřebují registraci. Ta zabere několik málo minut, zároveň je ale natolik zdlouhavá, aby alespoň trochu odradila problémového uživatele.

Registrace může mít různou podobu. Někdy stačí zadat jen přezdívku a heslo. V jiných případech požaduje server poskytnutí e-mailu. V těchto případech si uživatel zvolí přezdívku a heslo, zadá svůj osobní e-mail, kam server pošle aktivační odkaz nebo jen potvrzení o registraci. Jindy přidělí přístupové heslo pomocí e-mailu přímo server. Způsobů je mnoho.

Na rozdíl od jiných diskusí si však uživatel může zvolit jakoukoliv přezdívku a nemusí vystupovat pod svým pravým jménem. Identitu uživatele nikdo neověřuje (jako je tomu u jiných serverů).

Tento typ je nejrozšířenější skupinou webových diskusí. Najdeme je například na serverech *Blesk.cz*, *Reflex.cz*, *Živě.cz*, *Lidovky.cz* nebo *Respekt.cz* a dalších.

Facebookové diskuse

Některé zpravodajské servery přešly na diskuse vedené skrz *facebook*. Tento systém má několik nesporných výhod: snadné nasazení do stránky, nezatažování zpravodajských serverů, odrazení anonyma od založení falešného účtu⁴ či možnost diskutovat nejen pod články, ale i pod fotogaleriemi a infografikami. Dále pak fakt, že je *facebook* mezi uživateli masově rozšířen.

Navíc jsou na *facebooku* všichni uživatelé zaregistrováni pod svým pravým jménem a příjmením (a pokud ne, *facebook* se časem postará o to, aby byl falešný profil odstraněn). Právě od tohoto faktu si zpravodajské servery slibují, že omezí urážky, vulgarismy a rasistické projevy v příspěvcích. Pokud jsou příspěvky anonymní, nikdo se nebojí použít ty nejhrušší nadávky, když jsou ale pod nimi autoři podepsáni svými skutečnými jmény, riskují to, že za hrubé porušování zákonů budou tvrdě potrestáni – nejen moderátory, ale i policií.

Na *facebooku* existuje postranní lišta, v níž si může uživatel přečíst cokoli, co jeho přítel okomentuje (musí to být na veřejných stránkách, což zpravodajské servery jsou). A jelikož problémoví uživatelé nestojí o to, aby jejich přátelé, sousedi nebo příbuzní věděli o tom, že se na diskusních fórech chovají nevhodně, jsou diskuse vedené přes *facebook* brzdou, která uživatele donutí přemýšlet nad tím, jestli problémový příspěvek zveřejní.

Podoba diskusního rozhraní na *facebooku* znemožňuje více než dvoustupňovou stromovou diskusi, což způsobuje nepřehlednost – není vidět, kdo na co reaguje. Toto omezení má ale i svou výhodu: toto diskusní rozhraní neumožňuje rozsáhlé souboje názorů. Při několika stech příspěvcích za hodinu ztrácí problematičtí uživatelé přehled o tom, kde je odpověď jejich „soka“, a než stačí zareagovat, jeho příspěvek se ztratí ve vlně dalších.

I přes maximálně dvoustupňové větvení diskuse se dají snadno najít příspěvky, které stojí za přečtení. Komentáře, které obsahují zajímavý názor, uživatele „*likují*“ (z ang. *like* – líbit se) pomocí proslulého facebookového tlačítka „*líbí se mi*“. Příspěvek může být ohodnocen i 250 „*liků*“, což uživatel při „rolování“ stránky lehce postřehne.

Nevýhodou facebookových diskusí je to, že je redakce nemají zcela pod svou kontrolou. Nemohou mazat příspěvky. Toto právo náleží pouze samotnému uživateli. Přesto však mohou moderátoři diskuse skrýt problémové příspěvky před ostatními čtenáři. V naprosté většině případů toto skrytí funguje jako samotné smazání – běžní uživatelé skrytý příspěvek nevidí, zobrazuje se pouze uživateli samotnému a jeho přátelům. V krajních případech nevhodného chování existuje možnost zablokování příspěvků z daného profilu.

Další nevýhodou facebookových diskusí je to, že i přes masové rozšíření *facebooku* existuje spousta těch, kteří tuto sociální síť nepoužívají. V takových případech uživatelům nezbyvá nic jiného než se registrovat (jako je tomu na ostatních serverech s registrací), pokud chtějí do diskuse přispívat.

Mezi české zpravodajské servery, které využívají facebookových diskusí, patří například *Aktuálně.cz*, *E15.cz* nebo *iSport.cz*.

⁴Sám *facebook* uvedl, že počet falešných účtů se pohybuje ve výši 8,7 %. Uživatelé si vytváří duplicitní účty kvůli hrám, šmírování, potřebě skrýt skutečnou identitu, z pracovních důvodů (správa stránek) nebo proto, že původní účet byl uživateli zablokován. (Dočekal, 2012)

Jan Lipold, pracovník zpravodajského serveru *Aktuálně.cz* a vedoucí tamějších diskusí, píše v e-mailu ze 17. 10. 2012: „V roce 2011 přešel server *Aktuálně.cz* na facebookové diskuse, do té doby používal diskuse anonymní. Příspěvků výrazně ubylo, ale diskuse zato mají větší hodnotu a smysl. Moderování diskuse se provádí ‚ručně‘. Moderátoři díky úbytku komentářů stíhají pročítat všechny příspěvky. Během těch tří-nácti měsíců, kdy používáme facebookové diskuse, jsme ani jednou nemuseli sáhnout k tomu, abychom diskusi zavřeli, což se nám v minulosti občas stávalo.“

Diskuse s tzv. korespondenční registrací

Nejnavštěvovanějším českým webem je *Seznam.cz*, v lednu roku 2012 jeho stránky měsíčně navštívilo 5,6 milionu uživatelů. (SPIR, 2012)

Seznam.cz provozuje od r. 2003 spolu se svým spoluvydavatelem Borgis (mimo jiné vlastník deníku *Právo*) online deník *Novinky.cz*.

Návštěvnost online deníku *Novinky.cz* se v lednu 2012 pohybovala na 4 milionech uživatelů za měsíc, což z něj dělá nejnavštěvovanější zpravodajský server České republiky. (SPIR, 2012)

Právě kvůli vysoké návštěvnosti měly *Novinky.cz* problém ve svých diskusích: připomínaly více chaty. Urážlivé, vulgární a rasistické projevy byly na denním pořádku.

V roce 2009 se *Novinky.cz* rozhodly zásadně změnit podobu svých diskusí. (Fiala, 2009) Zpravodajský server prohlásil, že umožní přístup k diskusi pouze těm uživatelům, kteří projdou tzv. korespondenční registrací.

Při korespondenční registraci žádají uživatelé o přidělení účtu a přístupového kódu pomocí předem daného formuláře. Na oplátku musí serveru poskytnout své jméno, příjmení a obec, s těmito údaji pak v diskusi vystupují. Server si pravost osobních údajů ověřuje tím, že přístupové údaje posílá uživateli na uvedenou adresu.

Celý proces korespondenční registrace je zdlouhavý a pro zpravodajský server administrativně velmi náročný. Nové registrace bývají často kvůli velkému náporu uživatelů pozastavovány. Proto se na serveru často setkáte s upozorněním: „Přijímání nových registrací do názorových fór bylo 8. 2. 2013 pro velký zájem pozastaveno. Jakmile vyřídíme došlé žádosti, začneme přijímat nové.“ (Hlavní strana *Novinky.cz*, 2013)

Další obměnou, která doprovázela změnu diskusí na *Novinkách.cz*, bylo zrušení stromové diskuse a přechod k lineární diskusi (podobně jako je tomu na facebookových diskusích). Tato změna byla přijata kvůli tomu, aby byly diskuse slušnější. Uživatelé nemají možnost reagovat na sebe přímo, čímž se alespoň trochu zamezilo osobnímu napadání, nepřehledným chatům nebo diskutování mimo téma.

Diskuse pod „skutečnými jmény“

Druhý nejnavštěvovanější český zpravodajský portál *iDnes.cz*, který je mimo jiné online deníkem novin *Mladá fronta DNES*, bojoval s velkým množstvím příspěvků v diskusích, kde uživatelé diskutovali pod přezdívkami. Jak uvádí Jan Dvořák, správce diskusí serveru *iDnes*, v e-mailové komunikaci (1. 3. 2013), zpravodajský portál má okolo 40 tisíc registrovaných uživatelů, z nichž velká část diskutuje.

Od listopadu 2012 přistoupil zpravodajský portál na přísnější politiku – dovolil čtenářům diskutovat pouze pod svými skutečnými jmény. Na rozdíl od *Seznamu.cz* však nevyžadoval žádnou korespondenční registraci, ale pouze změnu přezdívky na jméno a příjmení.

Po provedení změny se začala objevovat spousta pochybných identit: Josef Stalin, Klement Gottwald a jiní. V takových případech požadoval server po uživateli zaslání nascenovaného identifikačního průkazu. Pokud tak uživatelé neučinili v předem stanovené době, jejich účet byl odstraněn.

Server se snaží podchytit užívání pochybných identit. Pokud si dnes uživatel během registrace vybere jméno Adolf Hitler, okamžitě se zobrazí varování informující o tom, že uživatel zadal jméno, které se v České republice nepoužívá, a že server pozastaví registraci uživatele, dokud nedoloží svou totožnost pomocí průkazu. Ovšem jméno Vladimír Lenin zaregistrovat lze. Takováto problémová jména jsou ale automaticky posílána k ručnímu schválení, které provádějí administrátoři diskusí.

Diskutujícím, kteří si do stanovené doby nezměnili své přezdívky na skutečná jména, byl jejich účet smazán (byli upozorněni několika e-maily).

Přesto však nelze předpokládat, že diskutující uvedli svá skutečná jména: pokud si někdo založil falešnou identitu se jménem Petr Novák, *iDnes.cz* nezjistí, jestli je toto jméno skutečné nebo ne.

Nová podoba diskusí *iDnes.cz* se setkává s vlnou odporu. Jako hlavní argument uživatelé uvádějí, že se obávají o svou bezpečnost a o to, že si bude moc zaměstnavatel vyhledat názory zaměstnance nebo dokonce lustrvat zájemce o práci. I na takové případy je *iDnes.cz* připravený: „*Kvůli ochraně diskutujících před nežádoucím monitoringem nedovolíme vyhledávačům indexovat jména diskutujících – vyhledávače je nenajdou.*“ (MAFRA a. s., 2012)

2.5.3 Diskuse využívající automaty

Server *iDnes.cz* využívá mimo moderátorů, registrací a používání „skutečných jmen“ v diskusích i automat, který rozpoznává kořeny nebo určité části nevhodných slov. Pokud takové slovo odhalí, nedovolí uživateli příspěvek do diskuse odeslat (a zveřejnit ho).

Existují ovšem i nijak nezávadná slova, která automat do diskuse nepustí, protože obsahují části nevhodných slov. Takovým případem je například *opice* (obsahuje vulgarismus *pice*, *piče*) nebo *sextant*⁵ (obsahující slovo *sex*).

2.5.4 Diskuse, v nichž na závadné příspěvky dohlíží sami uživatelé a programy

Zajímavé řešení nevhodných příspěvků přináší americké zpravodajské médium *Huffington Post*, které bylo založeno v roce 2005 Ariannou Huffingtonovou (v roce 2003 kandidovala na guvernérku Kalifornie, zvítězil Arnold Schwarzenegger), Kennethem Lerem a Jonahem Perettim. (Encyklopædia Britannica, ©2013)

Návštěvnost tohoto webu je velmi vysoká. Jen pro srovnání: zatímco u nás navštíví náš nejnavštěvovanější server *Seznam.cz* okolo pěti milionů lidí za měsíc, *Huffington Post* navštíví stejné číslo čtenářů během jednoho dne, měsíčně to dělá něco okolo 60 milionů čtenářů. (Quancast, ©2013; SPIR 2012) Slogan propagačního videa zní: „*Huffington Post, conversation starts there*“ (*Huffington Post, právě tady začíná diskuse*). (AOL INC., © 2013) A právě počet příspěvků na diskusních fórech je to, co činí *Huffington Post* zajímavým.

Server umožňuje svým čtenářům připojovat komentáře pod články. Některé články mají okolo 180 tisíc komentářů. Propagační video serveru uvádí, že s každou hodinou přibývá 10 tisíc příspěvků, za týden to dělá milion. Od roku 2012 jejich počet dokonce překračuje 9 milionů komentářů za měsíc, ročně tedy více než 70 milionů příspěvků. Pro představu: tuzemský zpravodajský portál *iDnes.cz* při přechodu na nové diskuse uvedl, že se na internetových fórech objevovalo 150 tisíc příspěvků týdně, tedy asi 8 milionů komentářů za rok. (AOL INC., © 2013; Struhárik, 2012)

Proč mají uživatelé tak rádi diskuse na tomto serveru? Podle Justina Isafa, ředitele diskusí *Huffington Post*, za to může „*poutavost článků, poctivé moderování, které vytváří bezpečné místo pro všechny zúčastněné, a to, že má server obsah, z něž jsou lidé nadšeni.*“ (AOL INC., © 2013; Struhárik, 2012)

Huffington Post prohlašuje, že má své diskuse pod kontrolou. Pokud by diskuse serveru byly plně nevhodných příspěvků, těžko by mezi uživateli měly takovou oblibu: do diskusí, kde se objevuje spousta nevhodných komentářů, slušní uživatelé přispívat nechtějí.

Čím to je, že jiné servery nezvládají ani zlomek toho, co *Huffington Post*?

Mohou za to hned dvě věci. Ta první z nich je intenzivní spolupráce s uživateli serveru. V roce 2010 začal server udělovat svým uživatelům vyznamenání za aktivitu. Pokud uživatel nahlásil administrátorovi 25 nevhodných příspěvků, stal se moderátorem prvního stupně. Pokud nahlásil 100 nevhodných komentářů, stává se moderátorem druhého stupně a sám může mazat nevhodné příspěvky – do doby, než toto privilegium zneužije. (Struhárik, 2012)

Přesto se uživatelé nestávají součástí administrátorského týmu, který se o diskuse stará. Ten sestává pouze z 30 členů na plný úvazek, kteří se střídají v šestihodinových směnách. Administrátoři dokážou projít stovky příspěvků za hodinu. Počet administrátorů je nízký díky *Julii*. (Sonderman, 2012)

Julia je uměle inteligentní program. Vyvinula ji firma Adaptive Semantics a její název je zkratkou „*Just a Linguistic Algorithm*“. V roce 2010 tento program *Huffington Post* odkoupil. *Julia* se dokáže během svého chodu učit a umí několik jazyků. (Sonderman, 2012)

⁵Sextant je přístroj, který se používá k úhlové vzdálenosti dvou bodů či k určení výšky hvězd nad obzorem.

Tento uměle inteligentní program prochází všechny komentáře na *Huffington Post* a rozpoznává nevhodné příspěvky. Svým prohledáváním napomáhá moderátorům, aby se jejich práce zrychlila. (Sonderman, 2012; Struhárik, 2012)

Julia však nepomáhá moderátorům jen s rozpoznáváním nevhodných příspěvků. Umí moderátory upozornit na zajímavé příspěvky, které stojí za povšimnutí, nebo odkázat na části diskusí, které jsou inteligentní. (Sonderman, 2012; Struhárik, 2012)

Jak *Julia* funguje však ředitel diskusí na *Huffington Postu* Justin Istaf odmítá sdělit. Principy *Julie* zůstávají tajemstvím firmy i *Huffington Postu*. (Sonderman, 2012)

2.5.5 Další řešení nevhodných diskusních příspěvků

Jsou i další řešení, jak zacházet s nevhodnými příspěvky. Například *Sme.sk*, největší slovenský zpravodajský server, nesmaže celý příspěvek, který obsahuje urážlivé výroky – jednoduše je celé „vyhvězdičkuje“ a příspěvek ponechá.

Dalším častým problémem diskusí je to, že počáteční vlákno diskuse obsahuje nevhodné výroky, reakce na tato vlákna jsou však bez problémů. Smazání počátečního vlákna s sebou však nese problém: zmizí i reakce. Některé weby tento problém neřeší, jiné ano: například *Blesk.cz* smazaný příspěvek nahradí upozorněním: „*Autor příspěvku dlouhodobě nerespektoval pravidla diskuse...*“, reakce na smazané vlákno jsou však ponechány.

Jiné servery zase umožňují zobrazení vadného příspěvku jenom po přihlášení.

2.6 Shrnutí

V této podkapitole bychom rádi shrnuli přehled informací, o kterých jsme doposud mluvili. K tomu slouží následující tabulka, která ukazuje, jaké typy diskusí se vyskytují na českých zpravodajských portálech.

typ diskuse	servery
anonymní diskuse	<i>iHned.cz</i> <i>Deník.cz</i>
diskuse s (jednoduchou) registrací	<i>Blesk.cz</i> <i>Respekt.cz</i> <i>Lidovky.cz</i> <i>živě.cz</i>
facebookové diskuse	<i>Aktuálně.cz</i> <i>E15.cz</i> <i>iSport.cz</i>
diskuse s korespondenční registrací	<i>Novinky.cz</i>
diskuse pod "skutečnými jmény"	<i>iDnes.cz</i>

Tabulka 2.2: Typy diskusí na českých zpravodajských portálech.

Navíc přidáváme tabulku, která ukazuje návštěvnost českých zpravodajských serverů. Tabulka uvádí období, název serveru a počet reálných uživatelů⁶ za měsíc.

Na tomto místě bychom rádi zdůraznili, že ne všichni návštěvníci zpravodajských portálů do diskusí přispívají. Ne všichni se registrují a ne všichni registrovaní do diskusí přispívají.

I přesto to není zanedbatelná částka, vzhledem k tomu, že někteří návštěvníci denně napíší několik desítek komentářů, které musí moderátoři projít.

U počtu reálných uživatelů *iDnesu.cz* uvádíme pouze počty v sekci „zprávy“, neboť do celkového počtu téměř 4 milionů měsíčních návštěv *iDnes.cz* započítává návštěvnost jízdnicí řádů (cca 1,7 milionů návštěv), fotogalerie na *Rajčeti* (cca 900 tisíc návštěv) atd. (Dočekal, 2013)

	leden 2006	leden 2008	leden 2010	leden 2012	leden 2013
<i>Novinky.cz</i>	1 993 413	2 928 299	3 349 348	4 051 176	3 986 948
<i>iDnes.cz</i>	767 610	995 084	1 232 184	1 542 623	1 811 760
<i>Aktuálně.cz</i>	633 566	1 055 477	1 278 702	1 538 996	1 710 346
<i>iHned.cz</i>	383 110	548 045	592 182	759 119	842 225
<i>Blesk.cz</i>	521 212	711 462	836 743	1 152 924	1 862 627
<i>iSport.cz</i>	84 715	157 171	301 518	421 648	498 822
<i>Respekt.cz</i>	27 602	81 428	46 377	74 171	189 700
<i>Deník.cz</i>		487 459	657 422	1 006 383	1 261 097
<i>Lidovky.cz</i>		579 855	664 924	887 016	1 021 657
<i>Parlamentnílisty.cz</i>				323 928	509 927

Tabulka 2.3: Tabulka, která ukazuje počet reálných uživatelů, již navštívili hlavní stránky zpravodajských serverů v lednu roku 2006, 2008, 2010, 2012 a 2013. (SPIR)

⁶Reálný uživatel označuje skutečného (reálného) člověka, který daný server navštívil ve zkoumaném období. Jeden reálný uživatel může používat více počítačů (a prohlížečů), tudíž je pak identifikován jako vícero unikátních uživatelů. (SPIR, 2011).

2.7 Závěr

Zdaleka ne všechny zpravodajské servery mají prostor, který by vyhradily uživatelům k diskusi. Důkazem toho je například *BBC* – ta u článků možnost diskutovat nemá. V tuzemsku tuto skupinu reprezentuje například *ČT24.cz*.

Tyto zpravodajské portály si nejspíš až moc dobře uvědomují, kolik práce (a financí) dá udržování slušných diskusí.

Zatímco zahraniční diskuse často využívají uživatelů, kteří hlásí nevhodné příspěvky, nebo speciálních programů, jež vyhledávají závadné komentáře, u nás nic podobného není. Náš jazyk je poměrně bohatý – urazit člověka lze různými způsoby, i za použití neutrálních slov. Na rozvíjení učících se programů, které by se naučily rozpoznávat urážlivé, vulgární a rasistické výrazy, chybí finance.

Objem příspěvků na českých diskusních fórech není tak velký, jako je tomu například na amerických serverech. Proto naše zpravodajské portály stále využívají moderátorů. Co se ale stane, až bude objem příspěvků neúnosný?

Někteří se však programům striktně brání, i když by na ně peníze našli. Jejich důvodem je nedůvěra k takovýmto programům. Nevěří totiž, že by odhalily to, co je potřeba. Nebo aby nesmazaly něco, co nemají. Jisté urážlivé nebo vulgární komentáře jsou v určitém kontextu přijatelné. A proto se ne všechny urážlivé vulgární příspěvky odstraňují.

Co se ale stane ve chvíli, kdy začne příspěvků přibývat a moderátoři je nebudou stíhat procházet? Jak dlouho zabere vývoj programu, který by si dokázal poradit s nevhodnými příspěvky uživatelů, a dokázal si dokonce poradit s takovými, které sprosté výrazy nepoužívají, a přesto jsou urážlivé? Čeština je bohatá. A uživatelé kreativní.

Zvláště v době, kdy moderátoři začali tvrdě zasahovat vůči nevhodným příspěvkům, se objevil na českých diskusních fórech nový trend – urážení nevulgárním způsobem, pomocí běžně používaných slov. Ta jdou odhalit velice těžce, protože neutrálním českým slovům byl přidán nový význam.

Používat nejčastější česká urážlivá a vulgární slova umí každý, ale vymyslet si svá vlastní slova, která navíc pobaví i ostatní uživatele, to už je umění.

A právě komentáře, které obsahují novotvary, jenž vtipným způsobem uráží, si získávají veliké oblíbenosti a stávají se čím dál používanějšími.

3. Urážlivé, vulgární a rasistické projevy na internetových diskusních fórech

V této části bakalářské práce se zaměříme na urážlivé, vulgární a rasistické projevy na internetových diskusích a pokusíme se je zařadit do kategorií. Kompletní seznam všech výrazů, které jsme během našeho sběru získali, je k nalezení v abecedně řazené příloze. V ní lze najít i význam hesla, jeho použití v kontextu a lexikografické zmapování, k němuž jsme používali *Slovník spisovného jazyka českého (SSJČ)*, *Slovník spisovné češtiny pro školu a veřejnost (SSČ)* a *Slovník nespisovné češtiny (SNČ)*; v případě víceslovných spojení *Slovník české frazeologie a idiomatiky (SČFI)*.

Zdroje, ze kterých jsme vybrané výrazy vybírali, jsou různé, převážně ale pocházejí z diskusí pod českými zpravodajskými servery.

Od *Novinek.cz* a tištěného deníku *Práva*) jsme získali několika set stránkový seznam nevhodných příspěvků, které byly z diskusí odstraněny. Jelikož seznam obsahoval pouze nevhodné komentáře, posloužil nám nejvíce. Velká část hesel ve slovníku v příloze pochází z těchto dat.

Novinky.cz však nebyly jediným zdrojem, který jsme při práci používali. Během doby, kdy tato práce vznikala, jsme intenzivně procházeli internetové diskuse a mezi desetitisíci příspěvky jsme vyhledávali ty, které obsahovaly nevhodná slova. Mezi hlavní zdroj patřil hlavně bulvární online deník *Blesk.cz*, který má hodně diskutérů, jejichž slovník byl daleko hrubší než slovník jiných uživatelů serióznějších médií (*Aktuálně.cz*, *iDnes.cz*). Občas nám byl nápomocný i online bulvární magazín *Aha.cz*.

Ke sběru dat jsme používali i jiné zpravodajské portály, mezi něž bychom mohli zařadit *Novinky.cz*, *Aktuálně.cz*, *iDnes.cz*, *iHned.cz*, *Deník.cz*, *ParlamentníListy.cz*, *EuroZprávy.cz* a *Lidovky.cz*.

Mezi nejčastější témata článků, v nichž se nevhodné příspěvky vyskytovaly, patřila hlavně témata týkající se celebrit – počínaje Ivetou Bartošovou, Lucií Bílou, Agátou Hanychovou a konče kauzou Ornelly Štikové. Dalším častým tématem byla politika – diskutéři se rozčilovali nad zákony, růstem DPH nebo nad politickými stranami a politickými osobnostmi. Velkou roli hrál hlavně odchod českého prezidenta Václava Klause a dále pak přímá volba prezidenta nového, jimž byla na zpravodajských serverech věnována velká pozornost. Největšími terči útoku se stávali kandidáti druhého kola – Miloš Zeman a Karel Schwarzenberg. Mnozí z nich (nejen poslední dva kandidáti) si vysloužili od uživatelů diskusí spoustu přívěsků.

Ani současná vláda nebyla ušetřena kritiky. Petr Nečas, Miroslav Kalousek, Jan Kubice nebo bývalí ministři Vít Bárta a Radek John; o těchto politicích se po prezidentských kandidátech mluvilo nejvíce. Z opozičních stran se častým námětem stával Bohuslav Sobotka. V zapomenutí neupadli ani dřívější rivalové Jiří Paroubek a Mirek Topolánek.

3.1 Kategorizace urážlivých, vulgárních a rasistických výrazů

Slovník urážlivých výrazů, příloha této bakalářské práce, zahrnuje více než pět set výrazů. Aby byla slova se stejnou tematikou u sebe, rozhodli jsme se je alespoň z části zařadit do dvaceti kategorií a dalších podkategorií.

První z nich jsou kontextově závislé výroky. Nevhodnost těchto příspěvků se projeví až po zařazení do kontextu. Další kategorií jsou výrazy se sexuální tematikou, které zahrnují sexuální orientaci nebo transsexualitu; v diskusích se často objevovalo označení prostitutek, mužských a ženských pohlavních orgánů, lidí se sklonem k promiskuitě nebo výrazy týkající se ženských prsou nebo samotného pohlavního styku. Proto jsme se rozhodli zařadit tyto výrazy jako samostatné podkategorie.

Třetí kategorii představují lexémy urážející inteligenci člověka, čtvrtá náboženství. Do páté kategorie jsme zařadili výrazy urážející vzhled – s podkategoriemi malý vzrůst, plnoštíhlá postava, ženy s velkými nadry, lidé trpící akné a výrazy označující ošklivou osobu obecně.

Šestá kategorie shromažďuje slova, jejichž cílem je urážet člověka. Kategorie sedmá se vztahuje k povaze člověka – zařadili jsme zde výrazy pro člověka lakomého, bázlivého, nemravného nebo nevychovaného, podlézavého a pro osobu, která mluví zbytečně nebo neobsačně, v některých případech i pomlouvá.

V osmé kategorii se vyskytují výrazy pro osoby závislé na návykových látkách, tedy na drogách a alkoholu. Devátá kategorie představuje výrazy, jež označují hloupost, kategorie desátá výrazy pojmenovávající obličej.

Častým předmětem diskusí byly i lexémy označující podřadné skupiny a nemanželské, zlobivé děti nebo děti s negativním podtónem. Pro tyto lexémy jsme vytvořili kategorii jedenáctou a dvanáctou.

Velkou kategorií představuje skupina s číslem třináct, kterou jsme nazvali politické výrazy. Spadají do ní pojmenování příslušníků politických stran a výrazy pro politiky samotné.

Čtrnáctá kategorie uvádí příklady slov označujících národnost člověka.

Patnáctou kategorií ztělesňují xenofobní výrazy, mířící proti cizincům na našem území. Jako podkategorie jsme zvolili Romy, Vietnamce a Japonce, neboť jejich jiná barva pleti upozorňuje na odlišnost od ostatních obyvatel nejvíce.

Kategorií šestnáctou reprezentují nejčastější nadávky, které se u nás používají.

Kategorií sedmnáct tvoří přízviska známých osob – nejčastěji pro prezidenty a prezidentské kandidáty, své místo zde ale našel i Karel Gott.

Do osmnácté kategorie jsme zařadili výrazy, které odráží jazykovou hru se jmény známých osob. Mezi oblíbené patřil Karel Schwarzenberg, Bohuslav Sobotka, Tomio Okamura, ale i Mirek Topolánek nebo bulvárně známá Ornella Štiková.

Předposlední, devatenáctou kategorií, představují grafické značky. Poslední kategorií tvoří speciálně upravené výrazy kvůli blokovacím automatům.

3.1.1 Kontextově závislé

Do této kategorie se řadí příspěvky, které se na první pohled nezdají ničím závadné, po zařazení do kontextu je okamžitě vidět jejich neetičnost.

Jako příklad uvádíme článek o nehodě motorkáře, kterému náraz do svodidel utrhl hlavu. V diskusích se objevovaly příspěvky typu: *bezhlavý jezdec; ta dnešní mládež hned ztrácí hlavu; nemá jezdit bezhlavě; jde mi z toho hlava kolem.*

Takovéto příspěvky neobsahují žádné nevhodné lexémy, program by je tudíž sám o sobě těžko odhalil. V takovýchto případech je potřeba ručního zásahu moderátora.

3.1.2 Sexuální tematika

V této kategorii se objevují výrazy různými způsoby spjaté se sexuální tematikou. Časté jsou narážky na lidi s homosexuální orientací, prostitutky, ale taky různé druhy výrazů, které slouží k označení mužských či ženských pohlavních orgánů, ženských nader nebo samotného pohlavního styku. Právě do takových podkategorií jsme výrazy rozčlenili. U některých podkategorií uvádíme i víceslovná spojení.

- ▷ **výrazy narážející na sexuální orientaci:** *bomerant, buk, bukál, bukvice, buzerant, buzik/buzík, buzna, homokláda, homouš, hulpéro, kuřbuřt, řit'opich, spermohlt, teplouš*
víceslovné lexémy: *zateplená menšina*
- ▷ **výrazy spojené s transsexualitou:** *trand'ák, transka, transvestil*
- ▷ **výrazy pro označení prostitutek:** *běhna, děvka, kurva, šlapka, štětka*
víceslovné lexémy: *vyjetá macocha*
- ▷ **výrazy označující promiskuitu:** *běhna, coura, čubka, čúza, dávačka, děvka, flundra, fuchtle, kuřevník, rajda, rozhod'nožka*
- ▷ **výrazy používané pro označení mužských pohlavních orgánů:** *brko, čulin, čurák, frantík, hadice, chobot, klacek, kláda, kladélko, klobása, kokot, ocas, péro, pindík, pind'our, pták, šulin*
- ▷ **výrazy používané pro označení ženských pohlavních orgánů:** *bobr, frnda, kunda, piča/piča, pi-pina, prcka, prcina*

- ▷ **výrazy používané pro označení ženských prsou:** *cecky, dudy, kozičky, kozy, melouny, plast'áky, vemená*
- ▷ **výrazy používané pro označení pohlavního styku:** *drbat, jebat, klátit, mrdat, ojet, píchat, prcat, šoustat, šukat, trkat*
- ▷ **další výrazy pojící se se sexuální tematikou:** *honibrk, jebka, kazišuk, kazimrd, klátič, mamrd, masturbant, mrdka, mrdoun, onanista, ošoust/vošoust, vymrdanec, zjeb, zmrđ*

3.1.3 Výrazy urážející inteligenci

Existuje celá řada výrazů, které urážejí inteligenci. Do této početné skupiny jsme zařadili následující slova, která jsou určena ženskému nebo mužskému pohlaví či oběma pohlavím.

blb, blbec, debil, degen, dement, dilina/dylina, husa, idiot, ichtyl, imbecil, jelito, koza, kráva, kripl, krůta, magor, nána, nanynka, negramota, pabl, pitomec, retard, slepice, trdlo, trol, trol, trubka, tupec, tupoun, tydýt, vymaštětec, vypatlanec

víceslovné lexémy: *mentálně pokulhávající*

3.1.4 Výrazy týkající se náboženství

Ačkoliv existují země, které jsou mnohem více netolerantní vůči jiným náboženstvím, neznamená to, že i u nás se jiná náboženství zcela respektují. Patříme mezi nejateističtější národy vůbec, proto nemáme výrazy zaměřené jenom proti vyznavačům jiného náboženství, než je křesťanství, zejména muslimské (*musulman*) a židovské (*židák, židle*), ale i slova mířená proti věřícím obecně (*pámbíčkář*) nebo proti kněžím (*černoprdelník*).

černoprdelník, musulman, pámbíčkář, židák, židle

3.1.5 Výrazy týkající se vzhledu

Častým cílem posměšků se stává i samotný vzhled člověka – ať už je to malá výška jedince, člověk s nadváhou, žena s bujným poprsím, osoba trpící kožní nemocí akné nebo ošklivý člověk obecně. Tyto příspěvky se objevovaly zejména na bulvárním serveru *Blesk.cz*, který se zaměřuje na známé osoby.

- ▷ **malý vzrůst:** *připrclík, skřet, trpaslík*
- ▷ **plnoštíhlá postava:** *špekoun, tlustoprd*
- ▷ **velká prsa:** *kozatka, prsatice*
- ▷ **akné:** *bed'arák, pod'obanec, uherák, uhrin*
- ▷ **výrazy označující ošklivou osobu:** *držka, kreatura, mužatka, příšera, ropucha, smažka, zrůda*

3.1.6 Výrazy urážející věk

Ani postarší lidé nejsou ušetřeni pošklebků, jež se nelichotivě vyjadřují o jejich věku. Do této kategorie řadíme některé z nich.

dinosaurka, fosílie, kraksna, mastodontka, rašple, senil, škatule, vykopávka, zkamenělina

3.1.7 Výrazy týkající se povahy

Tato kategorie se zaměřuje na výrazy vztahující se k negativní povaze člověka. Častým tématem je lakomost, bojácnost, nemravnost a nevychovanost, podlézavost nebo pomlouvačnost.

- ▷ **lakomec:** *držgrešle, krkoun, skrblík, škrt*
- ▷ **bojácný člověk:** *počůránek, posera, posránek, srab, srágora, srágotka, uplakánek, zbabělec*
- ▷ **nemravný nebo nevychovaný člověk:** *buran, čuně, čuňák, dobytek, hňup, hulvát, chlívák, křupan, prasák, svině*

- ▷ **podlézavý člověk:** *alpinista, eurohujer, ližprdelka, patolízal, řit'olezec, šplhoun, vlezdoprdelka, vlezdobruselista*
- ▷ **osoba, která zbytečně nebo bezobsažně mluví nebo pomlouvá:** *drbna, mluvka, kecka, kecal, kecálek, klepna, žvanil, žvatlal*

3.1.8 Výrazy označující osoby závislé na návykových látkách

Čeština je poměrně bohatá na výrazy označující osoby, které jsou závislé na návykových látkách, tedy alkoholu a drogách. Přitom slov, která jsou posměšně mířena proti alkoholikům, je více.

- ▷ **alkohol:** *becher, becherovka, chlastoun, kořala, nácučka, násoska, norek, notor, notorik, ochlasta, ožrala, pijan, výpitek, závislák, zmetylovaný*
- ▷ **drogy:** *fet'ačka, fetka, smažka, vymaštěnec*

3.1.9 Výrazy označující hloupost

Tato kategorie zahrnuje slova, která vyjadřují hloupost nebo hloupý názor, hloupou věc.

blbina, blbost, hovadina, kravina, pičovina/píčovina, sračka

3.1.10 Výrazy označující obličej

V češtině existuje značné množství lexémů, které negativně pojmenovávají lidská ústa nebo obličej. Tato kategorie reflektuje takové lexémy.

čuňa, držka, huba, klapačka, ksicht, pysk, tlama, zobák

3.1.11 Výrazy označující podřadnou skupinu

Častým jevem na internetových diskusích je označování skupin s dávkou podřadnosti nebo pohrdání. Takovýmto způsobem bývají nazýváni například politici nebo jiná společensky neoblíbená skupina. V této kategorii uvádíme alespoň některé výrazy.

holota, chamrad', chátra, lůza, pakáží, plebs, podlidi, póvl, sebranka, verbež

3.1.12 Výrazy označující nemanželské nebo zlobivé dítě

Tuto kategorii tvoří označení nemanželského nebo zlobivého dítěte. V některých kontextech byly tyto výrazy použity i pro neoblíbené dítě nebo dítě, jehož rodiče nejsou v oblibě.

bastard, fakan, harant, parchant, spratek, zkurvysyn, zmetek

3.1.13 Politické výrazy

Jedněmi z nejčastějších terčů posměšků se na diskusních fórech stávali politici a politické strany. V posledních letech byly vytvořeny výrazy, které se staly mezi uživateli velice oblíbené, a my jsme se s nimi setkávali téměř v každé diskusi týkající se politiky. K takovým výrazům patřila označení člena ODS *modrobolševik/modrý bolševik*, dále *oranžáda* a *sockomouš* jako pojmenování člena ČSSD.

- ▷ **příslušníci politické strany:** *bolševik, fašista, fašoun, komanč, komančeros, komár/komáři, komouš, modrobolševik, nácek, oranžáda, socan, sockomouš*
- víceslovné lexémy:** *modré straky, modrý bolševik, přebarvený komunista*
- ▷ **politici:** *darmožrout, eurohujer, hrabáč, nenažranec, papaláš, patolízal, podržtaška, rusofob, tlučhuba, tunelář, tuneldědek, vlezdobruselista*

3.1.14 Výrazy označující národnost

I osoby národnosti jiné než naší si získaly svá pojmenování. Cílovými skupinami se nejčastěji stávali naši sousedé – Němci (*Dojčland'ák, Helmut, Němčour, Skopčák*), Poláci (*Pšonek*) a Slováci (*Čobol, Čobolák*), ušetřeni nebyli ani ne málo vzdáleni Francouzi (*Frantík*) nebo u nás žijící etnikum Vietnamců (*Vietnamčik*). Angličané (*Anglán*), Američani (*Amík*), Arabové (*Arabáč*), Italové (*Talián*) a Japonci (*Japončik*) mají vedle neutrálních forem i svá další pojmenování.

Amík, Anglán, Arabáč, Čobol, Čobolák, Dojčland'ák, Frantík, Helmut, Japončik, Němčour, Pšonek, Skopčák, Talián, Vietnamčik

3.1.15 Xenofobní výrazy

Na našem území žije několik menšin, nejpočetnější z nich jsou Slováci, Poláci, Němci a romská menšina. Četné zastoupení u nás má i vietnamské etnikum.

Nejčastějším terčem kritiky se stávají Vietnamci a Romové, zvláště kvůli odlišné barvě pleti, ale i kvůli odlišnosti kultury nebo životního stylu.

Nejneoblíbenější jsou Romové, zvláště kvůli tomu, že se mezi nimi vyskytuje velké procento nezaměstnaných jedinců a lidí závislých na sociálních dávkách a rovněž vysoká kriminalita.

U výrazů označujících Romy si můžeme všimnout spojitosti s tmavou pletí (*briketa, hnědoobčan, kofola, morgoš, negr, přičmoudlík, ušmudlánek*), s vysokou nezaměstnaností (*nefakčenko, nemakčenko, nemaklo, netahlo, nicnedělač*) nebo podle názorů některých i zneužívání sociálních dávek (*parazit, vyžírka*).

- ▷ **Romové:** *bakelit, bezbarák, briketa, cigán, cikán, cikorka, cinkle, hnědočech, hnědoobčan, hnědoslovák, kofola, morgoš, negr, nefakčenko, nemakčenko, nemaklo, netahlo, nicnedělač, nepřizpůsobivý, parazit, přičmoudlík, robič, temnočech, ušmudlánek, utiskovaný, vyvolený, vyžírka, znevýhodněný*

víceslovné lexémy: *cikán černý, černá huba, černá opice, černá tlama, opálená menšina*

- ▷ **Vietnamci, Japonci:** *Japončik, rákosník, rejžožrout, šikmoočko, šikmooký, t'aman, tit'an, Vietkong, Vietnamčik*

víceslovné lexémy: *kurva šikmooká*

3.1.16 Nejčastější a nejpoužívanější výrazy

Vedle základních rozdělení do skupin, které jsme uvedly výše, jsme vytvořili kategorii, do níž řadíme nejčastěji se vyskytující lexémy v diskusích.

blbec, buzna, buzerant, čurák, debil, dement, hajzl, hovno, idiot, kokot, kráva, kurva, piča/píča, prdel, řít', sračka, srát, svině, šukat, vůl, zmrda

3.1.17 Přezdívky známých osob

I slavné české osobnosti a politici si vysloužili svá přízviska. Například Karel Gott, sedmdesátitřetiletý český zpěvák a držitel třiceti sedmi zlatých slavíků, si získal označení *skřehota zombie*, která uráží Gottovo zpívání (*skřehotání*) a věk (*zombie*).

Václavu Klausovi, bývalému prezidentovi České republiky, se v diskusích přezdívalo *kleptomán* a *čorlpero*, čemuž vděčí za návštěvu v Chile v roce 2011, kdy na konferenci s chilským prezidentem „ukradl“ protokolární pero přímo před televizními kamerami.

Další Klausovy přezdívky – *klimatolog, meteorolog* – se vztahují k jeho výrokům, v nichž kritizoval to, že by člověk způsoboval globální oteplování.

I další bývalý český prezident Václav Havel, který měl problémy s respiračním ústrojím, v minulosti kouřil a kterému lékaři odstranili půlku plíce kvůli rakovině, nebyl ušetřen posměšků. Kvůli častému kašli si získal přezdívku *chrchloun Venca*.

Třetí český prezident Miloš Zeman si vysloužil hlavně přezdívky spojené s jeho zálibou v alkoholu – *becher, kořala*.

Karla Schwarzenberga, protikandidáta Miloše Zemana v první přímé volbě prezidenta, diskutéri označovali jako *mickey mouse*, což je kreslená postavička Walta Disneyho, která je občas vyobrazována s motýlkem okolo krku, stejně jako Karel Schwarzenberg.

I další kandidáti na prezidenta měli mezi lidmi svá přízviska. Nejznámější se stala přezdívka Vladimíra Franze – *avatar*. V jeho modře potetovaném obličejí viděli lidé spojitost s modrou barvou pleti avatarů ze stejnojmenného filmu Jamese Camerona Avatar. Janě Bobošíkové, další prezidentské kandidátce, přezdívali lidé *korytnačka*, protože kandidovala na mnoho postů – na poslankyni Evropského parlamentu, do vedení KSČM, Poslanecké sněmovny, Senátu, na primátorku Prahy a naposledy na prezidentku České republiky. Uživatelé internetových diskusí si z jejího „tlačení se ke korytům“ dělali legraci vtipnými obrázky, na nichž oznamovala kandidaturu na papeže.

Karel Gott: *skřehota zombie*

Václav Klaus: *čorlpero, kleptoman, klimatolog, meteorolog*

Václav Havel: *chrchloun*

Miloš Zeman: *becher, kořala*

Karel Schwarzenberg: *mickey mouse*

Vladimír Franz: *avatar*

Jana Bobošíková: *korytnačka*

3.1.18 Jazyková hra se jmény známých osob

Častěji než vymyšlením přezdívek se lidé na internetových diskusích bavili tím, že známým osobám přerucovali jména. Adresát se pak za novotvary snažil najít skryté významy nebo nějaké spojitosti.

Nejširší pole působnosti nabízel uživatelům jméno Karla Schwarzenbergera, prvního místopředsedy současné vlády Petra Nečase, zároveň ministra zahraničních věcí a jednoho z kandidátů na prezidenta České republiky. Za zmínku stojí například *Schlafenberg* (z něm. *schlafen* – spát), *Šláfík* a *Šlofik* posmívají- cí se Schwarzenbergerovu usínání v Poslanecké sněmovně. Další označení *Trautenberk* si Karel Schwarzenberg vysloužil podobným zněním jmen a svou plnoštíhlou postavou, kterou měl i Trautenberk, jedna z postav Krkonošských pohádek.

Jméno Tomia Okamury, dalšího kandidáta na prezidenta⁷, poskytlo diskutérům široké množství urážlivých verzí – mezi nejzajímavější patří například *Pitomio*, *Odkama můra* a *Kikomura*.

Uživatelé přerucovali jména i dalších politiků – například Petra Nečase, předsedu nejméně oblíbené vlády v historii Česka, nazývali *Petr Přesčas*, což má vyjadřovat touhu lidí, aby vláda, která je tu „přes čas“, skončila.

Obměny se dočkalo i jméno neoblíbeného ministra financí Miroslava Kalouska (*Miroslav Kradousek*, *Lihoslav Krad'ousek*) nebo jméno bývalé ministryně obrany Karolíny Peake (*Pikča*, podobnost se slovem *piča*), která byla ve funkci pouhých osm dní.

Velkou řadu možností nabízí i jméno předsedy ČSSD Bohuslava Sobotky (*Dluhoslav Sockobotka*, *Socbotka*, *Srabotka*). Nová podoba *Dluhoslav* je nejspíš motivována tím, že tato strana je spjata s narůstáním státního dluhu. Výrazy *Sockobotka* a *Socbotka* jsou zase spojeny s tím, že ČSSD usiluje o zajištění lidí s nižšími příjmy a lidí závislých na sociálních dávkách, které jiní označují jako *socky*.

Bývalý předseda ČSSD Jiří Paroubek si vysloužil označení *Parobejk* – nejspíš kvůli tomu, že se po téměř třiceti letech manželství rozvedl a vzal si novou ženu Petru, která je o jedenáct let mladší než Paroubkova bývalá žena Zuzana. Podle *Slovníku nespisovné češtiny* označení *bejk* znamená „sexuálně hyperaktivní muž“, což měli zřejmě uživatelé na mysli.

David Rath, další z členů ČSSD, byl v roce 2012 zatčen při přebírání sedmimilionového úplatku. Tento prostořeký poslanec byl mezi lidmi neoblíbený, zřejmě proto si vysloužil označení *David Srath*.

⁷Tomio Okamura byl z oficiálních kandidátů na prezidenta vyškrtnut ministerstvem vnitra, protože mezi podpisy na jeho petiční listině se vyskytovalo velké množství falešných jmen, ve výsledku tudíž neměl požadovaných 50.000 podpisů na zařazení do oficiální kandidatury.

Poslední politik, o kterém se zmíníme, je Mirek Topolánek, bývalý předseda ODS a předseda vlády (2006–2009). Mimo pojmy *Topol* a *Tupohlav* si vysloužil ještě označení *Topořánek*, nejspíše motivované obdobným případem jako u Jiřího Paroubka. Mirek Topolánek si vzal o dvanáct let mladší ženu, než byla jeho bývalá manželka.

Z osobností „šoubyznysu“ zmíníme kauzu Ornely Štikové, která v současnosti hýbe bulvárem. Ornella otěhotněla s Josefem Koktou (podnikatel), který byl v té době ženatý s moderátorkou a zpěvačkou Gábinou Partyšovou. Ornella na celé situaci začala vydělávat tím, že do *Blesku* poskytovala rozhovory a fotky (své, později i svého dítěte), čímž se celá kauza začala nafukovat a lidé na tuto „celebritu“ začali být velice alergičtí. Články v *Blesku*, které se týkaly Ornely, obsahovaly největší množství diskusních příspěvků, diskutéri v nich výhradně nadávali. Ornella (pro některé synonymum prostitutky) si vysloužila označení *Pornella*, *Ohnella*.

Další problémovou osobu představuje česká zpěvačka Iveta Bartošová, která má za sebou celou řadu skandálů. Diskuse pod články, které se týkají této ženy, zaznamenávaly velkou aktivitu. Iveta se stala terčem vtipů a nadávek, diskutéri ji nejčastěji nazývali *Bliveta*.

Za zmínku stojí i slovenský porotce Jaroslav Slávik ze soutěže Československo má talent. V nedávné době v médiích vyplulo na povrch, že v mládí točil gay-porno, proto ho začali diskutéri označovat *Jaro Gáyik*.

Zakladatel islámu, proroku Muhamedovi, se dočkal obměny svého jména – diskutéri mu říkali prorok *Muhammrd*. První dáma byla přejmenována na *první dá mi*. Nového názvu se dočkal i server *Novinky.cz* – *blbinky.cz* a *hovinky.cz* – který je podle mnohých považován za nedůvěryhodný a bulvární. Bulvárnímu *Blesku* lidé v diskusích říkali *Plesk*.

Různých variant se dočkaly i daňové reformy (*daňová deforma*) a penzijní reformy (*penzijní zdechforma*) a název politické strany TOP 09 (*TOI⁸ 0-9*).

Karel Schwarzenberg: *Schlafenberg, Schweinzenberg, Šláfík, Šlofik, Trautenberg*

Tomio Okamura: *Kikomura, Odkama múra, Pitomio*

Petr Nečas: *Petr Přesčas*

Miroslav Kalousek: *Lihoslav Kradousek, Krad'ousek*

Kateřina Peak: *Pikča*

Bohuslav Sobotka: *Dluhoslav Sockobotka, Socbotka, Srabotka*

Jiří Paroubek: *Parobejk*

David Rath: *David Drath, David Srath*

Mirek Topolánek: *Topol, Topořánek, Tupohlav*

Ornella Štiková: *Ohnella, Pornella*

Iveta Bartošová: *Bliveta*

Jaroslav Slávik: *Jaro Gáyik*

Justin Bieber: *Justin Bobr*

Tomáš Klus: *Tomáš Cval, Tomáš Flus, Tomáš Klas*

Agáta Hanychová: *Špagáta*

prorok Muhamed: *prorok Muhammrd*

Blesk.cz: *Plesk.cz*

daňová reforma: *daňová deforma*

Novinky.cz: *blbinky.cz, hovinky.cz*

penzijní reforma: *penzijní zdechforma*

první dáma: *první dá mi*

TOP 09: *TOI-09*

⁸TOI TOI je název mobilních záchodků používaných hlavně na stavbách a různých festivalech.

3.1.19 Grafické výrazy

Do této kategorie jsme zařadili speciální výrazy označující zadek nebo prsa. Tyto grafické značky bývají občas uživateli v diskusích využívány k urážení – například:

Zadek: (._.), (._!)

Prsa: (.) (.)

3.1.20 Speciálně upravené výrazy kvůli blokovacím automatům

Protože mnohé diskuse využívají automatů, které zabráňují odeslat příspěvek obsahující nevhodné slovo, obcházejí uživatelé toto nastavení různými způsoby: střídáním velkých a malých písmen (*bLBeC*), přidáním nebo odebráním znaku ve slově (*kkt, kua, perdel*), přidáním nebo odebráním diakritiky (*débil, pice*), nahrazením určitého znaku jiným – ale takovým způsobem, aby bylo stále pochopitelné, o jaké slovo jde (*curva, echtly, hofno, howno, kujva, kulfa*).

Někdy není potřeba měnit formu slova, stačí do slova vložit interpunkční znaménka (mezery, tečky, hvězdičky apod.) – buď zčásti (*bl.b, de bil, ko*kot*), nebo je úplně provázat interpunkčními znaménky⁹ (*b l b e c, m.r.d.k.a.*) či nahrazení jednoho písmene pomocí interpunkčního znaku (*ko*ot*¹⁰, *krá.a*).

*bl.b, bLBeC, b l b e c, cocot, curva, de bil, débil, debíl, echtly, hofno, howno, kkt, ko*kot, ko*ot, krá.a, kua, kujva, kulva, kurfa, kurv@, kurwa, perdel*

⁹Extrémními případy jsou slova, která mají víc interpunkčních znamének než samotných písmen, např.: *bl.....bka*.

¹⁰„Vyhvězdičkování“ představuje oficiální způsob cenzury nevhodných slov v médiích (hlavně tištěných a online médií), obdobou je „vypívávání“ v televizi a rádiu.

4. Vznik urážlivých, vulgárních a rasistických výrazů

Problémová slova, která obsahuje náš slovník v příloze, vznikala určitým způsobem: přidáním nového významu ke starému slovu, derivací, kompozicí nebo vytvořením ustáleného spojení. V této kapitole se zaměříme na způsoby vzniku těchto slov trochu podrobněji. Pomocí této analýzy chceme získat alespoň základní charakteristiky vzniku, které by nám byly nápomocné při vyhledávání nevhodných slov v příspěvcích za pomoci programů. Hledáme tedy pravidla pro program, jenž by zachytil co nejvíce nevhodných slov, ale zároveň aby zachycoval co nejméně slov, která jsou v pořádku. K tomuto nám nejvíce poslouží slovtvorné postupy, zejména derivace a expresivní sufixy, které hrají ve vytváření nevhodných slov velkou roli.

Rádi bychom zdůraznili, že tato práce si nečiní nárok na hloubkovou analýzu toho, jak tato urážlivá slova vznikají – chceme pouze zjistit, jestli existují určité typické způsoby tvoření těchto výrazů, které bychom mohli zahrnout do pravidel programu.

Nové lexikální jednotky se tvoří čtyřmi způsoby: přejímáním z cizích jazyků, slovtvornými procesy, mezi něž se řadí derivace (odvozování), kompozice (skládání) a abreviace (zkracování), sémaziologicko-onomaziologickou derivací (sémantickými procesy), a spojováním slov (hlavně ustálená spojení).

4.1 Přejímání slov z cizích jazyků

Jak jsme již řekli v předchozích řádcích, přejímání slov z cizích jazyků je jedním ze čtyř způsobů vzniku nové lexikální jednotky.

Náš slovník v příloze zahrnuje jen velmi málo slov, jež by byla přejata z cizího jazyka. Z této kategorie se na diskusních fórech ujalo slovo *shit* [šit] pocházející z angličtiny. Tomuto lexému bychom mohli přiřadit český ekvivalent *sračka*, tedy dílo nízké hodnoty (*Takže verze 7 je ještě větší shit než Visty?*).

Dále můžeme vidět i vliv němčiny – ve slovech *Schwein* [švajn], v češtině *prase*, a *Sheisse* [šajse], v překladu *hovno*, *sračka*. Tato slova se většinou nepoužívala samostatně jako nadávka, ale při jazykové hře se jménem Karla Schwarzenberga [švarcnberka, švarcnberga], jednoho z prezidentských kandidátů. Diskutéri mu přezdívali *Schweinzenberg* [švajcnberk] a *Scheizenberg* [šajznberk].

V případě nalezení takovýchto slov je stačí pouze přidat do databáze nevhodných lexémů, které program sám nalezne a odstraní je.

4.2 Sémantické procesy

Nové lexikální jednotky často vznikají v důsledku sémantických procesů. Důležitou úlohu zde sehrávají významové přenosy (metafora a metonymie) a sémantické obměny – významové posuny (generalizace a konkretizace). Přenášení významu může být i expresivního charakteru, které může být doprovázeno změnou (dodáním) konotačních složek.

4.2.1 Významové přenosy (sémaziologicko-onomaziologická derivace)

Důležitou část mezi neosémantizačními procesy představují významové přenosy, kam spadá metafora a metonymie. Mluvíme o tzv. sémaziologicko-onomaziologické derivaci – ze sémaziologického pohledu

přenášíme význam, z onomaziologického přenášíme pojmenování. (Rangelová, 2005, str. 162) Přenášení významu pomocí metonymie není tak časté jako přenášení metaforické.

Právě metafora hrála u mnoha slov, která jsme zachytili v našem slovníku, velkou roli. Přidávala totiž starým slovům nový význam, a to na základě vnější podobnosti. Nejčastěji jsme mohli přenášení vidět u pojmenování Romů, kde byla důležitá barva pleti (*briketa, kofola, hašlerka* a *ušmudlánek*). Metaforické přenášení se dále vyskytovalo u slov označující staršího člověka (*fosílie, zkamenělina*) nebo u nehezského člověka (*ropucha*), ale i označení členů ČSSD (*oranžáda*), u nichž vidíme podobnost barvy strany ČSSD a oranžové limonády.

Jako metonymické přenášení bychom mohli uvést spojení *modrobolševik*, tedy pojmenování pro stranu ODS, jejíž barva je modrá (svou roli zde hraje i metafora). Zde se jedná o metonymické přenášení pojmenování jednotlivce (*bolševik*) pro označení celku (strany, vlády). U označení Romů jsme byli svědky přenášení významu pomocí ironie (*utiskovaný, vyvolený, znevýhodněný*).

4.2.2 Významové posuny

Méně jsme se setkávali s významovými posuny, do nichž spadají generalizace (rozšiřování) a konkretizace (zuzování).

Jako příklad generalizace (rozšiřování významu) bychom mohli uvést označení *negramota*, které v původním významu znamená *ten, kdo neumí číst nebo psát*. S postupem času se však vyvinul generalizací význam *nevzdělanost, nedovednost v určitém oboru (informační negramota)*.

Jako konkretizaci (zuzování významu) uvádíme lexém *bezbarák*, které by mohlo označovat bezdomovce, avšak uživatelé diskusních fór zúžili význam na Romy (*černí bezbaráci*).

4.2.3 Změna konotačních složek

Jistá emotivně hodnotící a expresivní dávka byla přidána například neutrálním slovům *bolševik* a *komunista*. Zatímco SSČ a SSJČ slovo *bolševik* pojímá jako neutrální, SNČ mu přidává charakteristiku hanlivé. Podobný případ vidíme i u slova *komunista*, v SSČ a SSJČ má význam neutrální, SNČ jej ale vůbec nezachycuje.

Tato slova sama o sobě pejorativní význam nemají, ačkoliv mohou být pragmaticky užita jako nadávky.

Program nedokáže sám rozhodnout, jestli byl slovu přidán nový význam (*hašlerka, kofola*). V těchto případech nezbyvá nic jiného než sledovat, jestli se některá slova, která se v příspěvcích objevovala jen velmi málo, nezačala objevovat ve větším množství (nejen pod jednou diskusí). V takovém případě by mohl program odeslat příspěvky moderátorovi, který by o nevhodnosti rozhodl „ručně“. Pokud by moderátor usoudil, že se opravdu jedná o nevhodné slovo, mohl by jej později přidat mezi slova, u nichž je potřeba tzv. desambiguace (zjednoznačnění).

4.3 Slovtvorné procesy

Této podkapitole se věnujeme podrobněji, protože se domníváme, že nám může při vyhledávání nevhodných slov, zejména při slovech nově vzniklých, značně pomoci. Proto se zabýváme z velké části derivací, velkou pozornost věnujeme hlavně expresivním sufixům.

Pokusíme se zjistit, nakolik je efektivní, pokud dáme programu příkaz, aby vyhledával všechna slova tvořená pomocí expresivních sufixů a předával je moderátorovi k zjednoznačnění.

Rádi bychom předeslali, že místo pojmu „sufix“ raději používáme výraz „zakončení“, neboť po formální stránce přesněji vyjadřuje to, co ve slovech hledáme, a usnadní nám práci.

4.3.1 Derivace

Koncovkové tvoření

▷ **Nulová koncovka:**

Pomocí nulové koncovky byla utvořena slova *hňup, neřád, pošuk, škrť a tlustoprd*. Pro náš program je nemožné, aby slova utvořená tímto způsobem vyhledával, proto s tímto způsobem tvoření slov npracujeme.

▷ **Zakončení -a**

Mnoho slov zachycených ve slovníku v příloze bylo vytvořeno pomocí koncového formantu *-a*, tedy konverze vzoru *žena, předseda*. Patří k nim například *coura, cuchta, hejhula, kořala, mizera, negramota, ochlasta, ožrala, posera, trajda*. Vyhledávání tohoto zakončení by nám práci nijak neusnadnilo, neboť takto končí mnoho slov – od 1-ového příděstí sloves ženského rodu (*koupila*), přes předložky (*na, za*), tvary podstatných jmen (*doktora*), číslovek (*několika*), a tak bychom mohli pokračovat dlouhým výčtem dál. Program, který by vyhledával takovéto zakončení slov, by nám posílal ke schválení mnohonásobně víc nezávadných slov.

Sufixace

Cílem této práce není dopodrobna popsat slovtvorbu všech námi zachycených slov, ale spíše využít slovtvorných procesů k tomu, abychom zjistili, zda je vůbec možné zohlednit derivaci v programu, který bude vyhledávat urážlivá slova. Z tohoto důvodu se nevěnujeme sufixům (resp. zakončením), které jsou součástí tvoření spisovných slov, neboť takováto analýza by byla příliš obsáhlá a ve výsledku by ničemu nepomohla. Potřebujeme totiž najít určité specifické zakončení slov, které se objevuje z velké části pouze u slov urážlivých (nebo expresivních).

Ve strukturování této podkapitoly jsme se inspirovali u podobné diplomové práce Jany Mališové (*Slovtvorná expresivita*, 2006), která se zabývala slovtvornou expresivitou, se zaměřením na derivaci.

Dále jsme si dopomáhali synchronním korpusem SYN2010, který obsahuje 100 milionů synchronních slov. V korpusu jsme vyhledávali informace o tom, kolik dalších slov má stejná zakončení, a podle toho jsme se rozhodovali, zda je efektivní zahrnout je do našeho vyhledávacího programu.

Maskulina

▷ **Zakončení -och:**

Zakončení *-och* jsme našli pouze u slova *rudoch* (ve významu komunista¹¹). Mnoho dalších slov má podobné zakončení – například expresivně zabarvená slova *staroch, tlust'och*, ale i neutrální slova *běloch, černochoch*, která jsou frekventovanější. Kvůli tomu nepovažujeme vyhledávání *-och* za efektivní. V tomto případě by účinnějším řešením bylo přidat jej do seznamu blokovaných slov, i když by se ovšem dalo polemizovat o tom, nakolik je slovo *rudoch* závadné, abychom jej z příspěvku museli odstranit.

¹¹Slovo *rudoch* se používá i ve významu indián. (Nekula, 1996, str. 89) Podle SSIČ je toto slovo označeno jako poněkud zastaralé. V diskusích se lexém ve významu indián nevyskytoval.

▷ **Zakončení -oš:**

Zakončení *-oš* se vyskytlo u slova *cigoš* (vzniklého resufixací). Ani vyhledávání tohoto sufixu nepovažujeme za vhodné, neboť jím končí spousta substantiv (*kámoš, listonoš, matroš, rozkoš*). Opět je vhodnějším řešením přidat lexém do seznamu blokováných slov.

▷ **Zakončení -ouš:**

Toto zakončení nalezneme například u slov *homouš, komouš, sockomouš* a *teplouš*. Takovýmto způsobem ale končí i neexpresivní slova jako *bělouš* a *chocholouš*, jejichž výskyt je však na diskusních fórech mizivý. Na druhou stranu jsou tímto sufixem tvořena jinak nezávadná expresivní slova (*cukrouš, drahouš, starouš*). Proto se nedomníváme, že by vyhledávání tohoto zakončení bylo něčemu nápomocné.

▷ **Zakončení -áš:**

Zakončení *-áš* jsme našli pouze u slova *papaláš*. Vyhledávat toto zakončení by ovšem nemělo žádný význam, protože takto končí i jiné slovní druhy – například druhá osoba singuláru indikativu (*děláš, máš*) nebo zájmena (*náš, váš*).

▷ **Zakončení -as:**

Ačkoliv se u slov *mamlas, ocas* vyskytuje takovéto zakončení, není vhodné jej do programu zavádět – končí tak například substantiva (*čas, hlas, zápas*) nebo imperativ sloves (*nahlas* to šéfovi).

▷ **Zakončení -out:**

Zakončení *-out* se vyskytlo u slov *darmožrout, rejžožrout*. Na druhou stranu takto ale končí velká řada sloves (*rozhodnout, zapomenout*) a některá podstatná jména (*chomout*; skloňování *pouta – pout*). Protože zakončení *-out* nepředstavovalo nijak frekventovanou skupinu, zvláště uvažujeme-li, že obě slova skládající se ze slova *žrout* jsou utvořena kompozicí, není nutné, aby program vyhledával takovéto zakončení.

▷ **Zakončení -oun:**

Jedním z velmi frekventovaných zakončení v našem slovníku se stalo zakončení *-oun*: *fašoun, chlastoun, chrchloun, kapoun, krkoun, mrdoun, plešoun, rudoun, slizoun, špekoun, šplhoun, tupoun*. Takové názvy, které označují osoby, jsou výhradně expresivní. Vedle toho existují i spisovná slova (*meloun, pozoun*), jejich výskyt v diskusích je však malý.

Jelikož je toto zakončení velmi frekventované, považujeme za vhodné, aby jej program do svého vyhledávání zahrnul. Problém je ovšem v tom, že se toto zakončení vyskytuje i u expresivních slov, která sama o sobě závadná nejsou (*mlsoun, hezoun*).

V tomto případě bychom mohli přidat slova se zakončením *-oun*, která se vyskytují v našem slovníku, na seznam blokováných slov. Program by měl dále ignorovat nezávadná expresivní slova (*hezoun, msloun*), ty bychom mu předem nadefinovali. Další slova, která byla utvořena pomocí *-oun* a která se zároveň nevyskytují na seznamu slov (ať už blokováných nebo ignorovaných) by program posílal ke schválení moderátorovi diskuse.

▷ **Zakončení -an:**

Buran, fakan, hnědoobčan, křupan, musliman, pijan, socan, t'aman, tit'an; to jsou slova se zakončením *-an*, která náš slovník zachytil. Na druhou stranu existuje spousta spisovných substantiv (*Američan, pan*) a expresivních substantiv (*hafan, nosan*), která končí podobně, nemluvě o n-ovém přičestí (ukol byl *udelan*) v případě, kdy uživatelé nepoužívají diakritiku.

▷ **Zakončení -án:**

Výskyt tohoto zakončení jsme zaznamenali u slov *cigán, cikán* a *čuramedán*. Takto ovšem končí četná řada substantiv (*džbán, román*), ale i n-ová přičestí (*zaměstnán*). Z tohoto důvodu není efektivní taková slova v textu vůbec vyhledávat.

▷ **Zakončení -ián:**

Náš slovník zachytil zakončení *-ián* pouze u slova *Talián*. Vedle toho však existuje velká řada spisovných slov se stejným zakončením (*pavián*, skloňování *liána – lián*), tudíž ani toto vyhledávání by ničemu nepomohlo.

▷ **Zakončení -oň:**

Mezi slova zakončená *-oň* patřil *mimoň, šikmoň*. I práce Mališové zachycuje další slova, která jsme na diskusních fórech nezachytili – *blboň, hluchon, hlupon, plešon, slepon*, přesto by se dala považovat za nevhodná (některá z nich jenom v kontextu). Vedle nich existují i další slova s podobným zakončením (*alespon, aspon*).

Protože v tomto případě převažují nevhodná slova, rozhodli jsme se zařadit vyhledávání tohoto zakončení do programu – s omezením, že slova *alespon* a *aspon* bude ignorovat.

▷ **Zakončení -l:**

Zakončení *-l* se vyskytla u slova *žvanil*. Mališová navíc zachytila i další urážlivá slova: *čumil, fňukal, kokotal, rejpal*. Všechna uvedená slova jsou deverbativní substantiva a jejich vyhledávání v textu by představovalo problém, neboť jimi končí *l*-ové přičestí sloves (*otočil, poklepal*).

▷ **Zakončení -áč:**

Mezi substantiva se zakončením *-áč* patřila hlavně slova *Arabáč, hrabáč* a *sráč*. Mimo ně existují i častá substantiva (*boháč, hráč, pláč*), tudíž ani tady by automatické vyhledávání zakončení nemělo význam.

▷ **Zakončení -ák:**

Jednu z nejfrekventovanějších skupin tvořilo právě zakončení *-ák*, proto jsme se je rozhodli rozčlenit na desubstantivní substantiva (*čuňák, Dojčland'ák, estébák, gumák, chlívák, opičák, plast'ák, prasák, Rusák, Skopčák, sviňák, židák*), deverbativní substantiva (*bonzák, čurák, piják, rypák, somrák, vrták*) a deadjektivní substantiva (*trapák, úchylák, závislák*). Problém je, že zakončení *-ák* je velmi produktivní u podstatných jmen (*darebák, chudák, levák, pták, tabák, voják, žebrák*), tudíž se nedomníváme, že by bylo vhodné, aby program tato slova vyhledával.

▷ **Zakončení -ec:**

Zakončení *-ec* se vyskytovala u slov *blbec, nenažranec, pitomec, posranec, pošahanec, rudokožec, řit'olezec, tupec, vymaštenec, vymrdanec, vypatlanec* a *zbabělec*. Na druhou stranu takovéto zakončení obsahuje řada frekventovaných substantiv (*chlapec, lovec, Němec, otec*), ale i příslovce (*nakonec, vůbec*). Z toho plyne, že není účelné toto zakončení vyhledávat.

▷ **Zakončení -ka:**

Další vysoce frekventovanou skupinu tvořilo zakončení *-ka*, který se uplatňuje u tvoření obourodých substantiv (vzor *žena, předseda*). Do této skupiny bychom mohli zařadit slova *držka, fetka, jebka, kecka, ližprdelka, mluvk, mrdka, nácucka, náfuka, násoska, padavka, podržtaška, sebranka, socka, srágotka, transka, vlezdoprdelka, vtěrka* a *vyžírka*.

Ovšem toto zakončení má řada frekventovaných substantiv (*láska, otázka*), příslovčí (*dneska*), dále pak řada různých zdvojnásobení (*cestička*), ale i tvary substantiv skloňovaných podle vzoru *pán* (*člověka, pomocníka*), ale i *hrad, předseda, město* a *kuře*.

Ačkoliv je zakončení *-ka* velmi frekventované, vyhledávání pomocí programu by přidalo mnohem víc práce.

Feminima

▷ **Zakončení -na:**

I skupina se zakončením *-na* je poměrně obsáhlá, zahrnuje slova: *běhna, buzna, drbna, fíflena, fňukna, fošna, klepna, kraksna, machna, mrdna, šukna*. Přesto existuje řada slovních druhů s tako-

výmto zakončením: například řada frekventovaných substantiv (*žena*), nemluvě o tvarech při skloňování vzoru pán (*Američana, blázna*), tvarech množného čísla nebo genitivu singuláru vzoru město (*jména*); dále toto zakončení pozorujeme u zájmen (*všechna*), číslovek (*jedna*), n-ových přídělostí sloves (*nalezena, ukradena*), příslovcí (*zrovna*) nebo předložek (*na*).

Protože mnoho slovních druhů formálně končí tímto způsobem, nepovažujeme vyhledávání zakončení *-na* za vhodné.

▷ **Zakončení -nda:**

Ačkoliv frekvence se zakončením *-nda* není moc velká (převážně jména *Amanda, Linda*), je toto zakončení vidět u řady expresivních slov (*sranda, septanda*). V našem slovníku jsme jej zachytili u třech slov: *frnda, kunda, pinda*.

Ani u této kategorie nepovažujeme za dobré řešení, aby program takováto slova vyhledával, v tomto případě stačí, když budou součástí blokovaného seznamu slov.

▷ **Zakončení -čka:**

Toto zakončení se objevilo například u slov *dávačka, feťáčka, klapáčka* a *sračka*. Mnohem častěji se však vyskytuje u běžných substantiv (*svíčka, rvačka, značka*), tímto způsobem jsou tvořeny i zdobněliny (*ručka, tvářička*). Z tohoto důvodu náš program takové zakončení vyhledávat nebude.

▷ **Zakončení -ina:**

K tomuto zakončení bychom mohli přiřadit slova jako *blbina, hovadina, kravina, micina, pipina* a *prcina*, formálně i *pičovina* (sufix *-ovina*).

Ani toto vyhledávání by ničemu nepomohlo, tímto způsobem končí řada substantiv (*hodina, rodina, rovina*), proto ani toto zakončení program vyhledávat nebude.

Prefixace

▷ **Prefix kvazi-:**

Pomocí prefixu *kvazi-* se tvoří slova s hodnotící funkcí ve smyslu předstírání nebo nepravosti. V našem slovníku se vyskytuje slovo *kvazipolitik*. To, že je tento prefix používaný, dokládá i korpus: *kvazimozek, kvazináboženské sekty, kvazinezávislá policie*. Proto jsme se rozhodli, že program bude takové prefixy vyhledávat a posílat je moderátorovi ke schválení.

▷ **Prefix pa-:**

Ačkoliv náš slovník zachycuje slovo tvořené pomocí prefixu *pa-* (*pabl*), není vhodné, aby jej program vyhledával, neboť mnoho slov začíná stejným způsobem (*pamatuješ, pan, paže*). Program navíc slovo *blb* nalezne i tak.

▷ **Prefix pod-:**

Prefix *pod-* vyjadřuje nižší stupeň v hierarchii. S jeho pomocí bylo utvořeno slovo *podlidi*, jak ukazuje náš slovník v příloze. Ani v tomto případě není efektivní, aby program vyhledával každé slovo začínající *pod-*, protože takový začátek mají mnohá slova (*podat, podezření*).

▷ **Prefix pseudo-:**

S pomocí prefixu *pseudo-* (ve významu falešný) byl utvořen lexém *pseudopolitik*, jak dokládá náš slovník. I korpus ukazuje, že tvoření slov tímto prefixem je poměrně častý: *pseudogestapák, pseudověda*. Z tohoto důvodu jsme se rozhodli, že slova obsahující tento prefix bude program vyhledávat a posílat moderátorovi ke schválení.

▷ **Prefix ultra-:**

Náš slovník zachytil i slovo utvořené pomocí prefixu *ultra-* (s významem nejvyšší míry): *ultrapiča*. Jak ale dokládá korpus, prefix *ultra-* je součástí mnoha dalších slov (*ultrafialový, ultramoderní, ul-*

trazvuk), proto jej do vyhledávacího programu nezohledníme.

▷ **Slova začínající truc-:**

I když kompozitum *truc-* mezi prefixy nepatří, řadíme jej do této části alespoň z hlediska formálního, neboť se jedná o začátek slova (zjednodušeně řečeno).

Slovo utvořené pomocí začátku *truc-* bylo zaznamenáno v našem slovníku: *Zemanova trucstrana*, čímž uživatelé mířili na to, že poté, co ČSSD před lety přestala mít zájem o Miloše Zemana, založil si „na truc“ svou stranu – Stranu práv občanů Zemanovci (SPOZ).

Korpus dokládá používání slov *trucpark*, *trucsólovka* atd. Ačkoliv existují slova, která začínají pomocí *truc-* (*trucovat*, *truchlít*), myslíme si, že by program měl vyhledávat takovýto začátek slova – s tím omezením, že slova *trucovat*, *truchlít* a všechny jejich tvary bude ignorovat.

Překrucování

K derivaci se řadí i překrucování slov. I takovéto výskyty náš slovník zaznamenal. V takových slovech můžeme vidět hru s jazykem, jejímž cílem bylo pobavit adresáta. Do této kategorie bychom mohli zařadit slova *partadement* (*parlament*), *posranec* (*poslanec*), *poštěvačka* (*pošt'ačka*), *senilát* (*senát*), *serebrita* (*celebrita*) a *soudrůzka* (*soudružka*).

V tomto případě nemusíme programu přidávat další instrukce, neboť blokovaná slova jsou součástí překrucovaných slov.

4.3.2 Kompozice

Některá slova, která zachytil náš slovník, byla vytvořena skládáním (kompozicí): *babochlap*, *budižkničemu*, *černoprdelník*, *darmožrout*, *ekoterista*, *eurohujer*, *fantasmagor*, *hnědočech*, *hnědoobčan*, *hnědoslovák*, *homokláda*, *kurvadrát*, *ližprdelka*, *modroboševik*, *nicnedělač*, *otočkabát*, *pravdoláskář*, *rejžožrout*, *rozhod'nožka*, *šprcguma*, *titulblábolista*, *tuneldědek*, *vlezdoprdelka*, *z kurvysyn* a *zlatokopka*.

Aby program mohl takováto slova odhalit, musel by sledovat délku slov ve znacích. Centrum zpracování přirozeného jazyka Masarykovy univerzity uvádí, že průměrná délka slova je 5.54 písmen. (CZPJ, 2012) Z výše uvedených slov je nekratší výraz *šprcguma* se svými osmi písmeny. Je tedy otázkou, zda je vhodné sledovat slova delší než osm písmen, abychom odhalili kompozita. Kupříkladu, i často používaná slova *prezident*, *univerzita* nebo *komunismus* mají více písmen.

Mimoto se mezi uživateli najdou takoví, které netrápí, když v příspěvku zapomenou na mezeru, tudíž by pak mohl nastat problém u příspěvku typu „*Včera jsem bylveškole a nudil jsem se.*“; další skupiny uživatelů mezeru nepoužívají vůbec (*MoSseMiTolíbí*).

Z těchto důvodů nepovažujeme za vhodné vyhledávat slova, která jsou delší než stanovený počet znaků. Když se pozorněji podíváme na některá slova (*černoprdelník*, *fantasmagor*, *homokláda*, *kurvadrát*, *ližprdelka*, *vlezdoprdelka*, *z kurvysyn*), zjistíme, že jsou složena ze slov, které blokovací program bez potíží zachytí. A protože celá řada nevhodných kompozit už v sobě obsahuje závadná slova, domníváme se, že otázkou kompozit se nemusíme nijak důkladněji zabývat.

4.4 Víceslovná spojení

V diskusích se mimo jednoslovné lexémy objevovala i ustálená – nebo alespoň častá – spojení. Není v silách moderátorů sledovat, jestli se v záplavě příspěvků tato spojení nevyskytují. V podstatě to ani není nutné, protože součástí těchto spojení jsou slova, která program vyhledá sám o sobě: *kráva* (*kráva nebeská*), *hajzl* (*jít do hajzlu*), *hovno* (*je to na hovno*), *prdel* (*jít do prdele*, *políbit prdel*, *slepičí prdelka*), *srát* a slova odvozená (*mlít sračky*, *nasrat si*, *to mě poser*, *v hlavě nasráno*, *vysrat si oko*).

Přesto bychom měli programu přidat na seznam závadných slov alespoň některá spojení, u nichž jde na první pohled poznat, že jejich cílem je urážet druhého uživatele (*držet hubu*, *držet klapačku*) nebo spojení s rasistickým podtextem (*černá huba*, *černá opice*, *černá tlama*).

4.5 Závěr

Zkoumáním toho, jakými způsoby byla utvořena hesla, která zachycuje náš slovník, jsme došli k závěru, že z hlediska sémantického není možné, aby program bez ruční posteditace odhalil nové významy již vzniklých slov. Jediné, co může udělat, je to, že bude sledovat, jestli se méně používaná slova nezačala v poslední době objevovat častěji – pokud ano, znamenalo by to, že jim nejspíš byl přidán nový význam (např. *kofola*, *hašlerka* jako označení Romů). V takovýchto případech by program příspěvky odesílal moderátorovi, který by celou situaci posoudil.

Z hlediska slovtvorného jsme došli k závěru, že by bylo efektivní, aby program vyhledával zakončení *-oun* (ignorace *hezoun*, *mlsoun*; *meloun*, *pozoun* atd. + jejich tvary) a *-oň* (ignorace *alespoň*, *aspoň*) a z prefixů *kvazi-*, *pseudo-* a slova začínající *truc-* (ignorace *trucovat*, *truchlit* + jejich tvary). Výskyty těchto zakončení a začátků slov pak bude program posílat moderátorovi diskuse k posouzení. V oblasti kompozice není nutné program nijak upravovat, protože součástí kompozit jsou nevhodná slova, jež jsou ve slovníku definována.

Jsmo si vědomi, že vyhledávání zakončení *-oun* apod. nemusí být efektivní, neboť takové prohledávání vyhledá pouze základní tvar slova (lemma), v textu se ale slovo bude vyskytovat v různých slovních tvarech (*tupouni*, *tupouna*). Takovýto problém dokážou programy vyřešit způsobem, že k nalezenému slovu určí základní tvar slova (lemma), tzn. pokud se v textu vyskytuje slovo *tupouni*, program vrátí slovo *tupoun*.

Z této podkapitoly vyplývá, že při zpracovávání materiálu z diskusních fór není možné opírat se o formální vlastnosti jednotek. Příznaková slova jsou vesměs tvořena neutrálními slovtvornými prostředky, tudíž není možné je vyhledávat na základě formálních kritérií (zakončení), protože by to vedlo k masivnímu přegenerování. Hlavní váha tudíž spočívá na ruční posteditaci, tedy na posouzení moderátora.

5. Zpracování slovníku (komentář k příloze)

Tato kapitola obsahuje komentáře k příloze, tedy slovníku nalezených urážlivých, vulgárních a rasistických slov, který jsme během bakalářské práce zpracovávali.

Cílem slovníku je představit shromážděný materiál, vysvětlit význam hesel, uvést je v kontextu a zmapovat, zda jsou námi získané výrazy zachyceny i v některých jiných lexikografických příručkách, tedy ve *Slovníku spisovného jazyka českého (SSJČ)*, *Slovníku spisovné češtiny (SSČ)*, *Slovníku nespisovné češtiny (SNČ)* a v případě víceslovných spojení i ve *Slovníku české frazeologie a idiomatiky (SČFI)*.

Slovník v příloze se nesnaží o výklad původu slov (etymologii), protože je to nad rámec bakalářské práce. Nesnaží se ani o dokonalou mikrostrukturu hesla. Usiluje pouze o elementární přehled zachycených výrazů a o přehled charakteristik, které uvádí výše zmíněná lexikografická díla.

Tato kapitola se zabývá jednak slovníkem, kde se zaměříme na to, jak jsme získávali materiál, na kolik částí jsme slovník rozčlenili a jak jsme jej zpracovávali. V další části se pak zaměříme na bližší charakteristiku použitých slovníků a zhodnocení jejich užitečnosti. Ke konci kapitoly se pak podrobněji zaměříme na příznakovost lexikálních prostředků a kritéria, podle nichž slovníky přistupují k dělení charakteristik.

5.1 O slovníku

5.1.1 Získávání materiálu

Jak jsme již zmiňovali v úvodu práce, materiál, který jsme zpracovávali v našem slovníku, jsme si opatřili z internetových diskusí. Velké množství dat jsme získali od *Novinek.cz* (resp. *Seznamu.cz*), které nám poskytly několika set stránkový seznam smazaných příspěvků z diskusí.

Mimo to jsme získávali data prohledáváním diskusí s velkou uživatelskou aktivitou, zejména diskuse pod zpravodajskými servery, kde denně přibývají desítky článků, což je pro uživatele ideální – mají mnoho různorodých témat, o nichž se dá diskutovat.

Kromě *Novinek.cz* jsme čerpali hlavně z bulvárního online deníku *Blesk.cz* a občas i *Aha.cz*. Ze serióznějších médií jsme využívali hlavně *Aktuálně.cz*, *Deník.cz*, *iDnes.cz*, *iHned.cz* a *Lidovky.cz*. Dále pak *ParlamentníListy.cz* a *EuroZprávy.cz*. Tyto servery nám žádná data neposkytly, nevhodné příspěvky jsme vyhledávali sami. Z těchto webů jsme tedy získali hesla slovníků, ale i kontext, ve kterém byla použita.

5.1.2 Rozčlenění slovníku

Námi získaný slovník jsme rozčlenili do pěti částí.

První část, zachycující jednoslovné lexémy, je nejrozsáhlejším úsekem, zobrazuje pouze čtyři slovní druhy: podstatná a přídavná jména, slovesa a citoslovce. V největší míře se vyskytovala podstatná jména. Z cca 530 hesel představují asi 490 slov, tedy asi 94 % slovníku. Zbytek připadá na přídavná jména (ve většině případů ironicky označují Romy), slovesa (převážně synonyma pohlavního styku) a ve velmi malé míře citoslovce. Do první části jsme zařadili i jména obyvatelská, která mají vedle neutrální podoby ještě další podobu, většinou se záporným citovým zabarvením. Hesla obsahují význam, použití v kontextu a zmapování výskytu celkem ve třech lexikografických dílech: SSJČ, SSČ a SNČ.

Do části druhé jsme zařadili speciálně upravené jednotky (např. *bl.b*, *howno*). S jejich pomocí uživatelé obcházejí blokovací automaty nepouštějící do diskuse nevhodné LJ. Tato část není nijak obsáhlá, čítá 24 hesel s kontextem a pravou podobou slova.

Třetí část zachycuje víceslovná spojení¹², jež jsme rozdělili na část neslovesnou (24 hesel) a slovesnou (22 hesel). Není tolik obsáhlá jako první část, neboť tato spojení nebyla primárním zájmem naší práce, pojali jsme ji spíše jako část okrajovou. Výskyt víceslovných spojení jsme hledali ve čtyřech lexikografických dílech: *Slovníku české frazeologie a idiomatiky* (SČFI), dále pak v SSSJČ, SSČ a SNČ.

Čtvrtá část obsahuje přízviska osob, kam jsme zařadili 8 hesel. Do této části jsme zahrnuli i další tři hesla používaná pro označení motorkáře a prezidenta.

Poslední, pátá část, obsahuje jednotky, u nichž vidíme jistou jazykovou hru se jmény osob nebo věcí. Uvádíme vždy heslo (žertovně upravené), pravou podobu jména (věci) a kontext, v němž bylo heslo použito. Pátá část obsahuje celkem 32 hesel reflektující žertovně upravená jména osob a 9 hesel hrajících si s ostatními slovy (názvy stran, webů apod.).

5.1.3 Heslové slovo

Heslové slovo uvádíme v reprezentativním tvaru, tedy v nominativu singuláru, u sloves pak v infinitivu. Udáváme převážně jednoslovné lexémy. V případě víceslovných spojení a části slovníku, kde se objevuje jazyková hra se jmény osob a věcí, uvádíme víceslovná spojení.

Ne vždy bylo možné hesla v reprezentativním tvaru uvést, např. u některých substantiv uvádíme pouze nominativ plurálu (*kozy*, *pindy*), nebo oba tvary nominativu, tedy singulár i plurál (*kec*, *kecy*). V některých případech mají slova se stejným významem dva tvary, uvádíme je jako jedno heslo ve dvojí podobě (*pičovina*, *pičovina*; *vlezdoprdelka*, *vlezprdelka*).

Nominativ jsme neuváděli u jednoho hesla v třetí části s víceslovnými spojeními (*do plynu*). Uvedli jsme tvar, v němž bývá heslo používáno. Tento nereprezentativní tvar jsme označili hvězdičkou *, podobně řeší problém SSČ.

Problém se zachováním infinitivu představují slovesná víceslovná spojení, kdy ani SNČ nebyl schopen určit základní tvar, a tak zachovává formu, v jakém se slova vyskytují (*je to na hovno*; *to mě poser*). Proto jsme k heslu přistupovali stejným způsobem. Nereprezentativní tvar jsme opět označili hvězdičkou *.

5.1.4 Struktura hesla

Heslo nehnízduje. Naším cílem nebylo nasbírat co největší počet jednotek, ale spíše získat podstatná jména, popř. přídavná jména či slovesa, tedy nějaký základ nevhodného slova, s nímž bude moci program dál pracovat. Předpokládáme totiž, že program bude schopný vyhledávat i slova odvozená. I když zatím program není sestaven, už teď předpokládáme zádrhel např. u výrazu *piča*, s jehož základem *pič-* (*zpičený*, *pičovní*) nemůžeme dál pracovat, neboť se vyskytuje u mnoha dalších slov (viz předešlá kapitola).

Slovník mimo heslové slovo uvádí význam¹³ hesla. Ten jsme přebírali hlavně z použitých lexikografických příruček: SSSJČ, SSČ, SNČ a SČFI. Největší podíl významů jsme převzali právě ze SNČ (ať už doslovně nebo s obměnou) a SSSJČ. V některých případech jsme neuváděli význam, ale pouze synonymum slova. U hesel, která jsme ani v jednom slovníku nenašli, jsme buď převzali význam od podobného slova, nebo jsme jej sestavili sami. Pokud má heslo více (odlišných) významů, uváděli jsme jej jako dvě odlišná hesla označená horním indexem (např. *zmetek*¹ a *zmetek*²).

Kontext, který jsme získali z diskusních fór, jsme upravovali. Jednalo se hlavně o opravu chyb (pravopis, gramatika), doplnění interpunkce a diakritiky, převodu velkého písma na malé¹⁴. Nářeční a nespisovné jevy jazyka jsme příspěvkům ponechali. Nezasahovali jsme ani do stylistických chyb, protože by se jednalo o velkou změnu komentáře. V některých případech jsme přidali vysvětlení do hranatých závorek [], aby bylo možné zcela pochopit význam (např. [o Zemanovi; o Klausovi]).

Slovník dále obsahuje zmapování výskytu slov v lexikografických příručkách SSSJČ, SSČ a SNČ, v případě víceslovných spojení navíc i SČFI (o charakteristice příruček viz dále). Pokud se v nich slovo

¹²Záměrně užíváme pojmu víceslovná spojení místo výrazu frazém, neboť si uvědomujeme rozdíl. Některá z těchto spojení mohou mít krátkodobý charakter, jiná naopak ne (některé zachycuje SSSJČ, SNČ nebo SČFI).

¹³V případě žertovného pozměnění hesla uvádíme pravou podobu slova.

¹⁴Pokud bylo v textu zvýrazněno caps lockem pouze jedno slovo či dvě, příspěvek zůstal v původní podobě.

vyskytovalo, uvádíme „*ano*“, v dalším sloupci pak následuje charakteristika slova. Jestliže se hledané heslo ve slovnících vyskytovalo v jiné podobě (např. jiném slovním druhu nebo tvaru), uváděli jsme „*jen podobné slovo (charakteristika)*“. Jako příklad si uveďme heslové slovo „*pod'obanec*“, které se v SSJČ vyskytovalo jen v podobě „*pod'obaný*“. Zaznamenání v našem slovníku tedy vypadá následovně: „*jen pod'obaný (zast., expr.)*“. Dále se často stávalo, že slovník sice námi hledané heslo obsahoval, nezachycoval však stejný význam, proto uvádíme spojení „*v jiném významu*“.

5.2 Charakteristika užitých lexikografických děl

5.2.1 Slovník spisovného jazyka českého (SSJČ)

Slovník spisovného jazyka českého je čtyřdílný výkladový slovník, který byl zpracovaný kolektivem Bohuslava Havránka a vycházel v období 1960–1971, druhého vydání se dočkal v roce 1989. Obsahuje 192 908 lexikograficky zpracovaných slov a na rozdíl od *Příručního slovníku jazyka českého* (PSJČ) získal tento slovník kodifikační status, který platí dodnes. (Jelínek, 1961; Křístek, 1971; Šabatová, 2007; SSJČ, 1989, str. V.–X.; Ústav pro jazyk český AV ČR, v. v. i., @2005–2013)

Během tvoření využívali autoři materiál lexikálního archivu Ústavu pro jazyk český při ČSAV, který vznikl bezmála padesát let.

Tento abecedně řazený slovník se může pyšnit svou promyšlenou makrostrukturou (heslové slovo) i mikrostrukturou (heslový odstavec). Významy slov jsou dokládány typickými slovními spojeními. Oproti PSJČ však slovník kvůli zhuštění neuvádí citáty autorů. Od roku 2011 AV ČR zpřístupnila SSJČ i v digitalizované podobě.

Slovník se snaží zaznamenávat běžnou slovní zásobu, ale i frazeologická a další slovní spojení, čímž může alespoň z části plnit úlohu slovníku frazeologického. Vedle výkladu připojuje SSJČ i výrazy synonymní, čímž do jisté míry plní funkci synonymického slovníku.

SSJČ propracoval stylové hodnocení slov. Protože se jedná o slovník spisovné češtiny, nijak zvlášť spisovné výrazivo neoznačuje, přihlíží ale k stylovým vrstvám spisovným – tedy slovům hovorovým, knižním, básnickým, publicistickým a odborným. Naopak výrazivu nespisovnému věnuje větší pozornost. Mimo kritérium spisovnosti tedy uplatňuje hledisko nespisovnosti, frekvence, dobového výskytu a citového zabarvení.

Zaznamenává slova nářeční, oblastní, lidová, slangová a ve výjimečných případech i argotická. SSJČ dále rozlišuje výrazy obecné češtiny. Dle kritéria frekvence dělí slova na běžně užívaná, řidší a řídká. Z hlediska dobového výskytu odlišuje výrazy, které jsou označeny jako „jevy zanikající nebo zcela zaniklé“ a dále pak výrazy zastaralé a zastarávající.

Co se týká citového zabarvení slov, dělí výrazy na citově zabarvená expresivní slova, které nemají další příznak, a slova expresivní s příznakem zvláštním. Do kategorie slov expresivních se zvláštním příznakem řadíme slova důvěrná, familiární, domácí, mazlivá a dětská. Dále pak slova hanlivá, zhrubělá, vulgární a výrazy eufemistické a ironické.

SSJČ si všímá i významů přenesených, které se ale doposud neustálily jako samostatný význam, z toho důvodu se uvádějí mezi příklady jednotlivých významů.

SSJČ svým pojetím zdaleka přesáhl rámec běžné slovní zásoby. Pro běžné uživatele se navíc stal svým rozsahem velice nepraktický, a tak byla potřeba vytvořit něco méně obsáhlého a vhodnějšího – začal vznikat *Slovník spisovné češtiny* (SSČ), jehož počet hesel je na rozdíl od SSJČ čtvrtinový.

5.2.2 Slovník spisovné češtiny pro školu a veřejnost (SSČ)

Jednosvazkový normativní *Slovník spisovné češtiny pro školu a veřejnost*, jehož hlavními redaktory byli Josef Filipek a František Daneš, vyšel poprvé v roce 1978. Dalších přepracovaných vydání se dočkal v roce 1994 a 2004.

Abecedně řazený výkladový slovník obsahuje 45 336 hesel a 63 872 významů. První vydání se zakládalo na materiálech lexikálního archivu Ústavu pro jazyk český ČSAV.

SSČ je určen nejširším vrstvám společnosti. Snaží se o reprezentaci současné slovní zásoby, kterou aktivně i pasivně používají průměrní uživatelé jazyka, a základního odborného slovníku. Autoři kladli důraz

zejména na doklady stylu odborného, hovorového a publicistického (hlavně polistopadové publicistické texty). (SSČ, 2005, str. 641)

SSČ rozděluje faktory, které ovlivňují charakter příznakových lexikálních jednotek (LJ), do sedmi skupin (os), kde proti sobě stojí protikladné jazykové jevy. (SSČ, 2005, str. 645)

Na časové ose SSČ rozděluje LJ na současné, běžně užívané a nově užívané v protikladu s jednotkami, které jsou užívány mimo současný systém (zastarávající, dříve užívané a historické vázané). S tím je spojeno i hledisko frekvenční (jednotky běžné – řidší).

Z hlediska prostorového a geografického zaznamenává SSČ jen výjimečně nespisovné nářeční výrazy, které pronikly do obecného povědomí. Pozornost však věnuje slovům oblastním a obecně českým, jež se objevují v hovorové vrstvě spisovného jazyka. Z hlediska sociálního dělí jednotky na spisovné a slangové. V rámci spisovného jazyka používá SSČ opozici hovorovost – knižnost.

Důležitý protiklad představuje osa nociónálnost – pragmatičnost. Jednu stranu představuje běžná a odborná slovní zásoba (termíny), druhou stranu pak kladné (zdrobněliny, žertovné, meliorativní a mazlivé prostředky) a záporné (ironické a hanlivé prostředky) expresivní lexikum. SSČ řadí mezi expresivní výrazy i citoslovce. Zahrnuta jsou i slova dětská a zhrubělá.

Funkční hledisko představují termíny, výrazy básnické, publicismy a hovorové prostředky. Uváděny jsou i přenesené významy, ovšem pouze v příkladové části příslušného významu, nikoliv samostatně. Ve výjimečných případech se uplatňuje charakteristika *nevhodné*. Specifickou kategorií tvoří frazémy, kam SSČ řadí i přirovnání, přísloví a pořekadla.

Slovník neuvádí výrazy periferní a řídké, ani nářeční, zastaralé a z našeho pohledu důležitá slova vulgární. SSČ obsahuje přílohy, které zahrnují slovotvorný přehled, seznam zkratk a značek, rodných (křest- ních) jmen a příjmení a zeměpisných jmen.

5.2.3 Slovník české frazeologie a idiomatiky (SČFI)

Slovník české frazeologie a idiomatiky, pod redakčním vedením Františka Čermáka, Jiřího Hronka a Jaroslava Machače, vycházel postupně ve čtyřech svazcích mezi lety 1983 a 1994.

První svazek, který vyšel roku 1983, obsahuje česká přirovnání (obsahuje cca 2 050 plných hesel a 350 redukováných). Druhý svazek, vydaný v roce 1988, tvoří výrazy neslovesné (tedy frazémy nominální, adverbialní) a složené gramatické výrazy (4 039 hesel v základním hesláři slovníku, 743 hesel v gramatických přílohách). Poslední dva svazky vydané v roce 1994 zahrnují výrazy slovesné povahy (9 976 verbálních hesel, 9–10 000 frazémů s abstraktním substantivem). (Kučera, 1995; SČFI, 1983, str. 14; Šabatová, 2007)

Autoři SČFI nechtěli vytvořit kodifikační slovník, ale praktickou deskriptivní příručku pro běžného uživatele i odborníka. Proto se slovník úmyslně zabývá i nespisovnými výrazy. Zdrojem byly dostupné lexikografické práce (SSJČ, PSJČ) a výpisky z románových a dramatických textů, lexikální archiv Ústavu pro jazyk český ČSAV, ale i záznamy mluveného jazyka. Cílem autorů bylo vytvoření slovníku, který by zachycoval současný stav české idiomatiky a frazeologie, zároveň ale aby fungoval jako referenční příručka pro příležitostná i systematická ověření a zjištění.

Běžné heslo v SČFI obsahuje standardní součásti (stylová a gramatická charakteristika, výklad významu, exemplifikace), ale uvádí i transformaci, kontext, valenci, údaje o úzu, etymologii, motivaci a jiné souvislosti příslušného výrazu, dále pak synonyma, opozita, volnější ekvivalenty obdobného významu a cizojazyčné ekvivalenty. (Kučera, 1995)

SČFI charakterizuje každý frazém z hlediska stylové příslušnosti. Na rozdíl od ostatních lexikografických děl označuje i výrazy bez stylového příznaku. K prostředkům běžně mluveného jazyka zařazuje lexémy spisovné (hovorové) a nespisovné (obecně české).

Výrazy, které se používají převážně v projevech psaných a textech slohově nadnesených, jsou charakterizovány jako knižní. Do oblasti projevů psaných se dále řadí i lexémy publicistické a básnické.

Slovník dále vyděluje i výrazy slangové a argotické. Z hlediska expresivity připojuje SČFI další charakteristiky. Výrazy nabývají buď blíže nespécifikovaných expresivních příznaků, nebo uvádí specifický expresivní příznak z hlediska opozic (příznivý – nepříznivý, pochvalný – hanlivý, obdivný – pohrdlivý, žertovný – posměšný, mazlivý – zhrubělý, popř. vulgární). Svě označení mají i výrazy ironické, zdůrazňující a eufemistické. (Kučera, 1995) Frazémy jsou rozlišovány i podle frekvence (řidší, řídké, objevují se ojediněle) a dobového výskytu.

5.2.4 Slovník nespisovné češtiny (SNČ)

V roce 2006 poprvé vyšel *Slovník nespisovné češtiny*, který zachycuje především argot a slang, o čemž svědčí už podtitul díla (*Argot, slangy a obecná mluva od nejstarších dob po současnost. Historie a původ slov*). Dále se pak věnuje obecné mluvě, brněnské a ostravské městské mluvě a v menší míře i slovům nářečným. (SNČ, 2009, str. 9)

Třetí vydání abecedně řazeného slovníku z roku 2009 obsahuje více než 17 000 hesel a mimo jiné se věnuje i etymologii (původu) slov. Slovník byl zpracován autorským kolektivem, jehož vedení se ujal Jan Hugo. Do díla však byla zapojena i širší veřejnost, hesla zasílali čtenáři, některá se podařilo získat přímo od příslušníků Policie ČR. Autoři sbírali materiály pro dílo více než 20 let.

Původním předmětem zájmu SNČ byly nespisovné výrazy německého původu. Představa, jak by měl slovník vypadat, se v průběhu vývoje měnila, přibývalo i oblastí zájmu.

Ze sociálního hlediska pracuje slovník se slangovými a argotickými výrazy. Mezi slangy, které slovník uvádí, patří: administrativní, cirkusový, počítačový, divadelní, drogový, filmový, hasičský, hornický, hudební, karetní, kriminální, letecký, motoristický, policejní, rozhlasový, řemeslnický, sportovní, stavební, studentský, šachový, školní, technický, televizní, tiskařský, trampský, vězeňský, vojenský a železniční slang. (SNČ, 2009, str. 40–41)

Z hlediska časového uvádí slovník zastaralé výrazy, z hlediska frekvenčního pak řidčeji užívané výrazy.

Teritoriální hledisko zahrnuje nářeční výrazy (hanácké, chodské, lašské, moravské; i německé – jiho-německé, bavorské), lidové a obecně české. Na ose hovornost – knižnost uvádí SNČ charakteristiku knižní a hovorový výraz.

Osa nociónálnost – pragmatičnost představuje charakteristiky expresivní (bez zvláštního příznaku), zdrobnělé, dětské, eufemistické, žertovné, ironické, hanlivé, zhrubělé a vulgární výrazy.

SNČ pracuje i s pojmem xenofobní výraz, jímž označoval lexémy s rasovým podtextem. Takovéto charakteristiky ostatní lexikografické příručky nepoužívají. Dále používá například slovo *oslabené* ve smyslu oslabení významu (např. *hajzlík* oslabené -> *hajzl*).

5.2.5 Zhodnocení užitečnosti lexikografických příruček

K lexikografickému zmapování urážlivých výrazů, které se vyskytovaly na internetových diskusích, jsme využívali celkem čtyř výše uvedených a popsaných slovníků: SSJČ, SSČ, SNČ a v případě slovníků spojení i SČFI.

Velmi nápomocný nám byl hlavně *Slovník spisovného jazyka českého*, který je se svými téměř 193 tisíci hesly velice obsáhlý. Zachycuje velké množství námi nasbíraných výrazů. Častým problémem ovšem bylo, že hesla uvedená v SSJČ neuváděla význam, ve kterém jsme heslo našli (např. *hašlerka* v SSJČ v jiném významu). Na slovníku byla navíc znát doba, v níž byl sepsán. Během vývoje, kterým jazyk od vydání díla prošel, se měnila charakteristika lexémů. Například slovo *huba* reflektuje SSJČ jako zhrubělé, ale později vydaný SSČ tento lexém označuje jako obecně český a expresivní. Naopak slovo *hňup* pojímal SSJČ jako expresivní, kdežto SSČ jej označuje jako slovo hanlivé.

Obecně vzato by se dalo říct, že charakteristiky lexémů se ve slovnících ve většině případů neshodovaly.

Slovník spisovné češtiny nám moc nápomocný nebyl, jelikož se se svým obsahem hesel (45 tisíc) nemůže poměřovat se SSJČ (193 tisíc). Navíc sám SSČ uvádí, že vulgarismy nejsou předmětem jeho zájmu, tudíž je vůbec nezachycuje. V menší míře jsme ve slovníku nacházeli slova hanlivá a zhrubělá. Častým jevem bylo i to, že slovník námi hledané heslo obsahoval, ovšem v jiném významu.

Nejvíce nám pomohl *Slovník nespisovné češtiny*, který je navíc i aktuální (nové vydání vyšlo v roce 2009) a zachycuje hanlivé, zhrubělé, vulgární a xenofobní výrazy ve velké míře. Pokud námi hledané heslo nezachycoval SSJČ ani SSČ, velice často jej mapoval SNČ (nebo alespoň výraz podobný).

Slovníku české frazeologie a idiomatiky jsme využili jen velice málo, neboť frazémy a slovními spojeními jsme se v této práci zabývali jen okrajově. Přesto jsme v tomto slovníku některá hesla našli.

Za zmínku stojí i to, že zatímco SSJČ a SSČ uvádí některá slova jako neutrální, SNČ ne. Takový případ nastal například u slova *bolševik*, které SSJČ a SSČ uvádí jako slovo spisovné. Naopak SNČ jej popisuje jako hanlivé, neboť s postupem času byly slovu dodány jisté konotační složky. Podobný případ

bychom mohli pozorovat i u slova *komunista*, ale SSJČ, SSČ jej uvádí jako slovo spisovné a neutrální. Naopak SNČ jej neuvádí vůbec, což může znamenat, že u něj zatím žádné konotační složky nezaznamenal. Ale z kontextu použitého v našem slovníku je zřejmé, že „*komunista*“ prodělal podobný vývoj jako „*bolševik*“ a rovněž v sobě skrývá jisté konotační složky.

Mezi jednoslovné lexémy a jim podobná slova, které nezachycuje (ani v jiném významu) ani SSJČ, SSČ nebo SNČ, řadíme následující: *bezbarák, bomerant, cinkle, čulin, Dojčland'ák, ekoterorista, hnědočech, hnědoslovák, hnědoobčan, kundula, kvazipolitik, modrobolševik, musulman, nepřizpůsobivý, otočkábát, partadement, plast'áky, posranec, poštěvačka, prasopes, pravdoláskář, pseudopolitik, senilát, se-rebrita, shit, transvestil, tuneldědek, uhrin, vietkong, vlezdobruselista, všeználek, vymrdanec, zádel, zmetylovaný, znevýhodněný*; z neslovesných víceslovných spojení pak *cikán černý, černá opice, černá tlama, černý bimbo, kráva nebeská, kurva šikmooká, modrý bolševik, opálená menšina, osoba práce se štítící, přebarvený komanč, rudé komando, vepřová hlava, vyjetá macocha, zateplená menšina*; ze slovesných víceslovných spojení *dostat na budku, mít okenu v tykvi, mlít sračky*.

Po vzoru magisterské diplomové práce Marty Šabatové (*Příznaková slovní zásoba v soukromé korespondenci mládeže*) jsme se rozhodli uzavřít podkapitulu věnovanou charakteristice užitých lexikografických děl pomocí podobné tabulky, kterou ve svém díle Šabatová použila (Šabatová, 2007). Na rozdíl od ní jsme však klasifikaci příznakových lexémů v užitých lexikografických dílech sestavovali podle seznamu zkratk slovníku, nikoliv podle pojednání o uplatnění stylistické klasifikace jako Šabatová. Navíc jsme ponechali uvedené pouze takové LJ, které se vztahují k tématu naší práce, tj. LJ spojené s urážlivými, vulgárními a rasistickými výrazy, nebo zkratkami, které se v našem slovníku v příloze vyskytovaly.

SČFI	
Stylová příslušnost	
<i>LJ slang</i>	slang
<i>LJ argot</i>	argot
Expresivita	
<i>LJ expresivně bezpříznakové</i>	expr
<i>eufemismus</i>	euf
<i>LJ příznivé – nepříznivé</i>	přízn – nepřízn
<i>LJ pochvalné – hanlivé</i>	pochv – hanl
<i>LJ obdivné – pohrdlivé</i>	obd – pohrd
<i>LJ žertovné posměšné</i>	žert – posm
<i>LJ mazlivé – zhrubělé, vulgární</i>	mazl – zhrub, vulg
<i>LJ ironické</i>	iron

Tabulka 5.1: **Klasifikace příznakových lexémů ve Slovníku české frazeologie a idiomatiky (SČFI).**

	SSJČ	SSČ	SNČ
ČASOVÉ HLEDISKO			
<i>LJ zastarávající</i> <i>LJ zastaralé</i>	poněk. zast. zast./†	zast.	zast.
FREKVENČNÍ HLEDISKO			
<i>LJ řidší</i>	ridč.	ridč.	ridč.
TERITORIÁLNÍ HLEDISKO			
<i>LJ nářeční</i> <i>LJ oblastní</i> <i>LJ obecně české</i> <i>LJ lidové</i>	nář. obl. ob. lid.	nář. obl. ob. lid.	nář. ob. lid.
SOCIÁLNÍ HLEDISKO			
<i>LJ slangové</i> <i>LJ argotické</i>	slang. arg.	slang. arg.	slang. arg.
HOVOROVOST – KNIŽNOST			
<i>LJ hovorové</i> <i>LJ knižní</i>	hovor. kniž.	hovor. kniž.	hovor. kniž.
NACIONÁLNOST – PRAGMATIČNOST			
<i>LJ expresivní bez zvl. příznaku</i> <i>LJ zdrobnělé</i> <i>LJ žertovné</i> <i>LJ eufemistické</i> <i>LJ dětské</i> <i>LJ ironické</i> <i>LJ hanlivé</i> <i>LJ zhrubělé</i> <i>LJ vulgární</i>	expr. zdrob. žert. euf. dět. iron. hanl. zhrub. vulg.	expr. zdrob. žert. euf. dět. iron. hanl. zhrub.	expr. zdrob. žert. eufem. dět. iron. hanl. zhrub. vulg.
FUNKČNÍ HLEDISKO			
<i>LJ publicistické</i>	publ.	publ.	
PŘENESENÁ UŽITÍ LJ			
<i>LJ přenesené</i>	přen.	přen.	přen.

Tabulka 5.2: Klasifikace příznakových lexémů ve Slovníku spisovného jazyka českého (SSJČ), Slovníku spisovné češtiny pro školu a veřejnost (SSČ) a ve Slovníku nespisovné češtiny (SNČ).

5.3 Příznakovost lexikálních prostředků

Příznakové lexikální prostředky jsou takové prostředky, které se vyznačují charakteristickým a rozlišujícím rysem¹⁵. (Hladká, 2012, str. 7)

K příznakovým lexikálním prostředkům řadíme jednotky, jejichž příznakovost je založena na příslušnosti k některému útvaru nebo poloútvary národního jazyka, dále pak k typu komunikace nebo k jednotlivým stylům a stylovým vrstvám. V těchto případech se příznakovost lexikálních prostředků představuje jako rozložení na ose spisovnost – nespisovnost.

K příznakovým lexikálním prostředkům dále přiřazujeme jednotky, „jejichž příznakovost se zakládá na vyjádření citového, hodnotícího a volního postoje mluvčího, tedy prostředky s rysem expresivity, sociality, evaluace.“ (Hladká, 2012, str. 8)

Mezi příznakové prostředky se přiřazují i „výrazy zastaralé a zastarávající, výrazy dobové a prostředky příznakové z hlediska frekvence.“ (Hladká, 2012, str. 8)

V této části se krátce věnujeme kritériím, podle nichž slovníky dělí charakteristiky příznakovosti LJ. V krátkosti se zaměříme na hledisko časové, frekvenční, teritoriální, sociální, funkční, přenesení významu LJ a osu hovorovost – knižnost. Pokud tato hlediska námi zpracovaný slovník reflektuje, pokusíme se je doložit příklady¹⁶ ze slovníku.

Podrobněji se pak zaměříme na osu nociónálnost – pragmatičnost, kam spadají expresivní výrazy. Pozornost upřeme hlavně na slova hanlivá, zhrubělá a vulgární.

5.3.1 Časové hledisko

Jak už bylo řečeno v kapitole charakterizující jednotlivé slovníky, časové hledisko je jedním z kritérií, podle nichž slovníky dělí slova a slovní spojení.

Do této kategorie spadají lexikální jednotky současného systému (běžně užívané, neologismy) a výrazy mimo dnešní systém (zastarávající, dříve užívané a historické).

Charakteristiku zastaralosti (*zast.*) uvádí SSSJ u slov *pod'obaný, póvl, Talián*. Jako zastarávající (*poněk. zast.*) uvádí SSSJ slova *socan, somrák*. Oproti tomu SNČ reflektuje slova *socan (hanl.), somrák* i *Talián*, ale příznak *zast.* neuvádí, což znamená, že slova kdysi zastaralá se dnes opět začala používat.

5.3.2 Frekvenční hledisko

Prostředky příznakové z hlediska frekvence zahrnují charakteristiky řidší a řídké. Náš slovník zachytil řidčeji (*řidč.*) užívaný lexém *kecka*.

5.3.3 Teritoriální hledisko

Z hlediska teritoriálního bychom mohli rozčlenit lexikální jednotky na nářeční¹⁷, oblastní¹⁸, obecně české¹⁹ a lidové²⁰.

Jako lexém oblastní (*obl.*) je charakterizováno slovo *drbat* (ve významu mít pohlavní styk), používané na území Moravy. Obecně česká (*ob.*) se v našem slovníku objevovala občas (*estébák, kašpar*). Jako slovo nářeční (*nář.*) je charakterizováno např. *čuňa, jebat*. Slova lidová jsme nezachytili.

¹⁵Existuje několik různých pojetí příznakovosti lexikálních prostředků. Jejich klasifikace se zakládá na tom, zdali přistupujeme k jevu z hlediska lexikografického, lexikologického nebo stylistického, zdali uplatňujeme hledisko systémové nebo „situačně-kontextové“ nebo opírá-li se koncept příznakových lexikálních prostředků o přístup sociolingvistický. (Hladká, 2012, str. 7)

¹⁶Podotýkáme, že některá slova mají často více než jednu charakteristiku (např. *hajzl – vulg., hanl.*), proto mohou být jako příklad uváděny vícekrát.

¹⁷Jako **nářeční (dialektická) slova** se uvádějí takové lexémy, s nimiž se mluví jen na určitém (ohraničeném) území.

¹⁸**Oblastní slova (regionalismy)** jsou taková slova, která se často užívají na určitém území (Morava, Čechy, Slezsko), které je rozsáhlejší než území nářeční.

¹⁹**Obecná čeština (interdialekt)** je útvarem, který slouží k běžnému dorozumívání. Na západní části našeho území tvoří základ běžné mluvy. (Čechová, 2008, str. 59)

²⁰**Lidová slova** se používají v lidovém jazyce, který má hlavně komunikativní funkci. Omezuje se většinou na projevy ústní a soukromé.

5.3.4 Sociální hledisko

K příznakovým lexikálním jednotkám z hlediska sociálního se řadí slang²¹ a argot²².

Argotické výrazy (*arg.*) se objevovaly velmi zřídka – *práskač*, *šlapka*, *buzík* (podle SSJČ). Za zmínku stojí fakt, že zatímco slovník tato slova označuje jako argotická, jiné slovníky je pojímají jinak: například SNČ uvádí slovo *práskač* jako slovo policejního slangu a slovo *buzík* jako vulgární; SSČ označuje slovo *šlapka* jako hovorové.

Lexikální prostředky náležící k určitému slangu (*slang.*) uvádí pouze SNČ – své zastoupení má slang drogový (*fet' ačka*, *fetka*), policejní (*běhna*, *tunelář*), vězeňský (*kuřbuřt*) a školní (*šplhoun*).

5.3.5 Funkční hledisko

Z funkčního hlediska jsou výrazy označovány charakteristikou básnickou nebo publicistickou. Podle charakteristiky, kterou uvádí SSČ, se v našem slovníku vyskytuje pouze jeden publicismus²³ (*publ.*) – *tunelář*. SNČ tento výraz řadí do policejního slangu. Básnické výrazy, poetismy²⁴, se v námi zachycených slovníkových heslech nevyskytují.

5.3.6 Přenesení významu lexikálních jednotek

Ačkoliv příznak přeneseného²⁵ užití (*přen.*) mají v našem slovníku pouze výrazy *čarodějnice*, *čmáral* a *skřet*, mohli bychom říci, že velká část námi zachycených slov by měla obsahovat charakteristiku „přeneseně“, neboť jak vyplývá z prvního pohledu na náš slovník, uvedené lexikografické příručky nezachycují nový význam již utvořených slov (např. *vykopávka*, *vemeno*). Významovým posunům (přenesení významu) jsme se podrobněji věnovali v předešlé kapitole.

5.3.7 Hovorovost – knižnost

Frekventovanější kategorii tvoří slova hovorová²⁶ (*hovor.*) (*fašoun*, *kolohnát*, *mizera*, *náfuka*, *neřád*, *notorik*, *pysk*, *srab*, *šlapka*, *trdlo*), charakteristika knižnosti²⁷ (*kniž.*) se objevuje pouze u slov *skřet* a *bezmozkový*.

5.3.8 Nocionálnost – pragmatičnost

Podle SSČ představuje důležitý protiklad osa nocionálnost – pragmatičnost. Tato kategorie zahrnuje výrazy odborné na jedné straně (nocionálnost), na straně druhé pak výrazy z expresivní oblasti (pragmatičnost).

Expresivní slova jsou taková slova, která mají příznak citového zabarvení (emocionality). Dělíme je na expresivní slova s citově kladným příznakem a na expresivní slova s citově záporným příznakem.

Mezi expresivní slova s citově kladným příznakem řadíme slova důvěrná (familiární), domácká (hypokoristika), zdobněliny (deminutiva), eufemismy a dětská slova.

K expresivním slovům s citově záporným příznakem přiřazujeme slova hanlivá (pejorativní), zveličená (augmentativa), zhrubělá, vulgární a dysfemismy.

²¹ **Slang** bychom mohli definovat jako jazyk určitých sociálních skupin, jako něco, co je stmeluje. Slangy se vyznačují hlavně expresivitou a metaforičností. Jejich podstatou velmi často bývá jazyková hra.

²² Mezi **argotismy** se řadí výrazy, které používají příslušníci společenského podsvětí. Jejich charakteristickým rysem je snaha o utajení významu slova ostatním vrstvám společnosti.

²³ **Publicismy** jsou jednoslovné nebo víceslovné lexikální jednotky, které jsou nepříznakové v lexikální vrstvě publicistiky. V jiných vrstvách jsou však příznakové.

²⁴ **Poetismy** jsou slova, která jsou používána v uměleckém stylu, ve většině případů se jedná o slova zastarávající či zastaralá.

²⁵ Pomocí sémantických procesů vznikají nové významy slov, mluvíme o významech **přenesených**.

²⁶ **Hovorová čeština** je mluvená forma spisovné češtiny. Na rozdíl od psané formy se v hovorové češtině používají i nespisovné výrazy. Svou roli zde hraje i tzv. univerbizace, tedy tvoření jednoslovného výrazu z víceslovných pojmenování. (Filipec, 1985, str. 106)

²⁷ **Knižních slova** se používají v literárních a psaných textech s oficiálním charakterem. Vyskytují se i v mluvených projevech pronášených při slavnostních příležitostech.

Protože slova, která zachycuje náš slovník v příloze, jsou charakteristická svou „dávkou“ expresivity, věnujeme jim trochu větší pozornost (nikoliv pouze v poznámkách pod čarou). Nezabýváme se ale všemi zmíněnými podkategoriemi.

Zaměříme se pouze na ty, jejichž charakteristiky jsou uvedeny ve slovníku v příloze (tedy slova s příznakem expresivním; žertovné a posměšné prostředky ze slov citově kladných zdobněliny, dětská slova; ze slov citově záporně zabarvených pak slova hanlivá, zhrubělá a vulgární).

Charakteristikou expresivní (*expr.*) jsou označena slova, která jsou expresivní, nemají ovšem další zvláštní příznak (*blázen, fúrie, chrchloun, kolohnát, pijan, slinta, škatule, uplakánek, žvanil, žvatlal*).

Z hlediska expresivních slov se zřídka objevovaly i žertovné prostředky (*žert.*) u slov *blbštajn, dokum-damenty* a posměšné (*posm.*²⁸) u LJ *kancelářská krysa, stará škatule*.

Expresivní slova s citově kladným příznakem

Příznak zdobnělý (*zdrob.*) mají slova, jejichž význam je deminutivní. Typicky jsou tvořena pomocí zdobnělých přípon. Tento příznak slovník zachycuje pouze u tří lexémů: *hajzlík, Japončik, kozičky*.

Příznak dětský výraz (*dět.*) obsahují slova, která používají dospělí jedinci při kontaktu s malými dětmi. Tato charakteristika se vyskytuje u slova *pindík*.

Expresivní slova s citově záporným příznakem

Při bližším pohledu na slovník v příloze jasně vyplývá, že většina hesel obsahuje právě jeden z těchto citově záporných příznaků (hanlivý, zhrubělý, vulgární), neboť jejich cílem je urazit druhou osobu. Než se v krátkosti začneme věnovat rozdílu mezi slovy hanlivými, zhrubělými a vulgárními, rádi bychom upozornili na to, že charakteristiky se liší slovník od slovníku (co SSJČ uvádí jako vulgární, pojímá SNČ jako hanlivé a naopak).

Příznak pejorativnosti neboli hanlivosti (*hanl.*) obsahují slova, která vyjadřují negativní hodnotící nebo emotivní příznak. Tato slova mají často urážlivý charakter vyjadřující opovržení (např. *bastard, děvka, fašoun, krůta, machna, megera, pámbičkář, Skopčák, židák*).

Slova zhrubělá (*zhrub.*) představují takové lexémy, které se používají k vyjádření nebo vzbuzení odporu nebo při pohoršení nad určitou věcí (např. *blb, fňukna, hovado, idiot, mrcha, ožrala, štětka, tlama, vůl*).

Sprostá slova neboli vulgarismy (*vulg.*) jsou lexikální jednotky, jež vyjadřují negativní a silně emotivní postoj člověka k určité věci. Lidé tato slova používají při hrubém společenském vyjadřování, které je považováno za nevhodné.

Jelikož slova vulgární tvoří primární zájem naší práce, rozhodli jsme se vypsát všechna, která alespoň jednou slovníky (SSJČ, SNČ, SČFI²⁹) uváděly s vulgárním příznakem:

běhna, bobr, bordel, bréca, brko, buchta, buk, bukál, bukvice, buzerant, buzík, buzna, cecky, coura, cuchta, čurák, čůza, dávačka, drbat, držka, dudy, flundra, fošna, frantík, frnda, fuchtle, grázl, hadice, hajzl, hajzlbába, homokláda, homouš, hovno, huba, chlastoun, chlívák, jebat, jebka, klacek, kláda, klátiti, kokot, koule, kozičky, kozy, kripl, ksindl, kunda, kurevník, kurva, kurvadrát, kurvafix, ližprdelka, mamrd, melouny, mrdat, mrdka, nabíječka, ocas, ochlasta, ojet, (v)ošoust, parchant, péro, píča, pička, pičovina/pičovina, pičulin, pičus, píchat, pind'our, pipina, pizda, posera, posrat, prasák, prcat, prcina/prcinka, precka, prdel, pták, rajda, rozhod'nožka, rypák, řit'opich, sráč, sračka, (srágora), sviňák, šoustat, šprcguma, šprcka, štětka, šukat, šukna, šulin, teplomet, teplouš, tlustoprđ, trkat, vemena, vlez(do)prdelka, vůl, vyčuránek/vyčů- ránek, zjeb, zkurvysyn a zmrđ; z hlediska víceslovných spojení: slepičí prdelka, držet hubu, honit si ptáka/ho, jít do prdele, leštit si bambus, nesrat boha, políbit prdel, vysrat si oko.

Mimo výše zmíněná hlediska příznakovosti (hledisko časové, frekvenční, teritoriální, sociální, ...) pracuje *Slovník nespisovné češtiny s xenofobními výrazy (xenof.)* Tímto příznakem jsou označena hesla s rasovým podtextem. Patří k nim lexémy: *bakelit, cigoš, cikorka/cikorky, Čobol, Čobolák, kofola, morgoš, negr, přičmoudlý, rákosník, rejžožrout, šikmovoký, šikmoň, t'aman, t'iňt'ong a Vietnamčik*; z hlediska více- slovných spojení: *černá huba*.

²⁸Tento příznak používá SČFI.

²⁹Jak jsme již předesílali, SSČ nereflektuje vulgární výrazy.

6. Návrh programu vyhledávající urážlivé, vulgární a rasistické výrazy

Cílem této kapitoly je popsat slovně pravidla pro program, který bude vyhledávat ve vstupním textu (diskusních komentářích) nevhodná slova, jež bude porovnávat se svou vlastní databází slov. Pokud nalezne shodu, příspěvek buď odstraní, nebo jej pošle ke schválení administrátorovi (u hanlivých nebo vícevýznamových slov).

Zdůrazňujeme, že program doposud není vytvořen, je tudíž možné, že některá pravidla budou nerealizovatelná.

V návaznosti na tuto práci vznikne program, jehož úspěšnost se bude teprve testovat. Podle ní se bude program „doladovat“. Proto tato pravidla nejsou pevně daná a neměnná.

6.1 Databáze

Ve vstupním textu (diskusních komentářích) bude program vyhledávat nevhodná slova, která bude porovnávat s vlastní databází³⁰, kterou jsme rozdělili do čtyř kategorií: posouzení vhodnosti slov správcem diskusí, ruční zjednoznačnění, seznam problémových osobností a seznam zakázaných slov.

Konverzace na diskusních fórech je neformální, nevede se jen ve spisovném jazyce. Mimo neutrální slova používají uživatelé i slova expresivní nebo hanlivá³¹. Ne každé takové slovo je nutné hned mazat, do kontextu se může hodit, nemusí nikoho urážet a může další uživatele i pobavit. Z tohoto důvodu jsme se rozhodli vytvořit kategorii se slovy, jejichž vhodnost posoudí správce diskusí. Velká diskusní fóra mohou posoudit slovo od slova, a pokud usoudí, že některá z nich nejsou vhodná, mohou je lehce přesunout do seznamu zakázaných slov.

Záleží na administrátorech, jak velkou volnost vyjadřování chtějí uživatelům poskytnout. Obecně by se dalo předpokládat, že velké diskuse usoudí, že „ruční“ zjednoznačnění této kategorie může být administrativně velmi náročné, tudíž je rovnou přesunou do seznamu zakázaných slov, čímž se do velké míry sníží i počet příspěvků, neboť seznam zakázaných slov bude tak obsáhlý, že to pro uživatele bude znamenat tak velké omezení svobodného projevu, že přestanou přispívat do diskusí. To může být pro velké diskusní servery výhoda (zvláště pokud chtějí menší množství komentářů). Naopak menší servery pravděpodobně upřednostní ruční zjednoznačnění takových slov, neboť nemají tolik příspěvků a dokážou je všechny kontrolovat.

Při hlubším pohledu na databázi se zakázanými slovy by se mohlo zdát, že některá slova přeci jen nejsou tak hrubá, aby musela být zakázána. Přesto jsme se rozhodli je sem zahrnout kvůli tomu, že jejich úkolem je buď pouze urážet druhého uživatele (nebo třetí osobu), nebo kvůli tomu, že byla příliš často používána v silně urážlivém kontextu. Pokud s tímto zařazením má někdo problém, není těžké slovo z databáze smazat nebo jej přesunout do jiné kategorie.

³⁰Databáze by měla být jednoduchá a moderátorům diskuse volně přístupná, aby ji mohli sami aktualizovat a jednoduše spravovat (vhodná i pro lidi, kteří nejsou příliš technicky zdatní).

³¹Pokud navíc uživatel používá v příspěvku hanlivá a expresivní slova, zvyšuje se pravděpodobnost, že bude používat urážky, které program nemusí znát, moderátor je během schvalování příspěvku může zaznamenat a přidat do databáze.

6.1.1 Posouzení vhodnosti slov správcem

Amík, Anglán, Arabáč, blázen, blbina, blbost, bonzák, bréca, budižkničemu, buran, cigán, čarodějnice, Čobol, Čobolák, čoklař, čuňák, čuně, darmožrout, dinosaurka, Dojčland'ák, donašeč, drbna, držgrešle, ekoterorista, estébák, eurohujer, fakan/fagan, fantasmagor, fašista, fet'ačka, fetka, fiflena, flákota, fňukna, Frantík, gauner, hejhula, Helmut, holota, hovado, hrabáč, hulvát, chamrad', cháttra, chlívák, Japončik, kapoun, kašpar, kecal, kecka, kecy, klepna, kolohnát, kořala, kozatka, kozičky, kraksna, krkoun, křupan, ksicht, ksindl, kvazipolitik, kydy, lotr, ludra, lump, lůza, mamlas, maškara, mimoň, mizera, mluvka, modrobolševik, mužatka, nabob, nácučka, náfučka, nálevna, nána, nanynka, narcis, násoska, negramota, Němčour, nenažranec, neřád, notorik, nula, nýmand, ořežprut, otočkabát, padavka, pakáž, pako, pámbičkář, papaláš, patolízal, pazgrivec, pijan, pimpulín, pindík, pitomec, plebs, plešoun, počůránek, polidi, poleno, pometlo, potvora, pouzdro, póvl, práškač, pravdoláškař, prsařce, pseudopolitik, psycho, rašple, robič, ropucha, roštěnka, rudokožec, rudoch, Rusák, rusofob, sebranka, sketa, skrbílík, skřet, somrák, sou-druh, srab, srágora, srágotka, sůva, sviňák, svině, škatule, škrt, šmejda, špekoun, špindíra, šplhoun, Talián, titulblabolista, trapák, trapka, trdlo, trouba, trubka, truhlík, trumpeta, tunelář, tupec, tupoun, tydýt, úchyl, úchylák, uplakánek, verběž, Vietnamčik, všeználek, vtěrka, vůl (vole), vyčůránek, vychcánek, výpitek, vyžírka, závislák, zbabělec, zlatokopka, zmetek, zmetylovaný, zmije, zobák, zombie, zrůda, žabař

6.1.2 Ruční zjednoznačení

bakelit, becherovka, bobr, bolševik, briketa, brko, buchta, buk, bukvice, dobytek, čmáral, dežo, díra, dírka, drbat, fosílie, fošna, gumák, hadice, hašlerka, hlen, hnůj, husa, chobot, jelito, kanál, kecal, klacek, kláda, kladélko, klátit, klobása, kofola, komáři, komunista, koule, koza, kráva, kreatura, krůta, melouny, micina, močůvka, multikulturní, mūra, nabíječka, narcis, nepřizpůsobivý, ocas, opičák, oranžáda, osel, parazit, péro, píchat, pijavice, plácal, plast'áky, plevel, poleno, pouzdro, pták, pysk, ropucha, roštěnka, roura, sle-pice, škrt, tele, trkat, trouba, trpaslík, trubka, truhlík, trumpeta, uherák, udavač, ušmudlánek, utiskovaný, vemena, vemeno, vrták, vůl (vole), vykopávka, vyvolený, zkamenělina, zmije, znevýhodněný, zobák, židle, žumpa, žvanil, žvatlal

6.1.3 Seznam problémových osobností

Některé české osobnosti jsou mezi lidmi neoblíbené, u těchto jmen se objevovaly urážky ve velké míře, proto je vhodné tato slova automaticky vyhledávat v diskusních příspěvcích (např. jen příjmení, nebo obojí – jméno a příjmení) a posílat moderátorovi ke schválení. Protože je každý server jinak orientovaný, měl by si svou databázi sledovaných témat vytvářet sám. My jsme námátkou vybrali časté náměty diskusí na bulvárním serveru *Blesk.cz*.

Agáta Hanychová, Dagmar Havlová, David Rath, Ewa Farná, Gábina Partyšová, Iveta Bartošová, Jan Fisher, Jiří Paroubek, Jiří Pomeje, Karel Gott, Karel Schwarzenberg, Leoš Mareš, Lucie Bílá, Lucie Borhy-ová, Michaela Ochotská, Miloš Zeman, Mirek Topolánek, Miroslav Kalousek, Ornella Štiková, Petr Nečas, Radek John, Tomio Okamura, Václav Klaus, Vendula Svobodová

6.1.4 Seznam zakázaných slov

babochlap, bastard, bed'arák, běhna, becher, bestie, bezbarák, bezmozek, blb, blbec, blekota, bomerant, bordel, bukál, buzerant, buzík, buzna, cecky, cigoš, cikorky, cinkle, coura, cuchta, černoprdelník, čubka, čulin, čurák, čuramedán, čuromrd, čůza/čůza, dávačka, debil, degen, dement, děvka, držka, dudy, dy-lina/dilina, echtyl, fašoun, fekálník, flundra, frnda, fuchtle, fúrie, grázl, hajzl, hajzlbába, harant, hnědo-čech, hnědoobčan, hnědoslovák, hnojařka, homokláda, homouš, honibrk, hňup, hovno, huba, hulpéro, chlastoun, chrchloun, idiot, imbecil, jebat, jebka, kazimrd, kazišuk, klapačka, klátič, kokot, kokrhel, ko-manč, komančeros, komouš, komundíra, kozy, kretén, kripl, kunda, kundoment, kundula, kurevník, kurva, kurvadrát, kurvafix, kuřbuřt, ližprdelka, magor, machna, mamrd, mastodontka, masturbant, megera, men-tál, morgoš, mrdat, mrdka, mrdna, mrdoun, mrcha, musliman, nácek, nefakčenko, negr, nemakčenko, nemaklo, netáhlo, nicnedělač, norek, notor, ochlasta, onanista, ožrala, pablb, parchant, partadement,

piča, píča, pičisko, pičovina, píčovina, pičulin, pičus, pinda, pind'our, pindy, pipina, pipka, pizda, po-držtaška, posera, poslanec, posránek, posrat, pošahanec, poštěvačka, pošuk, prasák, prasopes, prcat, prcina, prcinka, prcka, prdel, přičmoudlík, připrclík, psycho, Pšonek, rajda, rákosník, rejžožrout, retard, rozhod'nožka, rudoun, rypák, řit'olezec, řit'opich, semetrika, senilát, serebrita, shit, Skopčák, slinta, sli-zoun, smažka, socan, socka, sockomouš, soudržka, spermohlt, sráč, sračka, šikmooký, šikmoň, šikmočko, šlapka, soustat, šprcguma, šprcka, štetule, šukat, šukna, šulin, t'aman, teplomet, teplouš, tiť'an, tiň'ong, tlama, tlučhuba, tlustoprđ, tragéd, trajda, trand'ák, transka, transvestil, trol, trotl, trusnice, uhrin, viet-kong, vlezdoprdelka, vošoust, vymaštětec, vymrdanec, vypatlanec, zádel, zjeb, zkurvysyn, zmrđ, židák, zpičený, z víceslovných spojení pak: držet hubu, držet klapačku, černá huba, černá opice, černá tlama

grafické:

Zadek: (._.), (._!_)

Prsa: (.) (.)

Seznam mazaných slov, které zohledňují kontextv

Program bude navíc zohledňovat kontext u vícevýznamových slov. Pokud některá z uvedených slov nalezneme v blízkosti klíčového výrazu (v úvahu se bude brát jen několik pozic/znaků), příspěvek bude automaticky odstraněn.

alpinista: anální, rektální, řitní

bobr: neoholený, zarostený

brko: dlouhý, hulit (hulila mu), kouřit (kouřila mu), malý, uschlý

husa: blbá, hloupá, jsi, seš, si, pitomá, ty

kráva: blbá, pitomá, děsná, hrozná, jsi, si, seš, ty

ocas, péro, pták: honit, hulit, kouřit, leštit, mastit, uříznout

pysk: zavřít, držet

dobytěk, hovado, husa, můra, osel, tele, vemeno, vůl: blbý, jsi, pitomý, seš, si, pitomý

6.1.5 Zohlednění derivovaných slov

V našem slovníku jsme nezohledňovali slova derivovaná, jelikož tento úkol zvládne náš specializovaný program. Bude pracovat s kořeny slov, které se vyskytují na seznamu zakázaných výrazů: např. *blb* (*blbý, nablblý, zblbý, blbnout, pablb*), *debil* (*debilní*), *kurva* (*zkurevný, kurevský, kurvička*). Ne všude to ovšem bude možné – například u slova *piča* (*zpičený, pičulky, pičovní*), kde by program mohl pracovat se znaky „*pič*“. Jak dokládá korpus SYN2010, takovéto spojení obsahují i další slova (*špička, vyslepičit, lupič, kapička*). Slovo *zpičený* jsme tedy doplnili na konec databáze zakázaných slov.

6.1.6 Zohlednění skloňování, časování

Program bude rovněž zohledňovat skloňování (např. *nácek – nácka, náckovi*) a časování (např. *šukat – šuká, šukali, šukají*), k čemuž využijeme morfologického analyzátoru. Ten umožňuje práci přímo se základním tvarem slova.

6.2 Práce s diskusními příspěvky

6.2.1 Sledování negativního hodnocení příspěvků uživateli

Program by měl brát ohled na příspěvky, které uživatelé hodnotí záporně („-“). Na mnohých fórech takto lidé příspěvky hodnotí, pokud obsahují nevhodné výrazy nebo nevhodný názor. Na druhou stranu mohou obsahovat „pouze“ názory, s nimiž uživatelé nesouhlasí, ty ale nemusí mít nevhodný podtext.

V případě, že fórum nemá takovéto hodnocení příspěvků k dispozici, mohl by program sledovat emotikony (smajlíky) zobrazující palec dolů (jako reakce na příspěvek).

6.2.2 Sledování méně používaných slov, které se začaly objevovat ve větší míře

Ačkoliv program nedokáže rozpoznat, jestli nebyl určitému slovu přidán nový význam, může alespoň sledovat, jestli se některá méně používaná slova nezačala v poslední době objevovat ve větší míře. Tímto způsobem pak dokáže odhalit nové urážky (např. *kofola*, *hašlerka* jako označení Romů).

Program tedy bude pracovat s databází slov, u nichž bude uvedena frekvence výskytu. Tato databáze může být sestavena například podle *Frekvenčního slovníku češtiny* a může obsahovat několik tisíc nejčastěji používaných slov. Pokud se některé z méně frekventovaných slov začne objevovat ve větším množství, program pošle moderátorovi upozornění, který situaci posoudí.

6.2.3 Sledování editační (Levenstheinovy) vzdálenosti

K odhalení nově utvořených urážek nám bude pomáhat tzv. *Levenstheinova vzdálenost*, která vyjadřuje, jak moc jsou si dva textové řetězce podobné. Tuto podobnost vyjadřují počty změn (tedy nahrazení, vložení nebo vypuštění některého ze znaků), jež jsou potřeba ke změně jednoho řetězce ve druhý. (NITS)

Jisté diskuse nedovolují odeslat příspěvek, pokud obsahuje některé ze zakázaných slov. Proto uživatelé obcházejí toto nastavením obměnou slova, přidáním nebo vypuštěním určitého znaku.

Takto nejspíše vzniklo slovo *bomerant* (podobné slovu *buzerant*). Náznornou ukázkou vložení znaku si můžeme demonstrovat na řetězci *perdel* (*prdel*), vypuštění znaků dokazuje častý výraz *kua* (*kurva*), který je na diskusních fórech oblíbený.

Tyto nové výrazy bude program porovnávat se seznamem zakázaných slov, a pokud nalezne shodu, bude je posílat moderátorovi ke schválení.

Levenstheinovu vzdálenost bychom mohli využít i při jazykové hře se jmény osobností (např. *Kalousek – Kradousek*, *Ornella – Pornella*), která je založena na obměně slova, přidání nebo odebrání znaku. Proto bude program sledovat editační vzdálenost i u seznamu problémových osobností a shodu bude odesílat moderátorovi ke schválení.

6.2.4 Rozpoznávání známých a neznámých slov

Program by měl navíc umět pracovat se slovy, které zná, a které nezná. Podle toho pak pozná nově utvořená slova. Musí tedy mít databázi pokrývající současnou slovní zásobu (spisovnou i nespisovnou), k tomu může použít například hesla SSSJČ a SSČ i SNČ v případě lexika nespisovného, popř. i korpusy.

Dále pak musí umět pracovat s chybami v textu, rozpoznat je a vyhodnotit je jako slova známá (např. *chibička*). Program rovněž musí umět pracovat s texty, jejichž uživatelé nepoužívají diakritiku (*vcera jsem si koupil obrovskou lodku*). K tomuto všemu opět může posloužit Levenstheinova vzdálenost.

Neznámá slova pak bude program posílat administrátorovi ke schválení. Pokud moderátor diskuse rozhodne, že se jedná o slovo nezávadné, program jej zařadí do databáze známých slov (např. pokud program vyhodnotí slovo *šeptanda* jako neznámé a moderátor rozhodne o jeho nezávadnosti, neznámé slovo se přesune do slov známých).

Zpočátku bude tato práce náročná, avšak po „odklikání“ nezávadných slov, které rozšíří databázi slov známých, začne příspěvků s potřebou ruční desambiguace ubývat a po nějaké době bude program posílat jen neznámá slova, mezi nimiž se budou objevovat hlavně urážlivé, vulgární a rasistické výrazy.

6.2.5 Práce se vstupním textem

Vyhledávání velkých písmen

Program by měl vyhledávat ty části vět, které jsou napsány velkými písmeny – s jejich pomocí uživatelé zvýrazňují části textu. Většinou se mezi těmito zvýrazněnými částmi nacházejí právě urážky: *přestaň tu vopruzovat*, *ty KRÁVO!* Nalezené příspěvky pak pošle administrátorovi ke schválení.

Převod na malá písmena

Program převede vstupní text na malá písmena, s těmi pak bude dále pracovat a vyhledávat mezi nimi nevhodné příspěvky.

BLbeC -> blbec
KRÁVO -> krávo

Rozpoznávání diakritiky

Náš program by měl brát v potaz i to, že někteří uživatelé nepoužívají diakritická znaménka (*ty ses uplnej blbecek*), někteří zase použitím diakritiky obcházejí blokovací automaticky (*seš debil*).

Z tohoto důvodu by měl program odstranit diakritiku jak u seznamu zakázaných slov, tak u všech vstupních textů (příspěvků), a pak je mezi sebou porovnávat. Pokud by se některé řetězce znaků shodovaly, program by je smazal.

blbecek -> blbecek
kráva -> krava
debil -> debil

Odstranění přebytečných interpunkčních znamének, zmnožení písmen

Program odstraní z textu všechna interpunkční znaménka nebo je alespoň nebude brát v potaz, protože uživatelé s jejich pomocí obcházejí blokovací automaty různými způsoby: *m.r.d.k.a*, *pi-ča*, *krá. va*, *kok*ot* atd. Pokud program odstraní interpunkční znaménka, jednoduše pak rozpozná výrazy (*mrčka*, *piča*), které jsou součástí zakázaného seznamu, a odstraní je.

m.r.d.k.a -> mrčka
pi-ča -> piča
krá. va -> kráva
krááááááva -> kráva
*kok*ot -> kokot*

Odstranění mezer v textu

Program by měl umět pracovat s textem ve dvojí podobě: bez mezer a s mezerami (viz níže – vyhledávání *-oň*, *-out*).

V podobě bez mezer z toho důvodu, že uživatelé obcházejí blokovací automaty pomocí použití mezer: *k r e t é n*, *d e b i l*. Po odstranění těchto mezer program ihned rozpozná nevhodná slova, a příspěvek smaže.

k r e t é n -> kretén
d e b i l -> debil

Doplnění písmene, které bylo nahrazeno jiným znakem

Podobně jako televize nebo rádia „vypívávají“ sprostá slova, tištěná nebo online média výrazy „vyhvězdičkovávají“. Podobné řešení volí i uživatelé diskusních fór, kdy jedno písmeno nahradí jedním interpunkčním znaménkem – *zm.d*, *deb*l*. Rozhodnutí, jestli odstraňovat i takové příspěvky, které si uživatelé sami „cenzurují“ nebo je v diskusích ponechat, je na správci diskusních fór.

Chceme, aby náš program uměl v těchto slovech doplnit chybějící znak.

zm.d -> zmrđ
*deb*l -> debil*

Rozpoznávání překlepů

Uživatelé se v nevhodných výrazech dopouštějí překlepů – ať už nechtěných (*kuvra*), úmyslných (*curva*, *hofno*, *hovno*, *kujva*, *kulva*, *kurfá*, *pyča*, *xicht*) nebo nevědomých (*krypl*, *drška*).

c nahradí písmeno k: curva -> kurva
f nahrazuje v: kurfa -> kurva, hofno -> hovno
j nahrazuje písmeno r: kujva -> kurva
l nahrazuje písmeno r: kulva -> kurva
w nahrazuje v: howno -> hovno
x nahrazuje ks : xicht -> ksicht
y nahrazuje i (a naopak): pyča -> piča, krypl -> kripl
z/ž nahrazuje s/š (a naopak): drška -> držka

Chceme, aby program dokázal pracovat i s tímto nahrazováním písmen (*c* za *k*, *l* za *r* apod.) a uměl takto utvořené výrazy najít.

Vyhledávání zakončení -oň, -out

V předminulé kapitole jsme došli k závěru, že z hlediska slovo tvorby by bylo efektivní, aby program uměl vyhledávat zakončení *-oun* a *-oň*. Nalezená slova pak bude posílat moderátorovi ke schválení.

U zakončení *-out* bude mít program nadefinováno, že bude určitá neproblémová slova ignorovat (*hezoun; mlsoun; meloun; pozoun*).

U zakončení *-oň* nebude program brát v potaz slova *alespoň, aspoň*.

Vyhledávání začátků slov kvazi-, pseudo-, truc-

Podobně jako u zakončení slov bude program vyhledávat slova začínající prefixy *kvazi-* (*kvazipolitik*), *pseudo-* (*pseudopolitik*) a začátek slova *truc-*³³ (*trucstrana*), kde bude ignorovat slova *trucovat* a *truchlit* a všechny jejich podoby.

Výskyty těchto výrazů bude program rovněž posílat ke schválení moderátorovi.

Vyhledávání tykání a vykání

Během získávání našich dat jsme zjistili, že velká řada urážek se vyskytuje tam, kde na sebe uživatelé reagují. Proto jsme se rozhodli, že náš program bude vyhledávat projevy tykání (*ty, tys, tvoje, bys, by jsi, jsi, si, seš*) a vykání (*vy, váš, vaše, jste, ste, byste, by jste, kdybyste, kdyby jste*).

Protože jsme se zakončení slov při tykání a vykání doposud nevěnovali, podíváme se na tento jev trochu podrobněji.

Při tykání, tedy oslovení 2. os. sg., má sloveso specifický tvar: například zakončení *-ej* u rozkazovacího způsobu (*koukej* vypadnout). Takové zakončení má však řada substantiv (*obličej*), zájmen (*jej*), ale i nespisovných výrazů (*slizkej, blbej*). Proto bychom při vyhledávání takového zakončení slov museli program omezit pouze na práci se slovesy.

Při tvoření přítomného času se uplatňuje zakončení *-eš* (1., 2., 3. třída sloves; *bereš, kupuješ*), *-íš* (4. třída; *prosíš*) a *-áš* (5. třída; *uděláš*). Toto zakončení existuje u dalších slov (*veš, náš, guláš, nejspíš*), proto bychom museli i v tomto případě zahrnout do kódu programu takové omezení, které by dovolilo programu pracovat pouze se slovesy.

Rozkazovací způsob u 2. os. pl. se tvoří pomocí zakončení *-te/-ete/-ěte/-ejte*, jednoduše bychom tedy mohli vyhledávat zakončení *-te*, neboť je všem třem společné. Přítomný čas je rovněž tvořen zakončením *-te* (*prosíte, čistíte*).

Podobně jako u předchozích případů bychom museli program naprogramovat tak, aby pracoval pouze se slovesy, protože takové zakončení nalezneme při skloňování vzoru kuře (*bez telete*). Další problém se slovem končícím *-te* nastane u 3. os. sg. některých sloves (*roste, doroste*). Přesto si myslíme, že vyhledávání zakončení *-te* bude užitečné.

Dospěli jsme tedy k názoru, že program bude zasílat moderátorovi k „ručnímu“ zjednoznačení tato slova: *ty, tys, tvoje, bys, by jsi, jsi, si, seš; vy, váš, vaše, jste, ste, byste, by jste, kdybyste, kdyby jste*. Dále pak zakončení slov *-ej, -eš, -áš, -íš* a *-te*. V těchto případech bude program pracovat pouze s jedním slovním druhem: se slovesy.

7. Závěr

Nevhodné příspěvky na internetových fórech jsou problémem, které se snaží diskusní servery řešit. Využívají několika způsobů, které se snaží alespoň z části zabránit problémovým příspěvkům. Patří k nim například povinná registrace uživatele, včetně uvedení e-mailové adresy. Unikátní způsob zavedly například *Novinky.cz*, kde mohou diskutovat pouze lidé, kteří projdou tzv. korespondenční registrací. Uživatelé musí serveru poskytnout své osobní údaje – jméno, příjmení a adresu, kam jim server pošle přístupové údaje do diskusí.

Pokud totiž server přinutí své uživatele, aby místo anonymní identity používali identitu pravou, donutí je k tomu, aby se chovali slušně. To, že je tento způsob účinný, potvrzuje podobný přístup *iDnesu.cz*, kde se místo přezdívek uživatelů začaly používat skutečná jména, tedy jména a příjmení. Na rozdíl od *Novinek.cz* však *iDnes.cz* pravost „skutečných“ jmen nijak neověřuje. Podobný a jednodušší přístup volí i další zpravodajské servery – přechází na diskuse přes *facebook*, kde jsou uživatelé rovněž nuceni vystupovat pod svou pravou identitou.

Registrace nebo vystupování pod pravou identitou nejsou zdaleka jedinými způsoby, jak servery bojují proti problémovým příspěvkům. Na řadu přicházejí moderátoři diskusí, kteří musejí pročítat diskuse a rozhodovat o nezávadnosti příspěvků.

Zpravodajské portály dále využívají například blokovací automaty, které do diskuse nevpustí komentáře obsahující nevhodná slova. Další způsob představují programy, jež dokážou vyhledat nevhodné komentáře, v nichž se žádná urážlivá, vulgární nebo rasistická slova nevyskytují.

A právě o sestavení pravidel pro program, který by byl schopen najít jak běžně používané urážlivé, vulgární nebo rasistické výrazy, tak i výrazy nově utvořené, jsme se v této bakalářské diplomové práci snažili. Tento program vznikne v návaznosti na tuto práci, podle pravidel popsaných v kapitole šesté.

Abychom mohli takový program sestavit, bylo potřeba nejdřív nashromáždit materiál, na němž bychom mohli pozorovat typické rysy, jimiž jsou tyto výrazy tvořeny. K tomuto účelu byl vytvořen abecedně řazený slovník v příloze.

Mimo to se tento slovník snaží zjistit, jestli se jedná o nově utvořené a neznámé pojmy, nebo o výrazy dlouhodobě používané a reflektované v českých slovnících. Tuto skutečnost jsme ověřovali ve čtyřech lexikografických příručkách: *Slovníku spisovného jazyka českého* (SSJČ), *Slovníku spisovné češtiny* (SSČ), *Slovníku nespisovné češtiny* (SNČ) a *Slovníku české frazeologie a idiomatiky* (SČFI).

Dospěli jsme k závěru, že téměř 90 % výrazů alespoň jeden z výše uvedených slovníků zachycoval (většinou *Slovník nespisovné češtiny*, v menší míře pak *Slovník spisovného jazyka českého*). Je ovšem pravdou, že ne vždy slovníky zachycovaly významy, které jsme ve slovníku hledali. Z toho vyplývá, že u některých jednotek se nejedná o nově vzniklé lexémy, ale o nově přidané významy.

Kromě toho jsme zachycovali i charakteristiky slov uvedených v těchto slovnících, abychom zjistili, jak moc jsou slova nevhodná – jestli jsou pouze expresivní, hanlivá, zhrubělá nebo jsou vulgárního charakteru. Nejčastěji se vyskytovaly charakteristiky hanlivé, expresivní a zhrubělé.

O tom, která slova nebudou moderátoři diskusí tolerovat, rozhoduje sám server. Zpravodajské portály mohou dospět k rozhodnutí, že jistá slova hanlivá (např. *budižkničemu*) a zhrubělá (např. *blbec*) budou tolerovat.

Co ale diskuse pravděpodobně tolerovat nebudou, jsou výrazy vulgární a rasistické. Z našeho slovníku v příloze vyplývá, že charakteristiku *vulg.* obsahuje jen malá část slovníku, a protože je velice pravděpodobné, že servery budou právě tyto LJ chtít zakázat, vypisujeme zde kompletní seznam LJ, které byly alespoň v jednom z výše uvedených lexikografických děl charakterizovány jako vulgární:

běhna, bobr, bordel, bréca, brko, buchta, buk, bukál, bukvice, buzerant, bužík, buzna, cecky, coura, cuchta, čurák, čúza, dávačka, drbat, držka, dudy, flundra, fošna, frantík, frnda, fuchtle, grázl, hadice, hajzl, hajzlbába, homokláda, homouš, hovno, huba, chlastoun, chlívák, jebat, jebka, klacek, kláda, klátiti, kokot, koule, kozičky, kozy, kripl, ksindl, kunda, kurevník, kurva, kurvadrát, kurvafix, ližprdelka, mamrd, melouny, mrdat, mrdka, nabíječka, ocas, ochlasta, ojet, (v)ošoust, parchant, péro, piča/piča, pička, pičovina/pičovina, pičulin, pičus, píchat, pind'our, pipina, pizda, posera, posrat, prasák, prcat, prcina/prcinka, prcka, prdel, pták, rajda, rozhod'nožka, rypák, řit'opich, sráč, sračka, (srágora), sviňák, šoustat, šprcguma, šprcka, štětka, šukat, šukna, šulin, teplomet, teplouš, tlustoprđ, trkat, vemena, vlez(do)prdelka, vůl, vyčuránek/vyčuránek, zjeb, zkurvysyn a zmrđ; z hlediska víceslovných spojení: slepičí prdelka, držet hubu, honit si ptáka/ho, jít do prdele, leštit si bambus, nesrat boha, políbit prdel, vysrat si oko

Příznak xenofobní (podle SNČ) měly tyto lexikální jednotky:

bakelit, cigoš, cikorka/cikorky, Čobol, Čobolák, kofola, morgoš, negr, rákosník, režořrout, šikmoň, t'aman, tiň-t'ong, Vietnamčík; z víceslovných spojení pak černá huba

K nim navíc přidáváme další jednotky, které za xenofobní považujeme, ale žádný slovník je neuvádí:

bezbarák, cinkle, hašlerka, hnědočech, hnědoobčan, hnědoslovák, multikulturní, nefakčenko, nemakčenko, nemáklo, nepřízřůsobilý, parazit, přidmoudlík, robič, šikmooký, šikmočko, temnočech, tit'an, ušmudlánek, utiskovaný, Vietkong, vyvolený, vyžírka, znevýhodněný; z víceslovných spojení pak cikán černý, černá opice, černá tlama, kurva šikmooká, opálená menšina, osoba práce se štítící

Ke xenofobním výrazům bychom mohli zařadit i následující slova. Podotýkáme však, že některé z nich slovníky reflektují, někdy jako slova neutrální, jindy jako slova hanlivá apod.

Anglán, Amík, Arabáč, Dojčland'ák, Frantík, Helmut, Japončík, Němčour, Pšonek, Skopčák, Talián

Vedle základního zmapování nevhodných výrazů vyskytujících se na diskusních fórech měl slovník další cíl: snažili jsme se na získaném materiálu sledovat způsoby, jakým nevhodné výrazy vznikaly. Došli jsme k závěru, že největší roli sehrály sémantické a slovtvorné procesy.

K sémantickým přenosům a přesunům docházelo u výrazů, u nichž lze v našem slovníku najít, že lexikografické příručky sice jejich výskyt zaznamenaly, avšak „v jiném významu“. To znamená, že vedle významu, které slovníky zachycují, byl slovu přidán význam nový, většinou nespisovný, expresivní, urážlivý, vulgární nebo xenofobní (*alpinista, buchta, hadice, oranžáda*). Svou roli zde sehrálo hlavně metaforické přenášení významu.

Jelikož program sám o sobě nemyslí a pracuje pouze podle pokynů, které mu zadá jeho tvůrce, nedokáže rozhodnout, v jakém významu se slovo vyskytuje a jestli v nedávné době nezískal význam nový. Tento problém může alespoň z části řešit tím, že bude sledovat, jestli se v poslední době nezačaly v diskusích objevovat málo používaná slova ve větší míře. Pokud ano, pravděpodobně byl slovu přidán nový význam (např. *bakelit, hašlerka, kofola* jako označení Romů). V takových případech by měl program posílat slovo k tzv. ruční desambiguaci, kdy moderátor diskusí rozhodne, zda je slovo závadné.

Další problém představují vícevýznamová slova, která mají vedle svého neutrálního, spisovného významu ještě další urážlivý, vulgární nebo rasistický význam (*briketa, bukvice, kráva, ropucha, židle*). I v těchto případech bude ležet největší váha na ruční posteditaci administrátora diskusí. Dalším řešením může být to, že program bude sledovat kontext, ve kterém se slovo vyskytuje, a pokud se blízko vícevýznamového slova budou vyskytovat další předem daná slova, program je může vyhodnotit jako nevhodná a smazat je (např. *kráva* ve spojení *blbá, pitomá; pták* ve spojení *s honit, leštit*).

Z hlediska slovtvorných procesů jsme dospěli k tomu, že by program mohl vyhledávat neznámá slova, která obsahují zakončení *-oun* a *-oň*. Tato zakončení se často objevovala u expresivně zabarvených slov, jejichž cílem bylo urážet člověka. Ale ani tato pravidla nejsou vždy stoprocentně úspěšná: takové zakončení obsahují i neutrální nebo obsahově nezávadná slova (*alespoň, aspoň, hezoun, mlsoun*). I tyto výskyty bude muset program posílat k ručnímu zjednoznačnění moderátorovi.

Vyhledávání takovýchto zakončení s sebou nese i další problémy: ne vždy se slova budou vyskytovat v reprezentativním tvaru (tj. v nominativu sg. nebo pl., v případě sloves v infinitivu), ale v různých tvarech (skloňování, časování). Počítačové programy však dokážou tento problém vyřešit: k nalezenému slovu (*tupoune*) umí najít základní tvar (*tupoun*).

Dále jsme zjistili, že by bylo vhodné, aby program pracoval se slovy začínajícími pomocí *kvazi-*, *pseudo-* a *truc-* (ignorace slov *trucovat*, *truchlit*), neboť nimi byly tvořeny urážlivé příspěvky. I v tomto případě bude program posílat příspěvky ke schválení.

Zohledňování dalších sufixů a prefixů není účelné, neboť jsou jimi tvořena neutrální slova. Příznaková slova jsou vesměs tvořena neutrálními slovotvornými prostředky, a tudíž není možné vyhledávat tato příznaková slova na základě formálních kritérií (zakončení), neboť by to vedlo k masivnímu přegenerování a neefektivnosti programu.

Kromě těchto jazykových poznatků se dají nevhodné komentáře vyhledávat i podle jiných jevů: například negativního hodnocení komentářů jinými uživateli, podle výskytu negativních emotikonů (smajlíků) u příspěvků (např. palec dolů), sledování editační (Levensteinovy) vzdálenosti, vyhledávání slov psaných kapitálkami (*KRÁVO*), přebytečných interpunkčních znamének (*bl. . . .bka*), zmnožení písmen (*krááááva*), velkého počtu mezer, kterými uživatelé provazují blokovaná slova (*k r e t é n*), výskytu interpunkčních znamének ve slově (*zm.d*).

Nevhodné příspěvky se často vyskytují tam, kde na sebe uživatelé reagují. Tyto reakce jsou doprovázeny buď tykáním, nebo vykáním. Rozhodli jsme se, že program bude vyhledávat takovéto projevy tykání a vykání, pro něž jsou charakteristická určitá zakončení (*-ej*, *-eš*, *-íš*, *-áš*, *-te*). U vyhledávání těchto zakončení dostane program navíc instrukce, že bude pracovat pouze se slovesy. Kromě těchto zakončení pojících se s tykáním a vykáním bude program zohledňovat výskyt dalších slov, které jsou spjaty s tykáním a vykáním (*ty*, *tys*, *tvoje*, *bys*, *by jsi*, *jsi*, *si*, *seš*; *vy*, *váš*, *vaše*, *jste*, *ste*, *byste*, *by jste*, *kdybyste*, *kdyby jste*). Všechny tyto výskyty ve vstupním textu pak opět pošle moderátorovi k desambiguaci.

Program bude mazat pouze takové příspěvky, které obsahují slova shodující se s databází zakázaných slov. Z výše uvedených poznatků vyplývá, že hlavní váha posouzení, zda se jedná o příspěvek vhodný či nevhodný, bude spočívat na moderátorovi diskuse, který celou situaci posoudí. Program, který by o posouzení všech výrazů rozhodoval sám, vytvořit nepůjde. Přesto do velké míry pomůže moderátorovi k rychlému odhalení nevhodných příspěvků a schválení těch problémových.

8. Seznam použité literatury

ČECHOVÁ, Marie, Marie KRČMOVÁ a Eva MINÁŘOVÁ. *Současná stylistika*. Vyd. 1. Praha: Lidové noviny, 2008, 381 s. ISBN 9788071069614.

FILIPEC, Josef a František ČERMÁK. *Česká lexikologie*. Vyd. 1. Praha: Academia, 1985, 281 s.

HLADKÁ, Zdeňka a Olga MARTINCOVÁ. *Slova v soukromých dopisech: Lexikografická sonda*. Brno: Masarykova univerzita, 2012.

JANDOVÁ, Eva. *Čeština na www chatu*. Vyd. 1. Ostrava: Ostravská univerzita, Filozofická fakulta, 2006, 262 s. ISBN 8073682532.

LAPÁČEK, Jiří. *Internet pro úplné začátečníky*. Vyd. 1. Praha: Computer Press, 2000, x, 198 s. ISBN 8072262262.

NEKULA, Marek. *Stereotypy v lexikalizovaných etnických označeních*. In: A. Macurová - I. Nebeská (vyd.) *Jazyk a jeho užívání*. Praha: Filozofická fakulta UK, 1996. ISBN 80-85899-19-1.

RANGELOVÁ, Alben. *Neosémantizmy a neosémantizační procesy*. In: MARTINCOVÁ, Olga. *Neologizmy v dnešní češtině*. Praha: ÚJČ AV ČR, 2005, 248 s. ISBN 808649621x.

ŠMAHEL, David. *Psychologie a internet: děti dospělými, dospělí dětmi*. Praha: Triton, 2003, 158 s. ISBN 8072543601.

Slovníky

ČERMÁK, František. *Slovník české frazeologie a idiomatiky*. 2. přeprac. a dopl. vyd., 1. vyd. v nakl. Leda. Praha: Leda, 2009, 507 s. ISBN 97880733521651.

ČERMÁK, František. *Slovník české frazeologie a idiomatiky*. 2. přeprac. a dopl. vyd., 1. vyd. v nakl. Leda. Praha: Leda, 2009, 547 s. ISBN 97880733521722.

ČERMÁK, František. *Slovník české frazeologie a idiomatiky*. 2. přeprac. a dopl. vyd., 1. vyd. v nakl. Leda. Praha: Leda, 2009, 1247 s. ISBN 97880733521893.

ČERMÁK, František. *Slovník české frazeologie a idiomatiky*. 2. přeprac. a dopl. vyd., 1. vyd. v nakl. Leda. Praha: Leda, 2009, 1267 s. ISBN 97880733521964.

FILIPEC, Josef. *Slovník spisovné češtiny pro školu a veřejnost: s Dodatkem Ministerstva školství, mládeže a tělovýchovy České republiky*. 4. vyd. Praha: Academia, 2007, 647 s. ISBN 9788020013477.

HUGO, Jan. *Slovník nespisovné češtiny: argot, slangy a obecná mluva od nejstarších dob po současnost: historie a původ slov*. 3. rozš. vyd. Praha: Maxdorf, 2009, 501 s. ISBN 9788073451981.

HAVRÁNEK, Bohuslav a Lubomír DOLEŽEL. *Slovník spisovného jazyka českého*. 1. vyd. Praha: Academia, 1971, xxviii, 1311 s.

HAVRÁNEK, Bohuslav. *Slovník spisovného jazyka českého*. Vyd. 1. Praha: Academia, 1971, 1191 s.

HAVRÁNEK, Bohuslav. *Slovník spisovného jazyka českého*. 1. vyd. Praha: Academia, 1971, 1079 s.

HAVRÁNEK, Bohuslav. *Slovník spisovného jazyka českého*. 1. vyd. Praha: Academia, 1971, 1011, xx s.

Diplomové práce

GAŽAROVÁ, Irena. Vybrané etické problémy administrace online diskusních fór [online]. Brno, 2010 [cit. 2013-02-25]. Dostupné z: http://is.muni.cz/th/103073/fss_b/Vybrane_eticke_problemy_administrace_online_diskusnich_for.pdf. Bakalářská diplomová práce. Masarykova univerzita.

MALIŠOVÁ, Jana. Slovtvorná expresivita [online]. Brno, 2006 [cit. 2013-03-31]. Dostupné z: http://is.muni.cz/th/17005/ff_m/Slovtvorna_expresivita_04-11.pdf. Magisterská diplomová práce. Masarykova univerzita, Filozofická fakulta, Ústav českého jazyka. Vedoucí práce Zdenka Rusínová.

ŠABATOVÁ, Marta. Příznaková slovní zásoba v soukromé korespondenci mládeže [online]. Brno, 2007 [cit. 2013-04-13]. Dostupné z: http://is.muni.cz/th/44110/ff_m/Text_M_Sabatova.pdf. Magisterská diplomová práce. Masarykova univerzita. Vedoucí práce Zdeňka Hladká

Elektronické zdroje

AOL Advertising [online]. AOL INC., © 2013 [cit. 2013-02-26]. Dostupné z: <http://advertising.aol.com/brands/huffington-post>

Arianna Huffington. Encyclopedia Britannica [online]. Encyclopædia Britannica, ©2013 [cit. 2013-02-26]. Dostupné z: The Huffington Post. Encyclopedia Britannica [online]. Encyclopædia Britannica, ©2013 [cit. 2013-02-26]. Dostupné z: <http://www.britannica.com/EBchecked/topic/1192975/The-Huffington-Post>

Borgis, a.s. Firmy.cz [online]. Seznam.cz, © 1996–2013 [cit. 2013-02-25]. Dostupné z: <http://firmy.cz/detail/151339-borgis-praha-vinohrady.html>

Český národní korpus – SYN2010. Ústav Českého národního korpusu FF UK, Praha 2010. Dostupný z: <http://www.korpus.cz>.

ČÍŽEK, Jakub. Jak se diskutuje na českých zpravodajských webech. In: Živě.cz [online]. Mladá fronta a.s., 2012 [cit. 2013-02-26]. Dostupné z: <http://www.zive.cz/clanky/jak-se-diskutuje-na-ceskych-zpravodajskych-webech/sc-3-a-166144>

Databáze heslářů: Slovník spisovného jazyka českého. In: Ústav pro jazyk český Akademie věd ČR, v. v. i. [online]. @2005-2013 [cit. 2013-04-13]. Dostupné z: <http://lexiko.ujc.cas.cz/heslare/dictionary.php>

DOČEKAL, Daniel. IDnes je prý jednička mezi zpravodajskými servery. Skutečně?. Lupa.cz [online]. 2013 [cit. 2013-04-22]. Dostupné z: <http://www.lupa.cz/clanky/idnes-je-pry-jednicka-mezi-zpravodajskymi-servery-skutecne/?labelsBox>

-labelId=1524&,do=labelsBox-switch

DOČEKAL, Daniel. Na Facebooku je 83.09 milionu falešných účtů. Spočítal to sám Facebook. Lupa.cz [online]. roč. 2012 [cit. 2013-03-20]. Dostupné z: <http://www.lupa.cz/clanky/na-facebooku-je-mezi-5-az-27-falesnych-uctu-zalezi-na-tom-koho-se-zeptate/>

FAQ. In: NetMonitor.cz [online]. SPIR, 2011 [cit. 2013-03-31]. Dostupné z: <http://www.netmonitor.cz/FAQ>

FIALA, Luděk. Diskuse na Novinkách čeká zásadní změna. Novinky.cz [online]. 2009 [cit. 2013-02-25]. Dostupné z: <http://www.novinky.cz/domaci/160046-diskuse-na-novinkach-ceka-vyrazna-zmena.html>

Hlavní strana Novinky.cz. Novinky.cz [online]. 2013 [cit. 2013-02-23]. Dostupné z: <http://www.novinky.cz/vasenazory>

Frekvence písmen, bigramů, trigramů, délka slov. In: Centrum zpracování přirozeného jazyka [online]. 2012 [cit. 2013-03-28]. Dostupné z: http://nlp.fi.muni.cz/cs/Frekvence_pismen_bigramu_trigramu_delka_slov

Huffingtonpost.com. Quancast [online]. Quantcast Corporatio, ©2013 [cit. 2013-02-26]. Dostupné z: <http://www.quantcast.com/huffingtonpost.com>

Internet. Oxford Dictionaries [online]. Oxford University Press, 2010 [cit. 2013-02-25]. Dostupné z: <http://oxforddictionaries.com/definition/english/Internet?q=internet>

JELÍNEK, Milan. První díl Slovníku spisovného jazyka českého. Naše řeč. 1961, roč. 44, 1-2. Dostupné z: <http://nase-rec.ujc.cas.cz/archiv.php?art=4807>

KNELL, Matthew. Talking Community with Justin Isaf, Huffington Post Community Manager. In: Aol blog [online]. AOL Inc, © 2013 [cit. 2013-02-26]. Dostupné z: <http://blog.aol.com/2012/03/01/talking-community-with-justin-isaf-huffington-post-community-ma/>

KŘÍSTEK, Václav. Slovník spisovného jazyka českého. Naše řeč. 1971, roč. 54, č. 4. Dostupné z: <http://nase-rec.ujc.cas.cz/archiv.php?art=5579>

KUČERA, Karel. Slovník české frazeologie a idiomatiky. Naše řeč. 1995, roč. 78, č. 2. Dostupné z: <http://nase-rec.ujc.cas.cz/archiv.php?art=7236>

LEINER, Barry M., Vinton G. CERF, David D. CLARK a Robert E. KAHN. The Past and the Future History of the Internet. Communications of the ACM [online]. 1997 [cit. 2013-02-25]. Dostupné z: <http://ccrg.soe.ucsc.edu/CMPE252A/FALL2012/PAPERS/history1.pdf>

Levenshtein distance: (definition). In: National Institute of Standards and Technology: U.S. Department of Commerce [online]. [cit. 2013-04-17]. Dostupné z: <http://xlinux.nist.gov/dads//HTML/Levenshtein.html>

Měsíční zpráva - Leden 2013. NetMonitor.cz [online]. SPIR z. s. p. o, 2012 [cit. 2013-02-25]. Dostupné z: http://www.netmonitor.cz/sites/default/files/vvnetmon/2013_01_netmonitor_offline_report.pdf

Měsíční zpráva - Leden 2012. NetMonitor.cz [online]. SPIR z. s. p. o, 2012 [cit. 2013-02-25]. Dostupné z: http://www.netmonitor.cz/sites/default/files/vvnetmon/2012_01_netmonitor_offline_report.pdf

Měsíční zpráva - Leden 2010. NetMonitor.cz [online]. SPIR z. s. p. o, 2012 [cit. 2013-02-25]. Dostupné z: http://www.netmonitor.cz/sites/default/files/vvnetmon/2010_01_netmonitor_offline_report.pdf

Měsíční zpráva - Leden 2008. NetMonitor.cz [online]. SPIR z. s. p. o, 2012 [cit. 2013-02-25]. Dostupné z: http://www.netmonitor.cz/sites/default/files/vvnetmon/2008_01_netmonitor_offline_report_2.pdf

Měsíční zpráva - Leden 2006. NetMonitor.cz [online]. SPIR z. s. p. o, 2012 [cit. 2013-02-25]. Dostupné z: http://www.netmonitor.cz/sites/default/files/vvnetmon/2006_01_netmonitor_offline_report.pdf

O firmě: Rok 2003. Seznam.cz [online]. [cit. 2013-02-25]. Dostupné z: <http://onas.seznam.cz/cz/o-firme/historie-firmy/2003/>

SONDERMAN, Jeff. How the Huffington Post handles 70+ million comments a year. In: Poynter.org [online]. The Poynter Institute, 2012 [cit. 2013-02-26]. Dostupné z: <http://www.poynter.org/latest-news/top-stories/190492/how-the-huffington-post-handles-70-million-comments-a-year/>

STRUHÁRIK, Filip. Huffington Post: Vieme spravovať tisíce príspevkov za hodinu. In: Smeblog [online]. Petit Press, a.s., 2012 [cit. 2013-02-26]. Dostupné z: <http://struharik.blog.sme.sk/c/313047/Huffington-Post-Vieme-spravovat-tisice-prispevkov-za-hodinu.html>

The Huffington Post. Encyclopedia Britannica [online]. Encyclopædia Britannica, ©2013 [cit. 2013-02-26]. Dostupné z: <http://www.britannica.com/EBchecked/topic/1192975/The-Huffington-Post>

Změny v diskotování – otázky a odpovědi. Idnes.cz [online]. MAFRA a. s., 2012 [cit. 2013-02-26]. Dostupné z: <http://redakcni.blog.idnes.cz/c/294046/Zmeny-v-diskutovani-otazky-a-odpovedi.html>