



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH
TECHNOLOGIÍ

ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF TELECOMMUNICATIONS

SYSTÉM PRO AUTOMATICKÉ VYHODNOCENÍ E-MAILOVÝCH ZPRÁV

SYSTEM FOR AUTOMATIC EVALUATION OF EMAIL MESSAGES

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. JAN FRKAL

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. MARTIN KOUTNÝ, Ph.D.

BRNO 2014



**VYSOKÉ UČENÍ
TECHNICKÉ V BRNĚ**

**Fakulta elektrotechniky
a komunikačních technologií**

Ústav telekomunikací

Diplomová práce

magisterský navazující studijní obor
Telekomunikační a informační technika

Student: Bc. Jan Frkal

ID: 106438

Ročník: 2

Akademický rok: 2013/2014

NÁZEV TÉMATU:

Systém pro automatické vyhodnocení e-mailových zpráv

POKYNY PRO VYPRACOVÁNÍ:

Účelem práce je realizace systému na automatické vyhodnocení a analýzu emailových zpráv. Systém bude realizován v prostředí Internet. Systém bude umožňovat automatizované načtení emailovaných zpráv ze zvolených POP3/IMAP účtů, jejich vyhodnocení a třídění do zvolených kategorií. Takto realizovaný poštovní klient bude umět automaticky rozeznávat reklamní emaily, které bude třídit do kategorií podle typu zprávy, odesílatele a informace. Systém bude mít víceúrovňovou uživatelskou správu (bude umožňovat registraci účtů a přidělování práv podúčtům), pokročilou statistiku a další funkce související s provozem aplikace.

DOPORUČENÁ LITERATURA:

[1] LECKY-THOMPSON, Ed a Steven D NOWICKI. PHP 6: programujeme profesionálně. Vyd. 1. Překlad Ondřej Gibl. Brno: Computer Press, 2010, 718 s. Programujeme profesionálně. ISBN 978-80-251-3127-5.

Termín zadání: 10.2.2014

Termín odevzdání: 28.5.2014

Vedoucí práce: Ing. Martin Koutný, Ph.D.

Konzultanti diplomové práce:

prof. Ing. Kamil Vrba, CSc.

Předseda oborové rady

UPOZORNĚNÍ:

Autor diplomové práce nesmí při vytváření diplomové práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

ANOTACE

Tato diplomová práce se zabývá návrhem a realizací systému pro automatické vyhodnocení e-mailových zpráv. Systém spolupracuje s jazykem PHP a databází MySQL. Současně umožňuje automatickou synchronizaci. Při synchronizaci se stáhnou a uloží e-mailové zprávy za pomoci protokolu IMAP nebo POP3.

Následně jsou zprávy analyzovány. Během analýzy jsou tyto zprávy rozřazeny do druhů, a to podle předdefinovaných klíčových slov. Systém dále spolupracuje se seznamy černé a bílé listiny. Je-li odesílatel zprávy nalezen při synchronizaci na černé listině, takový e-mail se přeskočí. Naopak při nalezení odesílatele na bílé listině je u takového e-mailu přeskočeno hledání shod klíčových slov a z listiny je načten nastavený druh a kategorie.

Většinu hodnot z provedeného vyhodnocení e-mailů lze přehledně sledovat v pokročilé statistice. K dispozici jsou koláčové grafy a číselná statistika. Vstup do systému je chráněn přihlašovacími údaji. Přihlásit se tedy mohou jen registrovaní uživatelé.

KLÍČOVÁ SLOVA

PHP, MySQL, e-mail, IMAP, POP3, automatické, vyhodnocení, analýza, statistika

ABSTRACT

This diploma thesis deals with the design and realisation of system for automatic evaluation of e-mail messages. The system works with PHP language and MySQL database. It also allows automatic synchronisation. During the synchronisation e-mail messages are downloaded and saved using IMAP or POP3 protocol.

Subsequently, the messages are analysed. During the analysis the reports are classified into types, according to pre-defined keywords. The system also works with black lists and white lists. If the sender of an e-mail is during the synchronisation found in the blacklist, that e-mail will be skipped. On the contrary, if the sender of an e-mail is found within the white list, that e-mail will be excluded from keyword matching and from the list is type and category loaded.

Most of the values from the carried out evaluation of the e-mails can be clearly seen in advanced statistics. Pie charts and numerical statistics are available. Access to the system is protected by a login. Therefore, login can only registered users.

KEYWORDS

PHP, MySQL, e-mail, IMAP, POP3, automatic, evaluation, analysis, statistics

FRKAL, J. *Systém pro automatické vyhodnocení e-mailových zpráv*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2014. 64 s. Vedoucí diplomové práce Ing. Martin Koutný, Ph.D.

Prohlašuji, že svou diplomovou práci na téma „Systém pro automatické vyhodnocení e-mailových zpráv“ jsem vypracoval samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené diplomové práce dále prohlašuji, že v souvislosti s vytvořením této diplomové práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědom následků porušení ustanovení S 11 a následujících autorského zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

Brno

.....
(podpis autora)

Na tomto místě bych chtěl poděkovat vedoucímu diplomové práce Ing. Martinovi Koutnému, Ph.D. za užitečnou metodickou pomoc a rady při zpracování práce. Také bych rád poděkoval své rodině a přítelkyni, kteří mě po celou dobu podporovali.

Výzkum popsaný v této diplomové práci byl realizován v laboratořích podpořených z projektu SIX; registrační číslo CZ.1.05/2.1.00/03.0072, operační program Výzkum a vývoj pro inovace.

OBSAH

Úvod.....	12
1 Technologie elektronické pošty	13
1.1 Agenti pro elektronickou poštu.....	13
1.1.1 MTA.....	13
1.1.2 MUA	13
1.1.3 MDA	13
1.1.4 Komunikace při doručování pošty	13
1.2 Program Postfix	14
1.2.1 Vlastnosti postfixu	14
1.3 DNS	15
1.4 Protokoly pro odesílání pošty	17
1.4.1 Protokol SMTP	17
1.4.2 Příkazy protokolu SMTP	17
1.4.3 Ukázka SMTP komunikace	18
1.4.4 Protokol LMTP	19
1.5 E-mailová zpráva	19
1.5.1 Formát e-mailové adresy	20
1.6 Standard MIME	20
1.6.1 Hlavičky standardu MIME	21
1.7 Nevyžádaná pošta	23
1.7.1 Obrana před nevyžádanou poštou dle umístění filtru	23
1.7.2 Obrana před nevyžádanou poštou dle typu filtru	24
1.7.3 Jak získají rozesílatelé spamu cílové adresy	25
1.7.4 Snahy o obejití antispamových filtrů	26
1.7.5 Nalezení původního odesílatele zprávy	27
1.8 Reklamní zpravodaj (newsletter)	27
1.8.1 Reklamní zpravodaj v elektronické podobě.....	28
1.9 Běžná pošta	29
1.9.1 Nedoručitelná zpráva	29
1.10 Přístup k e-mailovým zprávám	29
1.10.1 Přístup přes webové rozhraní.....	29
1.10.2 Přístup pomocí poštovního klienta	29
1.10.3 Srovnání protokolů IMAP a POP3	31
2 Vývojové systémy.....	32
2.1 Jazyk PHP	32
2.1.1 Srovnání server vs. klient.....	33

2.1.2	PHPMailer	34
2.2	Jazyk MySQL	34
2.2.1	Funkce systému MySQL	34
2.2.2	Základní příkazy MySQL	35
2.2.3	Výhody a nevýhody jazyka MySQL	36
2.2.4	Typy MySQL úložišť	36
3	Realizace systému pro vyhodnocení zpráv	37
3.1	Návrh systému	37
3.2	Popis tabulek MySQL databáze	38
3.3	Vstup do systému	38
3.3.1	Bezpečnostní prvky systému	38
3.4	Popis aplikace	39
3.5	Druh přijaté pošty	40
3.6	Přehled pošty	40
3.6.1	Poštovní účty	40
3.6.2	Uložené e-maily	41
3.7	Synchronizace e-mailů	42
3.7.1	Rozdíly synchronizace mezi účty IMAP a POP3	43
3.8	Automatické vyhodnocení	43
3.8.1	Výsledky provedené analýzy	44
3.8.2	Nová analýza	45
3.8.3	Spam skóre	45
3.8.4	Automatická kategorizace	46
3.8.5	Ruční přidání klíčového slova z e-mailu	46
3.8.6	Filtr pošty	47
3.9	Nastavení analýzy	47
3.9.1	Klíčová slova	47
3.9.2	Kategorie	48
3.9.3	Černá listina	48
3.9.4	Bílá listina	48
3.10	Poštovní účty	49
3.10.1	Šablony účtů	50
3.11	Předvolby	50
3.11.1	Víceúrovňová uživatelská správa	50
3.12	Pokročilá statistika	52
3.12.1	Grafy Google Charts	52
3.12.2	Grafy zobrazované ve statistice	53

4	Závěr	54
	Seznam literatury	55
	Seznam použitých zkratek	58
	Seznam příloh	59

SEZNAM OBRÁZKŮ

Obrázek 1: Komunikace při doručování pošty.....	13
Obrázek 2: MX záznamy pro doménu Seznam.cz z online nástroje [37].....	16
Obrázek 3: Ukázka reklamního zpravodaje	28
Obrázek 4: E-mail odeslaný Mailer Daemonem, že schránka neexistuje.....	29
Obrázek 5: Výsledek synchronizace účtu	42
Obrázek 6: Nalezená shoda klíčových slov v těle e-mailu	44
Obrázek 7: Analýzou vyhodnocené e-maily	44
Obrázek 8: Rozlišení druhu pošty dle vypočteného skóre	45
Obrázek 9: Výpis uložených poštovních účtů.....	49
Obrázek 10: Oprávnění uživatele s počtem spravovaných uživatelů.....	51
Obrázek 11: Ukázka Google Charts grafu v aplikaci.....	52

SEZNAM TABULEK

Tabulka 1: Číselné odpovědi SMTP serveru na příkazy.....	18
Tabulka 2: Význam hlaviček elektronické pošty	19
Tabulka 3: Popis různých typů kódování MIME	22
Tabulka 4: Vysvětlení spam hlaviček	25
Tabulka 5: Správné významy překlepů spamera	26
Tabulka 6: Čísla portů u známých protokolů pro příjem pošty	30
Tabulka 7: Vybrané příkazy protokolu IMAP	31
Tabulka 8: Srovnání IMAP a POP3	31
Tabulka 9: Popis tabulek aplikace v systému MySQL	38
Tabulka 10: Popis souborů aplikace dle umístění.....	39

ÚVOD

Elektronická pošta je jednou z nejstarších služeb nabízených na internetu. Jedná se o oblíbenou, levnou a především rychlou službu. V poslední době jde také o jeden z nejpobulárnějších způsobů komunikace po síti internetu.

Cílem této diplomové práce je vytvořený systém pro automatické vyhodnocení e-mailových zpráv v prostředí internetu. Pro návrh takového systému byl zvolen jazyk PHP a databázový systém MySQL. Přístup do systému je chráněn jménem a heslem a registrace uživatelů je volně přístupná.

Nejdříve musí být definovány a v databázi MySQL uloženy účty IMAP nebo POP3 pro připojení k e-mailové schránce. Dále je k dispozici synchronizace. Při synchronizaci se z připravených účtů stahují nové zprávy. Poté jsou tyto zprávy filtrovány a uloží se jen ty, jejichž odesílatel nebo doména není na černé listině. Tyto uložené zprávy jsou připraveny k vyhodnocení.

Pro vyhodnocení zpráv se využívá analýza. S pomocí této analýzy lze rozřídít e-mailové zprávy podle nalezených shod klíčových slov. Několik desítek klíčových slov je již v databázi k dispozici, a to v několika různých jazycích. Každé klíčové slovo je specifikováno vahou a pravděpodobnou kategorií. Během analýzy se sčítají váhy nalezených klíčových slov. Po dokončení se uživateli zobrazí rozpoznáný druh pošty dle výsledné váhy a nejpravděpodobnější kategorie. Tyto hodnoty se následně uloží do databáze.

Z uložených hodnot se sestavuje pokročilá statistika. V té jsou k dispozici čtyři grafy a dále statistika zpráv, účtů, klíčových slov a kategorií. Systém dále nabízí víceúrovňovou uživatelskou správu, při které jsou uživatelé rozlišeni dle oprávnění na oprávněné a základní uživatele. Oprávnění mohou ovládat základní uživatele a zároveň mají tito základní uživatelé v aplikaci jistá omezení.

1 TECHNOLOGIE ELEKTRONICKÉ POŠTY

1.1 Agenti pro elektronickou poštu

1.1.1 MTA

Mail Transfer Agent je program, který zajišťuje doručení pošty z jednoho účtu na jiný. Uživatel většinou pracuje právě s MUA. MTA se stará o odesílání a příjem. Rozdělení do jednotlivých uživatelských schránek má pak na starost MDA. Některá MTA umí však tuto funkci zajistit sama. Příkladem MTA jsou programy Postfix, Exim nebo Qmail. [2]

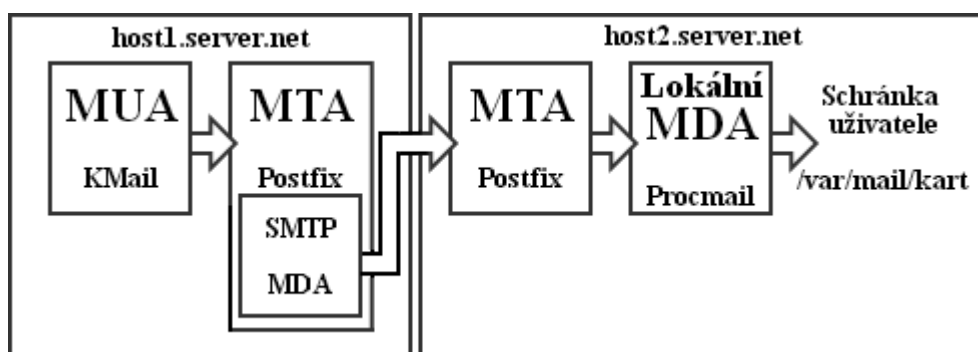
1.1.2 MUA

Mail User Agent je program vhodný pro čtení e-mailů (tzv. e-mailový klient). Takový MUA často spolupracuje s MTA či MDA. Většina MUA dokáže komunikovat přímo s poštovními servery, bez MTA. K nejznámějším linuxovým MUA se řadí KMail, Evolution a Mutt. [3]

1.1.3 MDA

Hlavním úkolem Mail Delivery Agenta je doručení e-mailů do schránky příjemce po přijetí zprávy na serveru. Také dokáže e-maily automaticky zpracovávat (odpovídat, uložit přílohy). MDA je často kombinovaný s MTA, lze však také využít MDA bez MTA. Dále vhodné k ukládání zpráv do schránek příjemců použít program, místo přímého ukládání zpráv. Při doručení lze poté zprávy dále filtrovat. Filtr pak může automaticky mazat nevyžádanou poštu (spam a viry). Typickým příkladem mohou být programy Procmail nebo Maildrop. [1, 4]

1.1.4 Komunikace při doručování pošty



Obrázek 1: Komunikace při doručování pošty

Z výše uvedeného obrázku je patrné, že MUA (KMail) na serveru *host1* odesílá zprávu do MTA (Postfix). Toto MTA zaznamená, že je zpráva určena uživateli na serveru *host2*. Zpráva má být doručena přes SMTP, takže ji MTA nejdříve předá SMTP MDA.

SMTP MDA se připojí k MTA serveru *host2* (Postfix) a zprávu zašle tomuto agentovi. Toto MTA zjistí, že je zpráva určena pro uživatele v místní síti, takže ji předá lokálnímu MDA. MDA uloží zprávu do schránky uživatele *kart*.

Jakmile se uživatel přihlásí k serveru *host2* a spustí své MUA, čeká na něj ve schránce zpráva připravená k přečtení. [1]

1.2 Program Postfix

Postfix je jedním z nejvyužívanějších programů z řady MTA. Vznikl již v roce 1998 jako open-source a vychází z programu Sendmail. Jedná se o bezpečný a rychlý SMTP¹ server. Autorem tohoto programu je Wietse Venema.

Program Postfix se vyznačuje jednoduchou instalací a rychlou konfigurací. Změny konfigurace se provádějí snadno úpravou dobře komentovaných konfiguračních souborů. Postfix je tvořen různými programy a podsystémy. Tento přístup zajišťuje vysokou flexibilitu. Díky těmto vlastnostem je tedy tento program vhodný pro splnění požadavku na snadnost obsluhy a instalace.

1.2.1 Vlastnosti postfixu

- **Spolehlivost**

Postfix dokáže pracovat spolehlivě i při vysoké zátěži. Mnoho jiných programů se může chovat nepředvídatelně, jakmile jim dojde volné místo nebo paměť. Postfix takové stavy předvídá a nabízí možnost vzpamatovat se.

- **Výkon**

Postfix je navržen pro dosažení vysokého výkonu. Zároveň limituje počet nových procesů, které je nutno vytvořit, tak i počet přístupů k systému souborů využívaných při zpracovávání zpráv.

¹ Simple Mail Transfer Protocol

- **Bezpečnost**

K obraně proti útočníkům zavádí Postfix několik vrstev. Zavádí bezpečnostní princip nejmenších oprávnění. Tak běží každý proces, který může běžet odděleně, s nejnižší sadou oprávnění, která skutečně potřebuje. Procesy s vyššími oprávněními nikdy nedůvěřují neprivilegovaným procesům.

Dále lze deaktivovat nepotřebné moduly. To vede ke zvýšení zabezpečení a zjednodušení instalace Postfixu.

- **Jednoduché používání**

Postfix je svým nastavením a správou jeden z nejjednodušších balíčků pro elektronickou poštu. A to protože pracuje s jednoduchými vyhledávacími tabulkami. Tyto tabulky zajišťují překlad adres a předávání zpráv. Během rozhodování se ve fázi návrhu autor přikláněl na tu stranu, která by se většinou zdála jako nejrozsudnější.

1.3 DNS

DNS² je rozsáhlá distribuovaná databáze. Jejím hlavním úkolem je mapování názvů hostitelů na IP³ adresy. Využity jsou zde porty 53 u protokolů TCP⁴, rovněž i u UDP⁵.

Pro lepší orientaci se běžně používají jména domén, adresy pro stroje jsou pak vyjadřovány následovně:

- pomocí 32bitových adres (v případě protokolu IPv4) jako A záznamy,
- pomocí 128bitových adres (v případě protokolu IPv6) jako AAAA záznamy.

Data se skládají ze záznamů, které se označují jako záznamy prostředků. Tyto různé záznamy prostředků nabízí různé druhy informací - např. adresy IP, směrování pošty nebo názvové servery či aliasy názvů hostitelů.

² Domain Name Server

³ Internet Protocol

⁴ Transmission Control Protocol

⁵ User Datagram Protocol

Známé jsou tyto záznamy prostředků:

- **A záznamy**

O mapování názvů na IP adresy starají záznamy A. Tyto obsahují název hostitele a IP adresu. Názvy, které se běžně používají na internetu pro odkazy na hostitele, je třeba převést na IP adresy pro směrování. Záznamy A umožňují tedy překlad názvu na IP adresu.

- **MX záznamy**

Tyto záznamy vycházejí z angl. Mail eXchangers a jsou tedy důležité pro směrování e-mailů v síti. Obsahují servery pro vyřizování pošty pro dané domény, tedy názvy poštovních serverů. Tato servery zajistí všechnu poštu pro daný název domény. Mluvíme-li o komunikaci s MTA, záznamy MX sdělují, kam se mají zprávy odeslat. A záznamy se jsou platné pro protokol IPv4, u protokolu IPv6 se objevuje alternativa v podobě AAAA záznamů.

Vybraná doména může mít také více poštovních serverů. Proto mají záznamy MX preferenční hodnotu k určení priority při volbě poštovního serveru, kam mají být zprávy doručovány. Platí, že čím nižší číslo, tím vyšší priorita, přičemž se začíná vždy od nejnižší priority (tj. nejvyššího čísla).

Pref	Hostname	IP Address	TTL
10	mx1.seznam.cz	77.75.72.42	5 min
20	mx2.seznam.cz	77.75.76.32	5 min

Obrázek 2: MX záznamy pro doménu Seznam.cz z online nástroje [37]

- **CNAME záznamy**

U některých názvů hostitelů se mohou vyskytnout aliasy, které ukazují na jiné názvy hostitelů (místo IP adres). Toto může být užitečné při směrování požadavků na služby (např. POP nebo HTTP). CNAME poskytuje kanonický (skutečný) název, na který alias názvu hostitele poukazuje. Např. název hostitele `www.firma.cz` může být ve skutečnosti CNAME záznam pro `server1.firma.cz`.

- **PTR záznamy**

U těchto záznamů je umožněno reverzní vyhledávání, tj. převod IP adres na názvy hostitelů. PTR záznamy běžně odpovídají A záznamům tak, aby dopředné vyhledání názvů hostitelů vrátilo IP adresu, jejíž reverzní vyhledání vrací stejný název hostitele.

Některé aplikace používají PTR záznamy k ověření, zdali je IP adresa připojujícího se klienta mapována na očekávaný název hostitele. [5]

1.4 Protokoly pro odesílání pošty

1.4.1 Protokol SMTP

Simple Mail Transfer Protocol je protokol vycházející z TCP protokolu. Jedná se o textovou komunikaci standardně využívající port 25. SMTP doručuje poštu přes přímé spojení mezi odesílatelem a adresátem. Běžně se používá v kombinaci s protokoly POP3 a IMAP⁶ pro uložení zpráv ze serveru.

1.4.2 Příkazy protokolu SMTP

Komunikace je založena na principu dotazů a odpovědí. Přičemž je vždy nutné čekat na odpověď aktuálního dotazu, teprve až poté lze vyslat další příkaz. Ke každému dotazu musí přijít právě jedna odpověď. Tyto odpovědi jsou charakterizovány trojmístným číslem (viz Tabulka 1).

Ukázka některých SMTP příkazů:

- **EHLO** - extended HeLO (využíváno u ESMTP serverů), ověření funkčnosti serveru na protistraně,
- **MAIL** - začátek přenosu e-mailové zprávy,
- **RCPT** - zadání adresy příjemce (příp. i jména), pro více příjemců se použije příkaz opakovaně,
- **DATA** - obsah vč. volitelného předmětu zprávy, ukončeno tečkou,
- **QUIT** - ukončení SMTP spojení.

⁶ Internet Message Access Protocol

Tabulka 1: Číselné odpovědi SMTP serveru na příkazy

Číslo	Význam	Popis odpovědi
1xx	Průběh	Přijata žádost, která bude dále zpracována. Čeká se na další příkazy od klienta.
2xx	Úspěch	Příkaz byl přijat a potvrzen.
3xx	Částečný úspěch	K dokončení je třeba dalšího příkazu (často se vyskytuje při odpovědi na příkaz DATA).
4xx	Dočasná chyba	Chyba je dočasná a po opakování může nastat úspěch (2xx).
5xx	Trvalá chyba	Chyba je trvalá a lze ji odstranit změnou předchozích příkazů.

1.4.3 Ukázka SMTP komunikace

Pro ukázkou komunikace byl použit program Telnet⁷ na portu 25:

```
telnet localhost 25
S: 220 localhost ESMTP Postfix (Ubuntu)
K: EHLO wifi-km.net
S: 250 localhost
S: 250 (výpis akceptovatelných příkazů)

K: MAIL FROM: <test@gmail.com>
S: 250 2.1.0 Ok
K: RCPT TO: <xfrkal02@stud.feec.vutbr.cz>
S: 250 2.1.5 Ok
K: DATA
S: 354 END DATA WITH <CR><LF>.<CR><LF>
K: Odeslání zkusební pošty do schránky.
K: .
S: 250 2.0.0 Ok: queued as 1KART023224
K: QUIT
S: 221 2.0.0 Bye
```

Význam zkratk a příkazů v komunikaci:

- K, S - označení pro příkaz od klienta a odpověď od serveru,
- EHLO - pozdrav klienta (ESMTP⁸ protokol; u SMTP HELO, u LMTP LHLO),
- MAIL FROM - e-mailová adresa odesílatele,
- RCPT TO - e-mailová adresa příjemce,
- DATA - sekvence těla zprávy (tělu může předcházet příkaz Subj : pro předmět) zakončená tečkou (.),
- QUIT - ukončení SMTP komunikace. [6]

⁷ Telecommunication Network

⁸ Extended Simple Mail Transfer Protocol

1.4.4 Protokol LMTP

Protokol LMTP⁹ vychází ze SMTP a je jeho zjednodušenou verzí. LMTP server nezařazuje zprávy do fronty a nezkouší opětovně doručit zprávy, které nemohou být doručeny ihned. Jestliže MTA doručuje e-mail na server SMTP a je tento e-mail určen pro více příjemců, z nichž některý nemůže e-mail přijmout, přebírá SMTP odpovědnost za doručení později. Zároveň dojde k hlášení úspěšného doručení odesílajícímu MTA.

Vzhledem k tomu, že LMTP servery nezařazují zprávy do fronty, musí tyto servery vrátit odpověď o stavu doručení všem příjemcům konkrétní zprávy. Těm příjemcům, kterým nemohlo být e-mail doručen, přebere odpovědnost za zařazení do fronty a pozdější pokus o doručení odesílající MTA, nikoliv LMTP server.

Výhodou LMTP je spolehlivost a vysoká škálovatelnost. Se zvyšující se zátěží, lze snadno přidat více prvků do LMTP systémů v popředí i na pozadí. [5]

1.5 E-mailová zpráva

Každá e-mailová zpráva má dvě důležité části – hlavičky, které jsou povinné a tělo, které je volitelné.

Tabulka 2: Význam hlaviček elektronické pošty

Položka hlavičky	Význam
To	E-mailová adresa příjemce
Cc	E-mailové adresy lidí, kteří dostanou kopii zprávy (z angl. Carbon Copy)
Bcc	E-mailové adresy lidí, kteří dostanou kopii zprávy, ale ostatní příjemci se o tom nedovědí (z angl. Blind Carbon Copy; slepá kopie)
From	E-mailová adresa odesílatele
Subject	Předmět zprávy
Date	Datum a čas odeslání
Received	Seznam serverů, přes které e-mail prošel. Zde je možné zjistit jméno serveru, doba, kdy server zprávu zpracovával, unikátní ID zprávy a komu je zpráva určena. Každý server opíše tento seznam a přidá svůj záznam. Tento údaj umožňuje určit, odkud zpráva putovala.
Message-ID	Unikátní ID zprávy, které např. umožní řazení zpráv do vláken.
Reply-To	Komu přijde odpověď, jestliže příjemce na zprávu odpoví.
Content-Type	Zprávy, které mají přílohu, obsahují informaci o typu přílohy (zvuk, video, archiv atp.) a také řetězec, podle kterého se určí začátek a konec přílohy (tzv. boundary).

⁹ Local Mail Transfer Protocol

Hlavičky začínající X* je možné ignorovat. Používají se např. pro informace o druhu e-mailového programu nebo rozesílacího seznamu apod. [7, 12]

1.5.1 Formát e-mailové adresy

E-mailová adresa se v STMP komunikaci vyskytuje v následujícím formátu:

místní_část oddělovač název_domény

- místní část - uživatelské jméno nebo pseudonym; levá strana,
- oddělovač - dle specifikace RFC5322 je oddělovačem znak zavináče (@),
- název domény - pravá strana.

Příkladem správné e-mailové adresy je: *petr.kratochvil@google.com* [13]

1.6 Standard MIME

MIME¹⁰ je důležitou součástí elektronické pošty. Umožňuje totiž psát zprávy také v jiných znakových sadách ASCII¹¹. Text e-mailové zprávy bez MIME striktně používá sedmibitovou znakovou sadu, která však zabraňuje použití všech znaků češtiny. Internetový standard MIME je popsán v dokumentu RFC 2045 a jeho principem je víceúčelovost.

Formát elektronických zpráv MIME tedy umožňuje:

- psaní textu zpráv v jiných znakových sadách než ASCII,
- podporu zpráv tvořených z více částí,
- podporu jiných znakových sad než ASCII v hlavičkách zpráv,
- použití netextových formátů v těle zpráv.

V elektronické poště je stále více využívána také novější verze – standard S/MIME (Secure/MIME). Tato verze kryptograficky zabezpečuje e-mailové zprávy, především integritu a nepopíratelnost původu pomocí elektronického podpisu. [36]

¹⁰ Multipurpose Internet Mail Extensions

¹¹ American Standard Code for Information Interchange

1.6.1 Hlavičky standardu MIME

Standard MIME zavádí do e-mailových zpráv tyto hlavičky:

- **MIME-Version**

Tato hlavička specifikuje verzi MIME normy a hlavním důvodem jejího zavedení je zajištění zpětné kompatibility. Toto je jediná hlavička nezačínající řetězcem „Content“ a znamená, že je zpráva sestavena podle doporučení RFC2045 až RFC2049. Verze rozšíření MIME je podle těchto doporučení 1.0 (`Mime-Version: 1.0`). V budoucnu jsou očekávány nové verze standardu MIME.

- **Content-Type**

Touto hlavičkou se popisuje typ dat v těle zprávy. To je důležité pro přijímajícího klienta, který tuto zprávu obdrží, aby mohl vhodně prezentovat obsah zprávy. Charakter obsahu zprávy je zde specifikován pomocí typu a podtypu. Případně mohou být zadány také doplňkové informace, které se umísťují za středník ve tvaru *parametr=hodnota*. Parametrů může být více (opět oddělené středníkem), přičemž nezáleží na pořadí.

Typ určuje, o jaký typ dat se jedná (ve zprávě je např. text, obrázek nebo binární soubor) a podtyp pak konkrétní formát textu nebo obrázku apod.

Do typů popisující přenášená data lze zařadit: *text*, *application*, *image*, *audio* a *video*. Mezi typy poukazující na složení zprávy z více částí pak patří *message* nebo *multipart*.

Ukázka použití hlavičky:

```
Content-Type: text/plain; charset=utf-8
```

- **Content-Transfer-Encoding**

Data posílaná e-mailem jsou často 8bitová nebo binární. Taková data však nelze zpravidla poslat přímo a je třeba je převádět. Použitý typ kódování je uveden v této hlavičce. Nejčastější algoritmy kódování jsou: 7bit, quoted-printable, base64, 8bit a binary.

7bitová zpráva je vhodná pro e-mail a žádné kódování tedy neprobíhá. Jedná se o implicitní metodu kódování, která se předpokládá, není-li tato hlavička uvedena vůbec. Také u hodnot 8bit a binary se ve skutečnosti žádné kódování neprovede. Tyto hodnoty jsou tedy užitečné jako indikace typu dat v objektu. Jednoduché srovnání několika výše uvedených typů kódování zobrazuje Tabulka 3.

Tabulka 3: Popis různých typů kódování MIME

Typ kódování	Popis
7bit	Data jsou v krátkých řádcích, obsahuje jen US-ASCII.
8bit	Data jsou v krátkých řádcích, vyskytují se i znaky jiné než ASCII.
binary	Data jsou v dlouhých řádcích, vyskytují se i znaky jiné než ASCII.

Rozdíl mezi 8bitovým a binárním kódováním je takový, že u 8bitu každých 8 bitů reprezentuje pouze jeden znak a u binary vystupuje spojitý tok bitů nerepresentující žádné znaky (např. spustitelný program). V případě binary musí být délka zprávy násobkem 8 bitů.

Hlavička se obvykle vztahuje k celému tělu zprávy. Pokud se ale objeví v konkrétní části zprávy, pak se vztahuje jen na tuto část.

Příklad použití hlavičky:

Content-Transfer-Encoding: base64

▪ **Content-ID**

V modernějších klientech lze vytvořit odkaz z jedné zprávy do druhé. Tělo zprávy je tedy možné označit identifikátorem v hlavičce Content-ID. Hodnota této hlavičky pak takovou zprávu jednoznačně identifikuje. Tato hlavička je volitelná, použití je ale povinné v implementaci, která generuje data typu message/external-body.

▪ **Content-Description**

Hlavička obsahuje informace popisující přenášenou právu, např. název obrázku, který je posílán jako tělo zprávy. Popis musí být vždy ve znakové sadě US-ASCII.

▪ **jiné hlavičky (např. Content-Length a Content-Disposition) [28]**

1.7 Nevyžádaná pošta

Tento druh pošty bývá často také nazýván jako spam. Jedná se o nevyžádanou reklamní poštu nebo jiné sdělení (často komerčního charakteru) rozesílané hromadně více příjemcům.

Jednoznačně lze za spam považovat takové zprávy odeslané adresátům, kteří si tyto zprávy neobjednali a především nemají možnost zrušit odběr zpráv. Nevyžádané zprávy často přicházejí ze zahraničních serverů a jazykem zpráv je často angličtina.

Zprávy také často přicházejí stejným příjemcům opakovaně. U každé další zprávy je její obsah velmi podobný až stejný. Jejich zdrojový kód je však upraven tak, aby se zpráva co nejvíc vyhnula antispamovým filtrům.

Běžní příjemci tak u dalších zpráv nepoznají žádnou změnu a zpráva se jím jeví jako stejná. Nezachytí-li antispamové filtry takovou zprávu, bude přijata do složky doručené pošty.

Ze strany serveru se lze spamu vyhnout následovně:

- pomocí nástroje SpamAssasin (viz níže),
- kontrolou e-mailu proti databázi nevyžádané pošty,
- zablokováním příchozích zpráv ze serverů typu open-relay (přes takové servery může odesílat zprávy kdokoli; dnes jich existuje už jen minimum).

Nevyžádaná pošta může být označována také jako UBE¹² nebo UCE¹³. Opakem spamu je pak „ham“ – vyžádaná pošta. [4, 8]

Nejpoužívanějším obranným nástrojem před nevyžádanou poštou jsou antispamové filtry. Úkolem těchto filtrů je rozpoznání spamových zpráv a jejich oddělení od běžné pošty. [9]

1.7.1 Obrana před nevyžádanou poštou dle umístění filtru

S nevyžádanou poštou lze bojovat s ohledem na umístění filtru na dvou místech, a to na serveru nebo na straně klienta.

¹² Unsolicited Bulk E-mail

¹³ Unsolicited Commercial E-mail

- **Filtr na serveru**

Nejideálnější místo pro umístění filtru je právě poštovní server. Snaží se včas odchytit všechnu nevyžádanou poštu. Tím šetří objem stahovaných dat a především čas uživatelů, kteří by spam museli filtrovat ručně sami.

Tento filtr přináší ale také nevýhodu. Tou je skutečnost, že vzhledem k velkému množství filtrovaných e-mailů dojde občas ke smazání žádané pošty.

- **Filtr u klienta**

Tento filtr je součástí poštovního klienta nebo je externí průchozí. Zde může uživatel vždy vidět smazané zprávy (přesunuté do složky spam) a tyto zprávy jako spam označit, v případě že jsou chtěné.

1.7.2 Obrana před nevyžádanou poštou dle typu filtru

- **SpamAssassin (slovníkový filtr)**

SpamAssassin je nástroj napsaný v jazyce Perl. Slouží k filtrování nevyžádané pošty na základě obsahu e-mailu. SpamAssassin je někdy nazýván jako slovníkový filtr a sestává z jednoduchých pravidel. Nevýhodou je relativně nízká spolehlivost.

Princip tohoto filtru spočívá v udělení bodového hodnocení každé zprávě v závislosti na tom, kolik prvků (běžně se u spamu vyskytujících) zpráva obsahuje. Tento počet bodů je následně porovnán s prahovou hodnotou, a pokud jej zpráva přesáhne, je označena jako spam.

Prahovou hodnotu filtru lze měnit dle potřeby. Pokud ji snížíme, začne procházet více nevyžádané pošty. Jakmile ji zvýšíme, může být běžná pošta chybně vyhodnocena jako spam. [5, 29, 32]

- **Bayesův filtr**

Tento filtr se dá také označit jako učení spamového filtru. Bayesův filtr zakládá na označování zpráv jako Spam nebo Ham. Tak označené zprávy se dále analyzují rozebráním na jednotlivé příznaky. Každý příznak má poté přiřazenu pravděpodobnost, s jakou může nevyžádaný e-mail obsahovat daný příznak.

Sečtením těchto pravděpodobností lze získat výslednou pravděpodobnost. Tak se zjistí přítomnost spamu. Jde však pouze o upřesňující informaci. Hodnota Bayesova filtru se запиše do hlavičky *X-Spam-Status* e-mailové zprávy. Bayesův filtr se jeví jako nejefektivnější metoda.

Ukázka hlaviček při detekci nevyžádané pošty:

```
X-Spam-Status: Yes, hits=16.48 tagged_above=2.5
required=6.3
tests=[BAYES_99=3.5, HELO_DYNAMIC_IPADDR=4.4,
HTML_30_40=0.01,
HTML_MESSAGE=0.01, MIME_HTML_ONLY=0.17,
RCVD_IN_BL_SPAMCOP_NET=1.26,
RCVD_IN_CBL=2.2, RCVD_IN_SORBS_DUL=1.97, RCVD_IN_XBL=3.06]
X-Spam-Level: *****
X-Spam-Flag: YES
```

Tabulka 4: Vysvětlení spam hlaviček

Název spam hlavičky	Vysvětlení
X-Spam-Status	Dosažený počet bodů (hits) a požadovaný počet bodů (required) k označení zprávy jako spam.
X-Spam-Level	Vyjádření míry úrovně nevyžádané pošty.
X-Spam-Flag	Závěrečná informace o tom, je-li zpráva spam (Yes/No).

[14, 21]

1.7.3 Jak získají rozesílatelé spamu cílové adresy

Existuje několik zdrojů, jak mohou rozesílatelé spamu (tedy spameři) získat adresy obětí nevyžádané pošty:

- z internetových stránek – cíleně vyvinuté aplikace prochází webové stránky a hledají na nich e-mailové adresy,
- ze soutěžních a různých formulářů – např. u soutěží slibujících výhru odpovězením na jednu nebo více jednoduchých otázek,
- z formulářů na vyřazení adresy reklamní pošty – především u zahraničních nabídek se nedoporučuje vyplňovat e-mail k vyřazení z databáze (takto si spamer ověří aktuálnost adresy a ví, že může rozeslat další spam),

- uhodnutím (metodou pokus-omyl) – náhodná kombinace několika znaků a čísel nebo využití slovníkového útoku k uhádnutí e-mailových adres (na kratší adresy přichází více nevyžádané pošty),
- zakoupením již existující databáze e-mailů. [31]

1.7.4 Snahy o obejití antispamových filtrů

V této kapitole je rozebráno, jak se rozesílatelé spamu snaží obelstít filtry tak, aby jejich nevyžádané zprávy pronikly do běžné pošty. Účelem umístění reklamního textu do zprávy je znemožnění nebo alespoň ztížení odhalení klíčových slov ve zprávě antispamovými filtry.

Snahou je tedy zobrazit reklamní text lidskému čtenáři zprávy a současně zneviditelnit nebo zpřeházet dotýčný text pro antispamový filtr. Takový filtr se snaží odhalovat negativně hodnocená klíčová slova.

▪ Překlepy a přesmyčky ve filtrovaných slovech

Nejjednodušším způsobem, jak ztížit odhalení klíčového slova, je záměrná chyba v zápisu. Může se jednat o překlep, chybný pravopis nebo prohození znaků ve slově. Člověk takovou chybu pochopí, ale stroj (filtr) je zmaten. V praxi již ale některé moderní filtry dokáží takové překlepy odhalit.

Tabulka 5: Správné významy překlepů spamera

Slovo spamera s překlepem	Správný význam
Pharkmncy	Pharmacy
Offir	Offer
Ctrl	Control
Nvestment	Investment

▪ Využití foneticky podobných číslic

Další možností přepisu textu jiným textem (dávajícím pro čtenáře smysl), je nahrazování některých slov číslicemi, které jsou stejně vyslovovány. Principem je shoda výslovnosti některých číslic s výslovností různých slov nebo jejich částí.

Nejobvyklejší je použití číslovky „4“ (four) ve významu „pro“ (for) a také číslovky „2“ (two) ve významu „k“ nebo „do“ (to).

Někdy je využívána i výslovnost jednotlivých hlásek, např. „B with U“ znamená „být s vámi/s tebou“. Velmi často se lze setkat s vyjádřením „4U“, tedy „pro vás/pro tebe“ (for you), případně „B4“ (before), „gr8“ (great) anebo „2C“ (to see).

V českém jazyce se takový princip příliš nevyužívá, výjimkou je „o5, z5“ (opět, zpět).

▪ **Využití graficky podobných číslic**

Jednou z metod, jak vnitřně upravit text, aby se uživateli zdál nezměněný, je grafická podobnost některých znaků. Konkrétně malé „L“ lze nahradit číslicí „jedna“, která vypadá téměř identicky:

„jedna“: 1,

„malé L“: 1.

Obdobně je tomu i s číslicí „nula“ podobnou písmenu „velké o“. Případně (ale již méně často) lze písmeno „S“ nahradit číslicí „5“. Dále např. v cyrilici může být písmeno „Z“ zastoupeno číslicí „3“ a v latince ruské „R“ použito místo „P“. Pro čtenáře je to opět totéž, nicméně pro počítač nikoliv. [9]

1.7.5 Nalezení původního odesílatele zprávy

Nejjednodušší cestou jak zjistit původního odesílatele je nahlédnutí do hlavičky *X-Originating-IP*. Tato hlavička je důležitá tím, že poskytuje IP adresu počítače, který e-mail odeslal. Bohužel se tato hlavička v poslední době již nevyskytuje příliš často, ale odesílatele lze zjistit z hlavičky *Received*. [30]

1.8 Reklamní zpravodaj (newsletter)

Newsletter je reklamní či informační zpravodaj. Cílem takového zpravodaje je informovat příjemce o novinkách, událostech a akcích. Zpravodaj může být zasíláný v elektronické nebo tištěné podobě:

- **elektronická** – distribuce po internetu (na e-mail); populární forma,
- **tištěná** – akční letáky, noviny; distribuci zajišťuje roznáška nebo pošta.

1.8.1 Reklamní zpravodaj v elektronické podobě

Elektronický zpravodaj může být rozeslán v grafické podobě v HTML¹⁴ formátu nebo formou obrazového letáku v PDF¹⁵. Rozhodujícím faktorem u rozesílaných zpravodajů může být grafická podoba či poutavost obsažených informací. Newsletter je často zasílán pravidelně přihlášeným odběratelům a zdarma. [34]

Seriózní firma by měla navíc do každého svého e-mailu zahrnout také stručný návod k odhlášení z newsletteru. Běžně se nabízí tyto varianty pro odhlášení:

- odpověď na e-mail se specifickým předmětem či tělem zprávy,
- odkaz vedoucí uživatele na stránku odesílatele k odhlášení.



Obrázek 3: Ukázka reklamního zpravodaje

Výhody reklamního zpravodaje:

- nízké náklady ve srovnání s tištěným zpravodajem a na provoz,
- cílené oslovení zákazníků, kteří jsou vedeni k objednání služeb či výrobků,
- firma získává zpětnou vazbu pomocí odkazů ve zprávě,
- newsletter připomíná existenci firmy.

Nevýhody reklamního zpravodaje:

- při nedodržení pravidel pro odeslání newsletteru nemusí být zpravodaj vůbec doručen (skončí např. ve spam složce).

Dle zákona smí být newsletter rozesílán pouze přihlášením čtenářům, tedy těm, kteří se zasíláním souhlasili. V opačném případě se jedná o zasílání nevyžádané pošty. [15, 16]

¹⁴ HyperText Markup Language

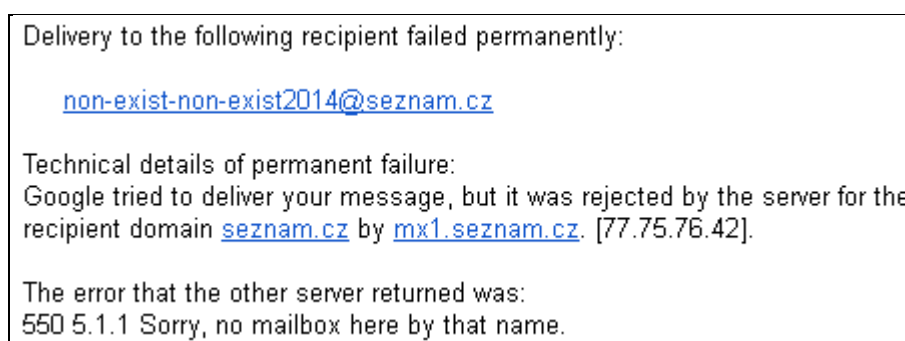
¹⁵ Portable Document Format

1.9 Běžná pošta

Za běžnou poštu se tedy označují zbylé zprávy – tedy ty, které nejsou ani nevyžádanou ani reklamní poštou. Sem patří také zprávy osobního charakteru.

1.9.1 Nedoručitelná zpráva

Někdy se může stát, že odeslaná zpráva příjemci nedorazí. Důvodem může být neexistující uživatel, chybná adresa nebo např. plná schránka. V takovém případě dostane odesílatel zprávu o nedoručení. Tu většinou rozesílá Mailer Daemon a jazykem bývá často angličtina. Náhled takové zprávy ukazuje Obrázek 4. Chyba na posledním řádku nastiňuje, že server neneviduje schránku s uvedeným názvem.



Obrázek 4: E-mail odeslaný Mailer Daemonem, že schránka neexistuje

1.10 Přístup k e-mailovým zprávám

Přístupovat ke zprávám a číst je lze pomocí webového rozhraní nebo poštovního klienta.

1.10.1 Přístup přes webové rozhraní

Pomocí webového rozhraní umožňuje většina e-mailových serverů přístup k e-mailu přes internetový prohlížeč uživatele. Výhodou je přístup odkudkoliv a kdykoliv, také není třeba schránku nijak konfigurovat. Nevýhodou může být pomalejší rozhraní a nižší komfort oproti poštovnímu klientovi.

1.10.2 Přístup pomocí poštovního klienta

Takového klienta představuje aplikace na počítači (MUA). Tato aplikace se připojuje k poštovnímu serveru protokolem IMAP nebo POP3. Výhodou je zde rychlost a komfort obsluhy.

Poštovní klient je vhodný pro uživatele, kteří přistupují k poště často. Navíc se nabízí možnost stažení pošty i z více schránek současně. Tabulka 6 zobrazuje čísla portů u protokolů IMAP a POP3 pro různá připojení. [29]

Tabulka 6: Čísla portů u známých protokolů pro příjem pošty

Druh připojení/Protokol	IMAP	POP3
Nezabezpečené	143	110
Šifrované SSL ¹⁶	993	995

▪ **Protokol POP3**

POP3 neboli Post Office Protocol je vhodný pro jednoduché a rychlé stažení e-mailových zpráv ze serveru. Tento protokol je vhodný při omezeném přístupu k síti nebo absenci trvalého připojení. Za takových podmínek dojde k připojení serveru, načtení všech zpráv a následnému odpojení.

Nevýhodou představuje ta skutečnost, že je třeba stahovat všechny zprávy. Tedy není možné stáhnout jen předměty zpráv a podle těch stáhnout odpovídající zprávy. Stažení všech zpráv může být nadbytečné a zdlouhavé. Další nevýhodou může také představovat fakt, že po přenesení zprávy do klienta se tato zpráva ze serveru smaže.

▪ **Protokol IMAP**

Protokol IMAP řeší několik omezení POP3 a umožňuje práci se zprávami přímo na serveru, kde jsou uloženy všechny zprávy. Tyto zprávy však není nutné stahovat všechny. Dále dokáže IMAP pracovat nejen v online, ale i offline režimu. Poslední uvolněná verze protokolu je IMAP4.

Výhodou je synchronizace, kdy IMAP stahuje pouze záhlaví zpráv a obsah se stáhne teprve, až když chce klient zprávu opravdu otevřít. Dále nabízí větší flexibilitu než protokol POP3. Navíc může být připojeno více uživatelů zároveň a u každé zprávy se ukládá také její stav (např. nepřečtená nebo zodpovězená).

Příkazy IMAP jsou zadávány pomocí textových řetězců. Seznam použitelných příkazů je mnohem rozsáhlejší, než u protokolu POP3. Výčet některých takových příkazů ukazuje Tabulka 7. [5, 11, 33]

¹⁶ Secure Sockets Layer

Tabulka 7: Vybrané příkazy protokolu IMAP

Název příkazu	Popis
LOGIN, AUTHENTICATE	Přihlášení uživatele
CAPABILITIES	Zjištění možností serveru
CREATE, DELETE, RENAME, LIST	Práce se složkami
COPY, FETCH, STORE, CLOSE	Práce se zprávami
SEARCH	Vyhledání zpráv
LOGOUT	Odhlášení, ukončení spojení

Mezi dostupné odpovědi protokolu IMAP lze zařadit:

- **OK** (úspěšné dokončení nebo doplňující údaje),
- **NO** (označená odpověď značí neúspěch, neoznačená varovné hlášení),
- **BAD** (označená odpověď značí nesprávný příkaz, neoznačená vážný problém),
- **PRAUTH** (netřeba zadávat LOGIN; vždy je součástí neoznačené odpovědi),
- **BYE** (ukončení relace serverem; vždy u neoznačené odpovědi).

1.10.3 Srovnání protokolů IMAP a POP3

Na základě popisu výhod a nevýhod protokolů IMAP a POP3 v předchozích odstavcích lze stanovit následující srovnání:

Tabulka 8: Srovnání IMAP a POP3

Činnost	IMAP	POP3
Uložení zpráv na poštovním serveru	Ano	Ne
Přístup k archivu zpráv a složek odkudkoliv (z domu, z práce)	Ano	Ne
Dostupnost zpráv a složek v případě havárie počítače	Ano	Ne
Synchronizace složek na serveru a v poštovních programech	Ano	Ne

U POP3 bylo uvažováno použití, při kterém se zprávy po stažení na serveru odstraní. [34]

Z uvedené tabulky tedy vyplývá doporučení využívat protokol IMAP ve srovnání s POP3, je-li to možné.

2 VÝVOJOVÉ SYSTÉMY

2.1 Jazyk PHP

Stránky, které v současné době využívají dynamického zpracování, se neobejdou bez komunikace se serverem. Výsledek je pak odeslán zpátky klientovi. K hlavním skriptovacím jazykům patří např. PHP, ASP ¹⁷ a další.

PHP (Hypertext preprocessor, původně Personal Home Page) je skriptovací jazyk vytvořený právě pro web a umístěný na straně serveru. Syntaxe PHP vychází z jazyka C a také se od ní mnoho neliší. Jazyk PHP je volně šířen. PHP příkazy lze jednoduše vkládat do jazyka HTML, avšak daný soubor musí mít koncovku *.php, případně *.phtml.

Každý příkaz musí být ukončen středníkem (;) a úsek s PHP kódem ohraničen některými z následujících tagů:

`<?php ?>` nebo starším `<? ?>`.

K vytváření PHP kódu postačí běžný textový editor. Pro PHP lze využít server Apache, který je vůbec nejvýkonnější. Apache často běží pod systémem Linux nebo také Windows. PHP pochází z roku 1994, kdy byl vytvořen jako jednoduchý projekt k vylepšení webových stránek. Dnes je již PHP velmi komplexním programovacím nástrojem, a to především díky autorovi Rasmusu Lerdorfovi.

PHP je vyvíjeno jako open source produkt a přístup k jeho zdrojovým kódům je tedy volný. PHP není závislé na platformě (Linux, Unix, Windows) a má rozsáhlé možnosti užití. Je jedním z nejrozšířenějších programovacích nástrojů k vývoji webových aplikací. Nejnovější verze jazyka PHP je 5.6.

Přednostmi jazyka PHP jsou:

- velký výkon,
- nízké náklady na provoz,
- jednoduchý jazyk k naučení,
- snadné použití a velké množství knihoven,
- rozhraní pro mnoho druhů databází,
- zdrojový kód jazyka k dispozici na webu.

¹⁷ Active Server Pages

Nutno uvést také nevýhody tohoto jazyka:

- názvy funkcí nejsou konzistentní, např. `imap_fetchbody()` s podtržítkem oproti `strtotime()` bez podtržítka,
- nejednotné pořadí parametrů u funkcí,
- potenciálně nebezpečné funkce (především u starších verzí PHP). [22, 23, 27]

2.1.1 Srovnání server vs. klient

V této kapitole je porovnáno vykonání skriptu na straně serveru a klienta. Vykonání na serveru lze popsat následujícími body:

- Nenáročné na software nebo hardware klienta – výstupem je běžně HTML stránka, kterou dokáže interpretovat i starší prohlížeč. Ze strany klienta tedy není třeba, aby prováděl skripty nebo podporoval zásuvné moduly atd.
- Snadná interakce s dalšími aplikacemi na serveru, jednodušeji se zajistí bezpečnost celého systému.
- Výrazně vyšší ochrana zdrojových kódů – díky interpretaci kódu na serveru se klient nedostane ke zdrojovým kódům.
- Menší objem přenesených dat – skripty na serveru, které se vkládají do kódu, se ještě na samotném serveru interpretují a ve výstupu se již neobjeví, zatímco skripty určené klientovi ano.

Oproti tomu má provádění kódu u klienta tyto výhody:

- Možnost interakce s objekty webové stránky. U klienta jsou možné reakce na stisk kláves, klepnutí myši nebo pohyb.
- Menší zátěž serveru. Server přijme požadavek a odešle stránku – zatímco v případě serverových technologií musí ještě spustit interpret jazyka, tomu předat stránku a klientovi se zasílá až výsledek. Jazyk PHP může běžet i jako modul webového serveru. Tím se sníží doba reakce a paměťové nároky, takže vliv interpretace PHP kódu na rychlost odezvy serveru je minimální. [23]

2.1.2 PHPMailer

PHPMailer je program k rozesílání elektronické pošty od firmy Worxware. Umožňuje také odeslání e-mailů s přílohou a poskytuje širokou škálu nastavení. Tento program je vytvořený právě v jazyce PHP jako třída, která je volně šiřitelná. Pomocí tohoto programu lze odesílat hromadné e-maily.

2.2 Jazyk MySQL

Databázový systém MySQL se vyznačuje mnoha užitečnými funkcemi:

- MySQL je interaktivní jazyk. Uživatelé zadají příkazy pomocí konverzačního programu a získaná data zobrazí ve výstupu programu.
- MySQL je administrační databázový jazyk. Administrátor využívá jazyk SQL k nadefinování databázové struktury a k řízení přístupu k uloženým datům.
- MySQL je administrační databázový programovací jazyk, pomocí kterého mohou programátoři zahrnout příkazy do svých programů a získat tak přístup k datům v databázi.
- MySQL je distribuovaný databázový jazyk. Systémům správy distribuovaných databází pomáhá MySQL při rozmísťování dat mezi mnoha počítačovými systémy.
- MySQL je jazyk aplikací typu klient/server. Programy na osobních počítačích používají jazyk MySQL pro komunikaci s databázovými servery umístěnými na síti. Tato klient/server architektura je velmi populární.
- MySQL je výhodný pro databázové brány. V síti tvořenou směsicí různých databázových produktů, je MySQL často použito jako brána umožňující jednomu systému komunikovat s jiným.

Velmi intuitivní správu MySQL databáze poskytují nástroje pracující na webovém rozhraní. Takovým nástrojem je např. PhpMyAdmin nebo při vývoji tohoto diplomového projektu využitý český projekt Adminer.

2.2.1 Popis systému MySQL

Jazyk PHP si tedy velmi dobře rozumí s databázemi. MySQL je relační databáze typu DBMS¹⁸ a již podle názvu lze usoudit, že vychází z programovacího jazyka SQL¹⁹.

¹⁸ Database Management System

¹⁹ Structured Query Language

Rovněž jako PHP je i MySQL open source projekt, veřejnosti přístupný od roku 1996. Projekt je standardně distribuován zdarma.

MySQL je velmi rychlý a robustní relační systém. Nabízí možnost efektivně vyhledávat, třídit a získávat data. Stará se o to, aby mohlo do ní přistupovat a být současně připojeno více uživatelů zároveň.

Práce s databází je mnohem rychlejší a pohodlnější než se soubory. Každá databáze v MySQL může obsahovat jednu či více tabulek. Sloupce takových tabulek jsou pak označovány jako atributy a řádky jako záznamy.

Abychom mohli s tabulkami a daty v nich uloženými pracovat, musí být tyto jednoznačně identifikovány. Každý sloupec má proto svůj název, stejně tak jako každá tabulka. [19, 23]

2.2.2 Základní příkazy MySQL

Jako nejzákladnější příkazy lze uvést `mysql_query()`, `mysql_fetch_array()`, `mysql_fetch_row()` a `mysql_num_rows()`. Příkaz `mysql_query()` zavolá do databáze dotaz. Příkazy `mysql_fetch_row()` a `mysql_fetch_array()` vrací hodnoty na právě odeslaný dotaz. Pomocí `mysql_num_rows()` lze jednoduše spočítat výsledky vrácené systémem MySQL.

Výhodný příkaz `mysql_query()` se používá v kombinaci s těmito varianty:

- `SELECT` – pro výběr hodnot z tabulky,
- `SELECT COUNT()` – ke zjištění počtu vrácených výsledků,
- `UPDATE` – pro aktualizaci hodnot v tabulce,
- `INSERT` – k uložení nových řádků do tabulky,
- `DELETE` – pro smazání řádků z tabulky.

V SQL dotazech lze dále také využít klauzulí *WHERE*, *GROUP BY* a *ORDER BY*. Tyto parametry zajistí omezení výběru hodnot z databáze a jejich seřazení. Počet vypisovaných řádků odpovědi lze omezit parametrem *LIMIT*.

2.2.3 Výhody a nevýhody jazyka MySQL

Jazyk SQL se tedy jeví jako výkonný a užitečný nástroj. A to jak pro propojení počítačových programů a systémů s daty uloženými v relační databázi, tak i k propojování lidí.

Mezi výhody jazyka SQL lze tedy uvést:

- rychlost,
- stabilita,
- uživatelská podpora,
- podpora přístupu z mnoha programovacích jazyků (PHP, Java, Perl, C a další).

A do nevýhod tohoto jazyka zařadit:

- nedostatečný výkon ve velice náročných aplikacích,
- složitější programátorské konstrukce nejsou podporovány. [17, 24]

2.2.4 Typy MySQL úložišť

Úložiště jsou také nedílnou součástí MySQL databáze. Do MySQL verze 5 patří:

- **MyISAM** – nejpoužívanější typ tabulek v MySQL. V MySQL databázích je zařazen od verze 3.23. Tyto tabulky nepodporují cizí klíče, výhodou je ale rychlost.
- **MRG_MYISAM** – nadstavba MyISAM sloužící k rozložení zátěže rozdělením dat do více identických tabulek. Tyto tabulky lze pak umístit na jiný pevný disk.
- **MEMORY** – data se uchovávají jen v paměti. Tento typ je sice nejrychlejší, ale data mohou být snadno ztracena vypnutím či pádem systému.
- **InnoDB** – tyto tabulky používají především cizí klíče a transakce. Jsou optimalizovány pro vyšší rychlost příkazů *UPDATE* a *INSERT* oproti MyISAM. Také jsou doporučeny pro práci s velkým objemem dat. [20]

3 REALIZACE SYSTÉMU PRO VYHODNOCENÍ ZPRÁV

3.1 Návrh systému

System byl navržen tak, aby stáhl e-mailové zprávy, dále je uložil do databáze a následně pak vyhodnotil. Stažení zpráv zajišťuje synchronizace, která kontroluje typ použitého protokolu vybraného účtu a také datum posledního staženého e-mailu tohoto účtu. Uvažuje se černá listina a na základě hodnot na této listině proběhne vyřazení pošty od nechtěných odesílatelů nebo domén.

Při ukládání e-mailových zpráv do databáze se zároveň stahuje hlavička zpráv. Z této hlavičky se zjišťuje odesílatel, doména, předmět a tělo zprávy. Na základě těchto dat pak může být provedeno vyhodnocení zpráv.

Také se načítá obsah bílé listiny, který předurčuje druh a kategorii e-mailu. Pokud není odesílatel nebo doména nalezena na seznamu této listiny, dojde k automatickému vyhodnocení. Při tomto vyhodnocení probíhá analýza klíčových slov v oblasti předmětu, těla a hlavičky e-mailu. Během analýzy se hledají shody těchto klíčových slov.

Na základě váhy u každého klíčového slova, dojde k vypočtení spam skóre zpráv. Podle tohoto skóre se pak určí druh e-mailové zprávy. Dále a také může být nalezena kategorie díky automatické kategorizaci pošty.

Data se vždy přehledně zobrazují do tabulky. Při výpisu dat dochází vždy k rozlišení sudých a lichých řádků tabulky, a to především pro lepší orientaci při větším počtu řádků. Všechna analyzovaná data lze pak přehledně zobrazit v rámci pokročilé statistiky a v grafech.

V práci je aplikována uživatelská správa rozlišující základní a oprávněné uživatele. Každý uživatel se může volně zaregistrovat a poté přihlásit do systému.

3.2 Popis tabulek MySQL databáze

Zde jsou popsány všechny tabulky databáze MySQL, se kterou spolupracuje realizovaná webová aplikace.

Tabulka 9: Popis tabulek aplikace v systému MySQL

Jméno v MySQL	Název tabulky	Popis tabulky
accounts	účty uživatelů	Definuje poštovní účty a jejich nastavení.
bwlists	černá a bílá listina	Rozlišuje odesílatele do černé a bílé listiny.
categories	kategorie e-mailů	Popisuje kategorie pro přiřazení zprávám.
emails	e-mailové zprávy	Ukládá podrobnosti stažených e-mailů.
keywords	klíčová slova	Definuje klíčová slova k analýze e-mailů.
templates	šablony	Uložené šablony pro nové poštovní účty.
types	druhy e-mailů	Obsahuje definice druhů e-mailů.
users	uživatelé aplikace	Ukládá nastavení uživatelů aplikace.

3.3 Vstup do systému

Přístup do systému je chráněn uživatelským jménem a heslem. Bez správné kombinace těchto údajů a úspěšného přihlášení není vstup do aplikace povolen.

Nový uživatel se může zaregistrovat, avšak zadávané uživatelské jméno musí být v rámci systému unikátní. Registrace do aplikace vyžaduje vyplnění několika vstupních údajů. Těmito povinnými údaji jsou uživatelské jméno, heslo a celé jméno uživatele.

Vzhled aplikace byl navržen tak, aby byl jednoduchý pro načítání a snadný pro ovládání uživatelem. Barevné schéma bylo zvoleno v neutrální modré barvě.

3.3.1 Bezpečnostní prvky systému

Aplikace má integrovány základní prvky pro bezpečnost uživatele při práci. Heslo uživatele je již při registraci do systému uloženo v šifrované podobě. Při tomto ukládání se využívá klíčování SHA1 (Secure Hash Algorithm).

Systém dále disponuje funkcí automatického odhlášení uživatele po neaktivitě. Doba této neaktivity vedoucí k automatickému odhlášení je implicitně nastavena na 30 minut. Po vypršení této doby se musí uživatel opětovně přihlásit.

3.4 Popis aplikace

Tato aplikace je rozdělena na tři části – horní, prostřední a spodní. Horní část obsahuje informace o aktuálně přihlášeném uživateli a aktuálním dni v týdnu, datu a čase.

Do nejdůležitější, prostřední, části okna aplikace se vždy načte obsah stránky odpovídající položce, která byla zvolena z horizontálního menu. Samotné menu se nalézá v záhlaví prostředního okna aplikace. Toto menu je přehledně uspořádáno právě do jednoho řádku, tak aby nezaujímal mnoho místa na stránce aplikace. Každou položku menu ilustruje pro ni specifická ikona. Popis načítaných souborů do aplikace ukazuje Tabulka 10.

Tabulka 10: Popis souborů aplikace dle umístění

Umístění a název	Popis souboru
index.php	Základní soubor celé aplikace
addKw.php	Přidání nového klíčového slova pro analýzu pošty
viewEmail.php	Zobrazení podrobností vybraného e-mailu
inc/config.inc.php	Konfigurační soubor celé aplikace
inc/func.inc.php	Soubor s definovanými funkcemi pro správný chod aplikace
pg/analyza.inc.php	Nová analýza e-mailových zpráv pro vyhodnocení
pg/analyza_set.inc.php	Nastavení klíčových slov pro analýzu e-mailových zpráv
pg/imap.inc.php	Nastavení IMAP a POP3 účtů pro synchronizaci zpráv
pg/evaluation.inc.php	Výsledky vyhodnocení e-mailů
pg/options.inc.php	Předvolby a konfigurace systému
pg/prehled.inc.php	Přehled přijatých a uložených e-mailů
pg/set_blist.inc.php	Nastavení černé listiny
pg/set_wlist.inc.php	Nastavení bílé listiny
pg/set_cats.inc.php	Nastavení kategorií pro přiřazení k e-mailům
pg/stats.inc.php	Zobrazení statistiky a grafů
pg/templates.inc.php	Nastavení šablon pro poštovní účty

Poslední spodní část obsahuje informace o názvu projektu, roku vytvoření a autorovi. Zcela napravo se pak nalézá šipka – odkaz pro přesun zobrazení stránky nahoru. Tato volba je užitečná, především je-li zobrazen např. velký počet zpráv.

Jakmile uživatel umístí kurzor nebo klepne myší na některou z položek menu, tak se tato podbarví. Podbarvení připomíná zobrazení ve 3D prostředí. U každé načtené položky menu se zobrazí i její podmenu pro upřesnění požadované volby.

3.5 Druh přijaté pošty

Tento webový systém je schopen automaticky rozlišit tyto druhy přijaté pošty:

- běžná pošta,
- reklamní zpravodaj (newsletter),
- nevyžádaná pošta (spam),
- ostatní a nerozpoznaná pošta.

3.6 Přehled pošty

Přehled pošty je úvodní stránkou, která se zobrazí uživateli ihned po jeho úspěšném přihlášení do systému. Poskytuje přehledný náhled na veškerou přijatou poštu, která byla stažena při synchronizaci.

3.6.1 Poštovní účty

V této části stránky přehledu pošty se vypisují detaily o účtech, které jsou přiřazeny aktuálně přihlášenému uživateli. V pravé části této tabulky lze zjistit datum přijetí posledního e-mailu a informace o tom, zdali je možné se ke schránce připojit – dále se také zobrazí:

- v případě úspěchu celkový počet e-mailů ve schránce,
- v případě neúspěchu podrobnosti o chybě připojení alternativním textem.

Systém se při vyvolání spojení zkusí připojit k serveru dvakrát. Tím se eliminuje případný neúspěch při prvním pokusu o spojení.

Během přidání nebo upravení poštovního účtu je nutné vyplnit všechna pole. Při těchto akcích lze také využít předdefinované adresy serveru pro připojení z dostupných šablon.

3.6.2 Uložené e-mailly

Uložené e-mailly všech účtů uživatele se zobrazí, pokud již alespoň jednou byla provedena synchronizace a byla během ní stažena nějaká pošta. V opačném případě dojde k výpisu chybového hlášení.

Na levé straně řádku s e-mailem se vypisuje ID účtu. Zároveň se rozliší řádek e-mailu nového účtu oddělovačem, kterou je jednoduchá černá linka. V případě první zprávy prvního účtu by byl oddělovač zbytečný, proto je jeho vykreslení potlačeno. Takové oddělení účtů zpřehledňuje výpis řádků s e-mailly. Po umístění kurzoru myši nad buňku s ID účtu se ukáže název účtu pro upřesnění.

Dále se v tabulce vypisuje předmět a tělo zprávy. Obě tyto hodnoty jsou vždy zkráceny, pokud je délka u předmětu větší než 24 a u těla větší než 28 znaků. Celou zprávu lze zobrazit v novém okně, viz níže.

Ve sloupci „UID“ se standardně zobrazuje hodnota UID zprávy. Tato hodnota se však zobrazí pouze v případě, že protokol použitý u zprávy je IMAP. V případě protokolu POP3, který parametrem UID nedisponuje, se zobrazí jen obálka s textem POP.

V pravé části tabulky s načtenými uloženými e-mailly ve sloupci „Akce“ se nalézají ovládací prvky. Těmito prvky lze:

- pomocí symbolu oka zobrazit náhled zprávy v novém okně (v tomto okně lze také přidat zprávu na černou nebo bílou listinu, či změnit kategorii),
- pomocí symbolu křížku vyřadit zprávu z databáze.

Před vyřazením zprávy se zobrazí uživateli dotaz k odstranění. Jakmile dojde k odstranění, nebude již tato zpráva opětovně stažena při další synchronizaci. Předpokládá se totiž, že o takovou poštu uživatel nemá zájem.

Zprávy lze také vyřadit označením jednoho či více zatrhávacích polí na řádku tabulky vlevo. Jakmile vybere uživatel takový e-mail, podbarví se jeho řádek modrou barvou pro lepší rozlišení od ostatních.

3.7 Synchronizace e-mailů

Navržený systém dokáže stáhnout e-mailové zprávy pomocí protokolu IMAP nebo staršího POP3. U stažených zpráv se následně kontroluje, zdali je hlavička odesílatele (konkrétně e-mailová adresa nebo doména) přítomna na černé listině. V takovém případě se zpráva ihned vyřadí. V opačném případě je uložena do databáze.

Při synchronizaci je důležité datum posledního staženého e-mailu poštovního účtu a UID zprávy. Tyto hodnoty se použijí při následující synchronizaci, kdy se porovnávají uložené a již stažené e-maily. Přitom se využívá již uvedeného identifikátoru UID, který je v rámci účtu pro každou zprávu unikátní. Nelze jej však využít u protokolu POP3, viz následující kapitola.

Během nové synchronizace se tedy načtou zprávy od data posledního e-mailu a poté se ověří parametr UID. UID je důležitý především proto, aby se odlišily již stažené a nové zprávy. Během synchronizace se využije funkce *imap_search* se specifickými parametry, které zabraňují označení zprávy jako přečtené během stažení.

V případě nálezů zpráv, se vždy vypisuje počet nalezených zpráv a datum, od kterého je tento počet aktuální (tedy datum posledního přijatého e-mailu). Synchronizace účtu nemusí vůbec uložit žádné e-maily, pokud jsou již tyto v databázi uloženy nebo pokud je vyřadí černá listina. O výsledku každé synchronizace je uživatel informován, stejně tak i v případě, je-li synchronizace aktuální nebo se naopak nemůže klient připojit.

Výsledek synchronizace je vždy vypsán do řádku k aktuálnímu synchronizovanému účtu. V tomto výsledku jsou počty e-mailů vždy barevně rozlišeny, a to následujícími barvami (viz Obrázek 5):

- zelenou pro nově uložené zprávy,
- modrou pro zprávy, které nebylo třeba ukládat,
- černou pro zprávy, které vyřadila černá listina,
- červeně pro případné chybně uložené zprávy.



✉ **Výsledek:** uloženo nových: 6, ?netřeba ukládat: 4, ?černá listina: 4

Obrázek 5: Výsledek synchronizace účtu

Po úspěšně dokončené synchronizaci se tedy zaktualizuje v databázi údaj o datu posledního přijatého e-mailu. V případě protokolu IMAP se přitom uloží také hodnota UID. Vývojový diagram synchronizace se nachází v příloze C1.

3.7.1 Rozdíly synchronizace mezi účty IMAP a POP3

Pro POP3 protokol nelze současně využít všech PHP funkcí určených pro IMAP. Především ale nelze použít jednoznačný identifikátor zprávy UID.

V případě POP3 protokolu se tedy oproti IMAP kontroluje u každého nového e-mailu, zdali jeho parametr *messageID* hlavičky již není uložen v databázi. Tento způsob příjmu zpráv může při větším počtu e-mailů (ať již uložených nebo nových) znamenat zpomalení aplikace.

Některé PHP funkce pro protokol IMAP lze sice aplikovat i na POP3, ale rozdílná je rychlost. Při synchronizaci zpráv za použití protokolu POP3 je zpracování pomalejší právě díky porovnávání hlavičky *messageID*. Při větším počtu nových zpráv nebo při větším počtu účtů, které jsou právě POP3, to může vést k prodloužení doby stažení nových e-mailů. Proto se doporučuje použití protokolu IMAP (je-li takové připojení dostupné).

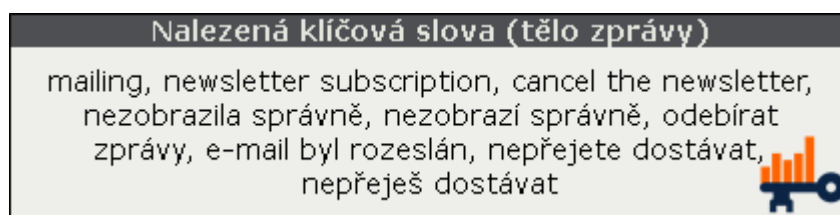
3.8 Automatické vyhodnocení

Automatické vyhodnocení se provádí pomocí nové analýzy e-mailů. Při analýze e-mailových zpráv se hledají shody klíčových slov postupně v oblasti předmětu, hlavičky a těla zprávy. Klíčová slova jsou testována bez ohledu na velikost písmen.

Jakmile se najde shoda klíčového slova, načte se také jeho váha pro určení spam skóre. Toto skóre se vypočítá tak, že se na konci analýzy každého e-mailu sečtou váhy nalezených slov. Vypisuje se do obdélníku s bílým pozadím.

Jestliže se v dané oblasti nalezne shoda, její výskyt se vypíše za anglickou iniciálu oblasti nálezů u e-mailu. Po umístění kurzoru myši nad tuto iniciálu se zobrazí nalezená klíčová slova v dané oblasti, jak ukazuje Obrázek 6. Oblasti nálezů jsou spolu s výskyty jednotlivých shod barevně odlišeny:

- předmět zprávy modrozelenou barvou (iniciála „S“),
- tělo zprávy hnědou barvou (iniciála „B“),
- hlavička zprávy fialovou barvou (iniciála „H“).



Obrázek 6: Nalezená shoda klíčových slov v těle e-mailu

Podle nalezené shody a skóre se určuje druh pošty. V tomto systému se rozlišují čtyři druhy pošty a pro přehlednost je každý z těchto druhů opět barevně odlišen:

- běžná pošta zelenou barvou,
- newsletter modrou barvou,
- nevyžádaná pošta (spam) červenou barvou,
- a ostatní pošta barvou oranžovou.

Pokud se shoda nenalezne, bude zobrazena ikona křížku a druh automaticky nastaven na ostatní (nerozpoznanou) poštu.

Každé klíčové slovo může mít také specifikovanou kategorii, která je nejpravděpodobnější. Podobně jako u spam skóre se dle nejvyššího výskytu kategorie stanoví ta nejvíce odpovídající. Obrázek 7 poskytuje náhled v prvním případě na nerozpoznaný e-mail a v druhém případě na e-mail s nalezenými shody a kategorií.



Obrázek 7: Analýzou vyhodnocené e-maily

3.8.1 Výsledky provedené analýzy

V této volbě se zobrazí výsledky již vykonané analýzy. Tato volba umožní rychlejší načtení, než při provádění nové analýzy. Také se zde již neprovádí testování na shody klíčových slov. Zde jsou zobrazeny výsledky analýzy e-mailů, které byly staženy během všech provedených synchronizací účtů uživatele.

3.8.2 Nová analýza

Během nové analýzy se přehledně do tabulky zobrazují detaily o e-mailové zprávě a v pravé části také oblast a počet nalezených shod (jestliže byly takové nalezeny). Dále se vypíše nalezený druh e-mailu a kategorie.

Byla-li nalezena u e-mailu shoda, lze takovou zprávu také zařadit do kategorie (nebo již přiřazenou kategorii změnit). Jestliže zvolí uživatel novou nebo změní již stávající kategorii, uloží se vždy tato do databáze. Současně dojde k nastavení filtru zobrazovaných zpráv podle zvolené kategorie.

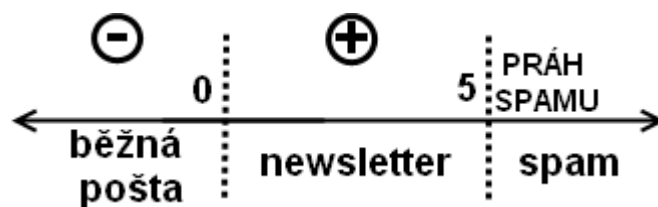
Klepnutím na ikonu žlutého slunce v řádku se zprávou napravo se rozbalí další řádek v tabulce s podrobnostmi pro přidání odesílatele nebo domény na bílou listinu. Černé slunce značí naopak možnost přidání na černou listinu.

Novou analýzu lze spouštět libovolněkrát, dokud uživatel nespustí další synchronizaci pošty. Poté již nelze provádět analýzu současných e-mailů, ale jen nově stažených e-mailů.

3.8.3 Spam skóre

Spam skóre charakterizuje nevyžádanou poštu (spam). Každé klíčové slovo má v systému nastavenou číselnou hodnotu – váhu pro spam skóre. Výchozí hodnotu pro nová slova lze měnit v předvolbách aplikace. Implicitně je přednastaveno: 1.

Nová analýza e-mailů prohledá shody na klíčová slova a při nalezené shodě se načte také váha slova. Jakmile se dokončí prohledávání shod, sečtou se všechny váhy těchto nalezených slov, přičemž se začíná od nuly.



Obrázek 8: Rozlišení druhu pošty dle vypočteného skóre

Dosáhne-li takový součet na hranici meze pro spamovou poštu nebo je-li vyšší, pak je taková pošta považována za spam. Výchozí práh je v aplikaci nastaven na 5 a lze jej měnit individuálně pro každý poštovní účet uživatele. Váhy slov lze také upravovat v nastavení analýzy. Druhy pošty dle vypočteného skóre znázorňuje Obrázek 8.

Výše skóre se vždy zobrazí v okně analýzy vedle výskytu shod. Po umístění ukazatele myši je možné ověřit si nastavený práh aktuálního účtu.

3.8.4 Automatická kategorizace

Systém disponuje funkcí, která se snaží roztrždit nalezené e-mailové zprávy do přednastavených kategorií. Automatická kategorizace se provádí spolu s testy na výskyt klíčových slov během analýzy.

Každé klíčové slovo může mít v databázi předvolenou vybranou kategorii. Několik kategorií je již v aplikaci přednastaveno. Další může doplnit sám přihlášený uživatel.

Princip kategorizace spočívá v tom, že se během automatického vyhodnocení zpráv počítají nalezená klíčová slova (stejně jako u spam skóre). Nejdůležitějším parametrem je však ID kategorie u takového slova. Kategorie nemusí být zadána, protože je volitelná. Následně se sečte výskyt nalezených kategorií. Kategorie, která bude mít nejvyšší hodnotu výskytu u dané e-mailové zprávy, bude nejpravděpodobnější. Tato kategorie bude zobrazena uživateli v roletovém menu a zároveň také podbarvena modrou barvou.

Kategorizace neprobíhá v případě, že odesílatel nebo doména e-mailu byla nalezena na seznamu bílé listiny. Vzhledem k tomu, že se u každé nové analýzy po dokončení uloží výsledek do databáze, bude také uložena i právě nalezená kategorie u e-mailu. V tom, že se na základě analýzy nalezla či nenalezla kategorie, je přihlášený uživatel vždy informován ikonou modrých šipek v kruhu.

Uživatel může nalezenou kategorii změnit výběrem jiné z roletového menu kategorií. Původní položka je vždy také označena alternativním textem. Kategorizaci nastiňuje Obrázek 7 včetně rozpoznané kategorie *Cestování*.

3.8.5 Ruční přidání klíčového slova z e-mailu

Další možností v tomto diplomovém projektu je přidat klíčové slovo ručně. Provede-li se analýza a shoda klíčového slova se nenalezne, zobrazí se uživateli modrá ikona symbolu „plus“ pro volbu ručního přidání.

V nově otevřeném okně pro ruční přidání se zobrazí detaily odpovídajícího e-mailu z oblasti hlavičky, předmětu a těla zprávy. Uživatel může přidat klíčové slovo do pole pro nové klíčové slovo přímo nebo jej vybrat intuitivně pomocí myši. V takovém případě stačí označit část textu v odpovídající oblasti zprávy. Tento text se zkopíruje do pole pro přidání a přepínač se označí podle oblasti slova. Dále je povinné uvést váhu slova (nebo ponechat výchozí hodnotu), volitelně lze také specifikovat druh pošty.

Pole pro přidání nového klíčového slova je omezeno na maximálně 30 znaků. V horní části okna pro přidání klíčového slova se nachází prvky pro znovunačtení a náhled e-mailu, dále pro zavření okna. Také je zde zobrazena nápověda pro uživatele.

3.8.6 Filtr pošty

V aplikaci je dostupná funkce pro filtrování zobrazení dle druhu pošty a kategorie. Nabídka druhu pošty je v roletovém menu navíc barevně odlišená. Filtrovat lze jen podle jednoho vybraného filtru nebo podle obou kombinací.

Jednotné vypnutí obou filtrů lze provést klepnutím na ikonu smetáčku napravo. Bez nastavených filtrů aplikace opět vypíše všechny e-maily s provedenou analýzou.

3.9 Nastavení analýzy

V tomto menu se provádí konfigurace, která ovlivňuje průběh analýzy e-mailových zpráv. Pokaždé, když vybere uživatel jednu z voleb nastavení, zobrazí se kromě názvu zvolené volby také počet uložených položek k dispozici (v závorce).

3.9.1 Klíčová slova

Při přidávání nového klíčového slova je vždy nutné vybrat a vyplnit právě jednu oblast s klíčovým slovem. Každá oblast s klíčovým slovem může také obsahovat více pojmů, přičemž se tyto pojmy oddělují mezerou.

Doporučuje volit 2 až 4 pojmy, maximálně však 5 pojmu. Takový počet je vždy kontrolován při přidávání nebo úpravě klíčového slova. Pro doplnění lze nalézt v okně žlutý obdélník s nápovědou.

3.9.2 Kategorie

V tomto menu lze nalézt kategorie, dle kterých se filtrují a zobrazují analyzované e-maily. Každá kategorie specifikuje vybranou poštu. Při vytváření nové takové kategorie je třeba zadat její název, volitelně i popis.

3.9.3 Černá listina

Tato listina zakazuje příjem zpráv, jejichž e-mail odesílatele nebo celá doména jsou uloženy právě na tomto seznamu. Při synchronizaci jsou pak takoví odesílatelé či domény vynechány a jejich pošta není uložena.

Do této listiny lze přidávat nebo upravovat položky manuálně v menu u nastavení analýzy nebo automaticky během vyhodnocení v analýze e-mailů.

3.9.4 Bílá listina

Tato listina upřednostňuje druh a kategorii e-mailových zpráv. Tyto parametry jsou při vytváření nové položky povinné.

Stejně jako v předchozím případě (u černé listiny) se uložení odesílatelů a domén vypisují společně do jednoho sloupce, přestože v tabulce MySQL využívají sloupce dva. Pro rozlišení se nachází u každé položky jedna ze značek:

- „O“ pro označení odesílatele,
- „D“ pro označení domény.

Oproti černé listině jsou zde tedy odpovídajícím odesílatelům a doménám přiřazeny položky druh pošty a kategorie. Změny těchto položek lze provést jednoduše výběrem z menu v editaci.

Obdobně i zde lze přidat nebo upravit položky manuálně nebo automaticky analýzou. Je-li však již e-mail umístěn na seznamu bílé listiny, bude ikona žlutého slunce pro přidání záměrně skryta. V opačném případě se zobrazí i řádek pro specifikaci vstupních parametrů bílé listiny.

3.10 Poštovní účty

Tato nabídka nabízí konfiguraci poštovních účtů pro synchronizaci. Každý z těchto poštovních účtů umožňuje nastavit následující položky:

- název samotného účtu,
- adresu serveru pro připojení,
- jméno a heslo pro připojení k tomuto účtu,
- protokol IMAP nebo POP3.

Uživatel také může nastavovat šablony k účtům. Tyto šablony poskytují adresu serveru právě k poštovním účtům. Usnadňují tak vytváření účtů nových a úpravy účtů stávajících. Během přidání nového účtu tak přihlášený uživatel jednoduše zvolí poštovní server ze seznamu předdefinovaných šablon, který může ještě před samotným uložením upravit na požadovanou hodnotu.

Blok poštovního serveru musí být vždy povinně ohraničen párem složených závorek. Za tímto blokem následuje název adresáře, kterým je často *INBOX* označující Doručenou poštu.

Bylo zaregistrováno několik poštovních účtů na různých serverech využívajících různých protokolů. Tyto účty byly vytvořeny pro účely diplomové práce a pro jsou již přednastaveny v databázi.

Ukázka platného zápisu adresy poštovního serveru POP3:

{pop3.seznam.cz:995/novalidate-cert/pop3/ssl}INBOX

Z výše uvedeného zápisu je patrné, že se bude klient připojovat na poštovní server Seznam.cz na zabezpečeném SSL portu 995 bez certifikátu a výchozím adresářem je Doručená pošta. Další příklady uvádí Obrázek 9.

Název	Server	Login
db99test@gmail	{pop.gmail.com:995/novalidate-cert/pop3/ssl}INBOX	db99test@gmail.com
djkart@centrum	{imap.centrum.cz:993/imap/ssl}INBOX	djkart@centrum.cz
db01test@seznam	{pop3.seznam.cz:995/novalidate-cert/pop3/ssl}INBOX	db01test@seznam.cz
db01@atlas	{imap.atlas.cz:143/imap/novalidate-cert}INBOX	db01@atlas.cz

Obrázek 9: Výpis uložených poštovních účtů

3.10.1 Šablony účtů

Specifikovat lze také šablony účtů, což zjednodušuje práci při vytváření a editaci IMAP a POP3 účtů. Zde je vždy potřeba zadat název šablony a server – tyto položky jsou povinné. Volitelně lze přidat popis šablony účtu (např. parametry připojení).

Šablony jsou v systému vedeny veřejně, modifikovat je může jakýkoli oprávněný uživatel (viz níže).

3.11 Předvolby

V tomto menu lze provádět změny ovlivňující chod samotného systému. Během synchronizace, a to především je-li definováno více účtů nebo účty protokolu POP3, může vypršet časový limit skriptu. Výchozí hodnota časového limitu pro vykonání skriptu v jazyce PHP je nastavena na 30 sekund. Pro účely této diplomové práce byla hodnota záměrně navýšena na 60 sekund.

Pokud však ani tato doba nebude dostačující pro dokončení synchronizace e-mailových zpráv, nabízí se zde předvolba pro úplné zakázání tohoto časového limitu. V tomto případě nebude vykonání samotného skriptu nijak limitováno a stáhnou se tak všechny nové e-mailové zprávy. Taková operace může trvat i několik minut, než se dokončí.

Pro aktuálně přihlášeného uživatele lze také nastavit zákaz výpisu spravovaných účtů v nabídce přehledu pošty. I pomocí této volby lze zrychlit běh aplikace na této stránce. Dále lze specifikovat výchozí váhu pro nově přidávaná klíčová slova, přípustné je i desetinné kladné či záporné číslo.

3.11.1 Víceúrovňová uživatelská správa

Registrace uživatelů do systému není nijak uzavřená a každý uživatel si tak může volně vytvořit nový uživatelský účet. Správu lze konfigurovat v menu předvoleb.

Nadřízení (oprávnění) uživatelé mohou spravovat podřízené uživatelské účty. Podřízení uživatelé mají základní oprávnění a jejich práce s webovou aplikací je omezena.

Přidání nového uživatele vyžaduje parametr UID a uživatelské jméno. Tyto hodnoty si musí oprávněný uživatel obstarat od podřízeného uživatele. Bez těchto hodnot nelze v přidávání pokračovat. Před přidáním nového uživatele do správy oprávněného se kontroluje:

- zdali byly parametry UID a uživatelské jméno zadány,
- zdali se dle zadaných parametrů vyhledá správný uživatel,
- zdali je vyhledaný uživatel již pod správou aktuálního uživatele,
- zdali není vyhledaný uživatel pod správou jiného uživatele.

Pokud se úspěšně projde touto kontrolou, dojde k přidání nového podřízeného (základního) uživatele do seznamu.

Oprávnění uživ.	Spravovaných uživ.
oprávněný	2

Obrázek 10: Oprávnění uživatele s počtem spravovaných uživatelů

V rámci správy podřízených uživatelů lze u těchto měnit heslo, ukončit párování, dále je upravovat nebo zcela odstranit ze systému. Pro oprávněné uživatele byla zvolena barva žlutá, pro základní uživatele barva bílá. Úroveň oprávnění lze ověřit při jakékoli zobrazené stránce v alternativním textu přihlášeného uživatele v řádku okna aplikace vpravo nahoře.

V případě oprávněného uživatele se v uživatelském panelu zobrazí počet spravovaných uživatelů a naopak u základního se vypíše název a celé jméno nadřazeného účtu.

Základní uživatel má v aplikaci oproti oprávněnému následující omezení:

- nemůže přidat, upravit či smazat šablony účtů,
- nemůže přidat, upravit či smazat kategorie,
- nemůže měnit prahovou hodnotu pro spam skóre účtů.

3.12 Pokročilá statistika

Vzhledem k tomu, že se do databáze ukládá velké množství dat, je vhodné tato data uceleně prezentovat. K tomu slouží menu s pokročilou statistikou.

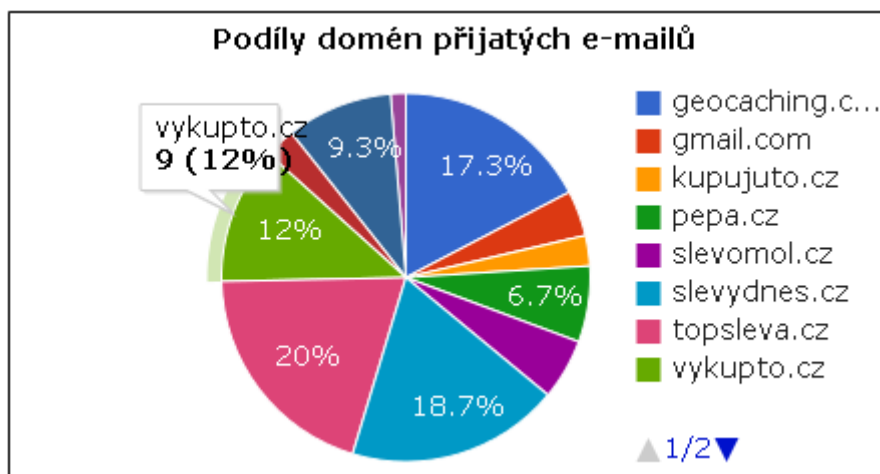
Pokročilá statistika zobrazuje koláčové grafy a spolupracuje s API²⁰ rozhraním Google Charts. Pro správnou funkci těchto grafů se do HTML stránky načte kód Javascript. Také je potřeba nastavit parametry grafu. [25]

Dále se pod vykreslenými koláčovými grafy vypisuje číselná statistika v tabulce. Tabulka nabízí přehled o počtech zpráv, účtů, dostupných klíčových slov, kategorií a položek na černá a bílé listině. Ukázka celé stránky s pokročilou statistikou je k dispozici v příloze B.

3.12.1 Grafy Google Charts

V parametrech se zadávají vstupní data a typ grafu. Google Charts pracují jen v online režimu, nelze je tedy použít pro offline využití.

Jako typ grafu byl zvolen 2D koláčový graf, a to především proto, že poskytuje dostatečný přehled i při větším počtu zobrazovaných položek. Náhled takového grafu přibližuje Obrázek 11.



Obrázek 11: Ukázka Google Charts grafu v aplikaci

²⁰ Application Programming Interface

Každá položka dostane přiřazenou náhodnou barvu pro jednoznačnou identifikaci. Názvy položek, stejně jako jejich barvy, se poté vypisují napravo od zobrazeného grafu do legendy. V případě, že se nevejdou všechny položky v legendě na jednu stránku, nastane stránkování. Mezi takto vzniklými stránkami lze procházet pomocí modrých šipek pod legendou.

Pokud je výšeč položky v grafu dostatečně velká v porovnání s ostatními, vypíše se přímo procentuální hodnota. Jakmile umístí uživatel kurzor na některou z výšečí grafu, zobrazí se bublina obdélníkového tvaru. V této bublině se nachází název položky, dále pak počet výskytů a procentuální poměr v grafu.

3.12.2 Grafy zobrazované ve statistice

Grafy se vykreslují na základě aktuálních vstupních dat. Tato data jsou vždy načtena z databáze. K dispozici jsou celkem 4 grafy:

- podíly domén přijatých e-mailů,
- nejčastější domény přijatých e-mailů,
- výskyt kategorií v e-mailech,
- výskyt druhů pošty.

Druhý graf vychází z hodnot prvního grafu. K vykreslení druhého grafu nemusí dojít v případě, že nebudou k dispozici dostatečná data. Musí být totiž splněna podmínka alespoň 20 různých domén přijatých e-mailových zpráv.

4 ZÁVĚR

Tato diplomová práce popisuje a realizuje systém pro automatizované vyhodnocení e-mailových zpráv. Systém byl navržen za použití jazyka PHP, frameworku jQuery a databázového systému MySQL. Systém pracuje v prostředí internet.

Tento systém stahuje pomocí protokolů IMAP a POP3 e-mailové zprávy, které se uloží do databáze pro další využití. Avšak v případě, že byl odesílatel nebo doména e-mailu nalezena na seznamu černé listiny, k uložení nedojde. Přístup do systému je zabezpečen a je nutné se autorizovat zadáním správné kombinace uživatelského jména a hesla. Rovněž je nabídnuta registrace pro nové uživatele.

Uložené e-maily se pomocí analýzy klíčových slov automaticky vyhodnocují. Analýzou se tedy určí druh pošty a případně také kategorie zpráv. Rozhodování o druhu pošty se zjišťuje na základě vypočteného skóre zprávy. Tak lze rozlišit běžnou, reklamní a nevyžádanou poštu. Je-li skóre vyšší než nastavený práh, jedná se automaticky o spam.

Dále se také uvažují odesílatelé a domény na bílé listině, která, pokud jsou tyto parametry nalezeny v hlavičce zprávy, přeskočí hledání shod klíčových slov a automaticky zobrazí druh a kategorii e-mailu. Kategorie, která má nejvyšší počet výskytů u shodujících se klíčových slov s analyzovanou zprávou, bude považována za nejpravděpodobnější.

Vyhodnocená data zobrazuje přehledně pokročilá statistika systému. Vykreslovány jsou čtyři 2D koláčové grafy a dále tabulka se statistickými hodnotami e-mailů, analýzy a uživatelských účtů. Systém disponuje víceúrovňovou uživatelskou správou a rozlišuje dva druhy oprávnění uživatelů – základní a oprávněné uživatele.

SEZNAM LITERATURY

- [1] ROTH, Mark D. *MUAs, MTAs, and MDAs* [online]. [cit. 2014-03-14]. Dostupné z: <http://www.feep.net/sendmail/tutorial/intro/MUA-MTA-MDA.html>.
- [2] KRÁTKÝ, Robert. *MTA* [online]. 20. 10. 2008 [cit. 2014-03-28]. Dostupné z: <http://www.abclinuxu.cz/slovník/mta>.
- [3] KRÁTKÝ, Robert. *MUA* [online]. 20. 10. 2008 [cit. 2014-03-28]. Dostupné z: <http://www.abclinuxu.cz/slovník/mua>.
- [4] KRÁTKÝ, Robert. *MDA* [online]. 2. 6. 2007 [cit. 2014-03-28]. Dostupné z: <http://www.abclinuxu.cz/slovník/mda>.
- [5] DENT, Kyle D. *Postfix: kompletní průvodce*. 1. vyd. Praha: Grada, 2005, 237 s. ISBN 80-247-1029-3.
- [6] TechTarget: Search Exchange. *SMTP(Simple Mail Transfer Protocol)* [online]. c2014 [cit. 2014-03-29]. Dostupné z: <http://searchexchange.techtarget.com/definition/SMTP>.
- [7] MÍČEK, Emil. *Elektronická pošta – SMTP* [online]. [cit. 2014-04-02]. Dostupné z: <http://www.fi.muni.cz/~kas/p090/referaty/2006-jaro/st/smtp.html>.
- [8] Techopedia. *Open Relay* [online]. c2014 [cit. 2014-04-03]. Dostupné z: <http://www.techopedia.com/definition/1699/open-relay>.
- [9] ADÁMEK, Martin. *Spam: jak nepřivolávat, nepřijímat a nerozesílat nevyžádanou poštu*. 1. vyd. Praha: Grada, 2009, 166s. ISBN 978-80-247-2638-0.
- [10] WOLFE, Paul, Mike W. ERWIN a Charlie SCOTT. *Antispam: metody, nástroje a utility pro ochranu před spamem*. Vyd. 1. Překlad Ivo Fořt. Brno: Computer Press, 2004, 375 s. ISBN 80-251-0479-6.
- [11] ONESolution s.r.o.: Znalostní báze. *Jaký je rozdíl mezi POP3 a IMAP?* [online]. c2014 [cit. 2014-04-11]. Dostupné z: <http://www.onehelp.cz/onebit/kb/jaky-je-rozdil-mezi-pop3-a-imap>.
- [12] STANIK, Rastislav. *Ako funguje e-mail* [online]. 23. 6. 2004 [cit. 2014-04-12]. Dostupné z: <http://www.abclinuxu.cz/clanky/site/ako-funguje-e-mail>.
- [13] RESNICK, P. *Internet Message Format* [online]. 2008 [cit. 2014-04-14]. Dostupné z: <http://tools.ietf.org/html/rfc5322>.
- [14] Technická univerzita v Liberci. *Elektronická pošta* [online]. 2013 [cit. 2014-04-15]. Dostupné z: http://liane.tul.cz/cz/Elektronick%C3%A1_po%C5%A1ta.

- [15] CentralNews. *Co je to newsletter?* [online]. c2013 [cit. 2014-04-16]. Dostupné z: <http://www.centralnews.cz/co-je-to-newsletter>.
- [16] Adaptic, s.r.o. *Newsletter* [online]. c2014 [cit. 2014-04-17]. Dostupné z: <http://www.adaptic.cz/znalosti/slovnicek/newsletter/>.
- [17] Adaptic, s.r.o. *MySQL* [online]. c2014 [cit. 2014-04-17]. Dostupné z: <http://www.adaptic.cz/znalosti/slovnicek/mysql/>.
- [18] GROFF, James R. a Paul N. WEINBERG. *SQL: kompletní průvodce*. Vyd. 1. Brno: CP Books, 2005, 936 s. ISBN 80-251-0369-2.
- [19] Artic Studio. *Co je to databáze MySQL?* [online]. c2011 [cit. 2014-04-20]. Dostupné z: <http://www.artic-studio.net/slovnicek-pojmu/databaze-mysql>.
- [20] WEDOS Internet, a.s.: Znalostní báze. *MySQL - typy tabulek* [online]. 2010 [cit. 2014-04-20]. Dostupné z: <http://kb.wedos.com/mysql/typytabulek.html>.
- [21] Kerio: Knowledge Base. *Jak pracuje Bayes filtr a SpamAssassin* [online]. [cit. 2014-04-22]. Dostupné z: <http://kb.kerio.com/product/kerio-connect-cz-/how-to/jak-pracuje-bayes-filtr-a-spamassassin-560.html>.
- [22] HORÁK, Michael. *Internetový systém pro odesílání novinek emailem*. Brno: Vysoké učení technické v Brně, 2011. 64 s. Bakalářská práce.
- [23] FRKAL, Jan. *Webový systém pro správu uživatelů sítě*. Brno: Vysoké učení technické v Brně, 2010. 49 s. Bakalářská práce.
- [24] KOSEK, Jiří. *PHP - tvorba interaktivních internetových aplikací: podrobný průvodce*. Vyd. 1. Praha: Grada, 1999, 490 s. ISBN 80-716-9373-1.
- [25] Google Developers. *Google Charts* [online]. [cit. 2014-04-25]. Dostupné z: <https://developers.google.com/chart/>.
- [26] PONKRÁC, Miloslav. *PHP a MySQL bez předchozích znalostí: průvodce pro samouky*. Vyd. 1. Brno: Computer Press, 2007, 221 s. ISBN 978-80-251-1758-3.
- [27] Adaptic, s.r.o. *PHP* [online]. c2014 [cit. 2014-04-29]. Dostupné z: <http://www.adaptic.cz/znalosti/slovnicek/php/>.
- [28] Slezská univerzita v Opavě. *Rozšíření MIME* [online]. [cit. 2014-05-16]. Dostupné z: <http://zam.opf.slu.cz/botlik/CD-dalsi/CD-mime/mime2.htm>.

- [29] PERNICA, Tomáš: Služby IT. *Jak funguje email* [online]. 20. 3. 2008 [cit. 2014-05-12]. Dostupné z: <http://www.tomp.eu/2008/03/20/jak-funguje-email>.
- [30] Media Temple, Inc. *Understanding an email header* [online]. [cit. 2014-05-12]. Dostupné z: <https://kb.mediatemple.net/questions/892/Understanding+an+email+header>.
- [31] HORÁK, Vladimír. *Úvod - Co je SPAM a jak se mu bránit* [online]. 2006 [cit. 2014-05-13]. Dostupné z: <http://uvt1.cuni.cz/email/spam/uvod.html>.
- [32] KRČMÁŘ, Petr. *Linux: postavte si počítačovou síť*. 1. vyd. Praha: Grada, 2008, 182 s. ISBN 978-80-247-1290-1.
- [33] Střední průmyslová škola elektrotechnická a informačních technologií Brno. *Elektronická pošta e-mail* [online]. [cit. 2014-05-13]. Dostupné z: https://moodle.sspbrno.cz/pluginfile.php/2273/mod_resource/content/0/email.pdf
- [34] Infonia, s.r.o. *Pošta IMAP nabízí více než POP (a vysvětlení, o co se jedná)* [online]. 2. 6. 2006 [cit. 2014-05-13]. Dostupné z: <http://www.infonia.cz/posta-imap-nabizi-vice-nez-pop3>.
- [35] NOVOTNÁ, Markéta. *Newsletter* [online]. 20. 2. 2012 [cit. 2014-05-14]. Dostupné z: <http://wiki.knihovna.cz/index.php/Newsletter>.
- [36] NOVOTNÝ, Pavel. *Řešení e-mailového serveru a aplikace pro jeho správu* [online]. [cit. 2014-05-15]. Dostupné z: <http://programatornavolnenoze.cz/pages/rmails/>.
- [37] MXToolBox, Inc. *SuperTool Beta7* [online]. c2014 [cit. 2014-05-16]. Dostupné z: <http://mxtoolbox.com/SuperTool.aspx>.

SEZNAM POUŽITÝCH ZKRATEK

API	Application Programming Interface
ASCII	American Standard Code for Information Interchange
ASP	Active Server Pages
DBMS	Database Management System
DNS	Domain Name Server
ESMTP	Extended Simple Mail Transfer Protocol
HTML	Hypertext Markup Language
IMAP	Internet Message Access Protocol
IP	Internet Protocol
LMTP	Local Mail Transfer Protocol
MDA	Mail Delivery Agent
MIME	Multipurpose Internet Mail Extensions
MUA	Mail User Agent
PDF	Portable Document Format
PHP	Personal Home Page
pop3	Post Office Protocol
SMTP	Simple Mail Transfer Protocol
SQL	Structured Query Language
TCP	Transmission Control Protocol
Telnet	Telecommunication Network
UDP	User Datagram Protocol

SEZNAM PŘÍLOH

A	Ukázka webové aplikace	60
B	Pokročilá statistika.....	61
C	Vývojové diagramy.....	62
C.1	Synchronizace e-mailů.....	62
C.2	Vyhodnocení (analýza) e-mailů.....	63
D	Obsah příloženého CD.....	64

A UKÁZKA WEBOVÉ APLIKACE



Automatické vyhodnocení e-mailových zpráv


[Přehled pošty](#)


[Q Vyhodnocení](#)


[Nastavení analýzy](#)


[Poštovní účty](#)


[Statistika](#)


[Předvolby](#)


[Odhlášení](#)

VÝSLEDKY

NOVÁ ANALÝZA

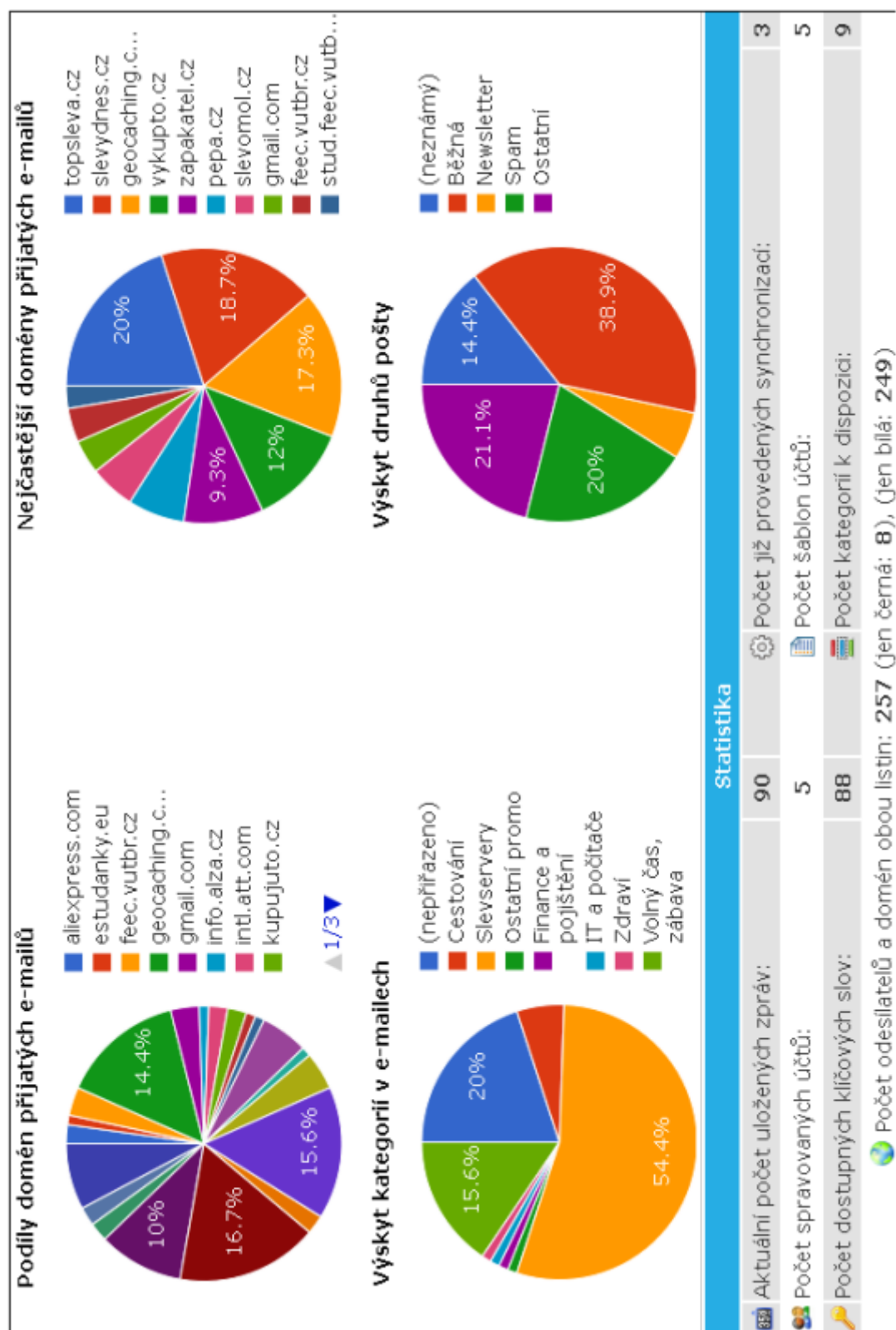
Nová analýza e-mailů pro vyhodnocení

Účet	Od	Čas	Předmět	Výskyty shod	Druh	Kategorie(změnit)
1	noreply@geocaching.com	25.5.14 15:24	[LOG] Owner: pavkas01 fou...	 Odesílatel	Běžná	Volný čas, zábava
1	noreply@geocaching.com	25.5.14 20:21	[LOG] Owner: marsikula01 ...	 Odesílatel	Běžná	Volný čas, zábava
1	noreply@geocaching.com	25.5.14 21:52	[LOG] Owner: Gisbourne fo...	 Odesílatel	Běžná	Volný čas, zábava
1	xpetra04@stud.feec.vutbr.cz	26.5.14 11:50	MGMP 13/14L (134001): Re:...	 -1.2 B1 H2	Běžná	Cestování
4	zpravodaj@topsleva.cz	25.5.14 15:40	Chcete se zařít zdravě...	 Doména	Newsletter	Slevservery
4	newsletter@slevydnas.cz	25.5.14 16:52	★ Dopřejte si oddych s...	 Doména	Newsletter	Slevservery
4	zpravodaj@topsleva.cz	25.5.14 19:08	♥ Vychytávky ve světě ...	 Doména	Newsletter	Slevservery
4	slevy@vykupto.cz	26.5.14 04:54	Dobrý den, máte u nás ...	 Doména	Newsletter	Slevservery
4	nabidky@zapakatel.cz	26.5.14 07:07	Thajský ráj Kadeřnic...	 Doména	Newsletter	Slevservery
4	zpravodaj@topsleva.cz	26.5.14 07:44	✓ Ideální dárek pro...	 Doména	Newsletter	Slevservery

Systém pro automatické vyhodnocení e-mailových zpráv

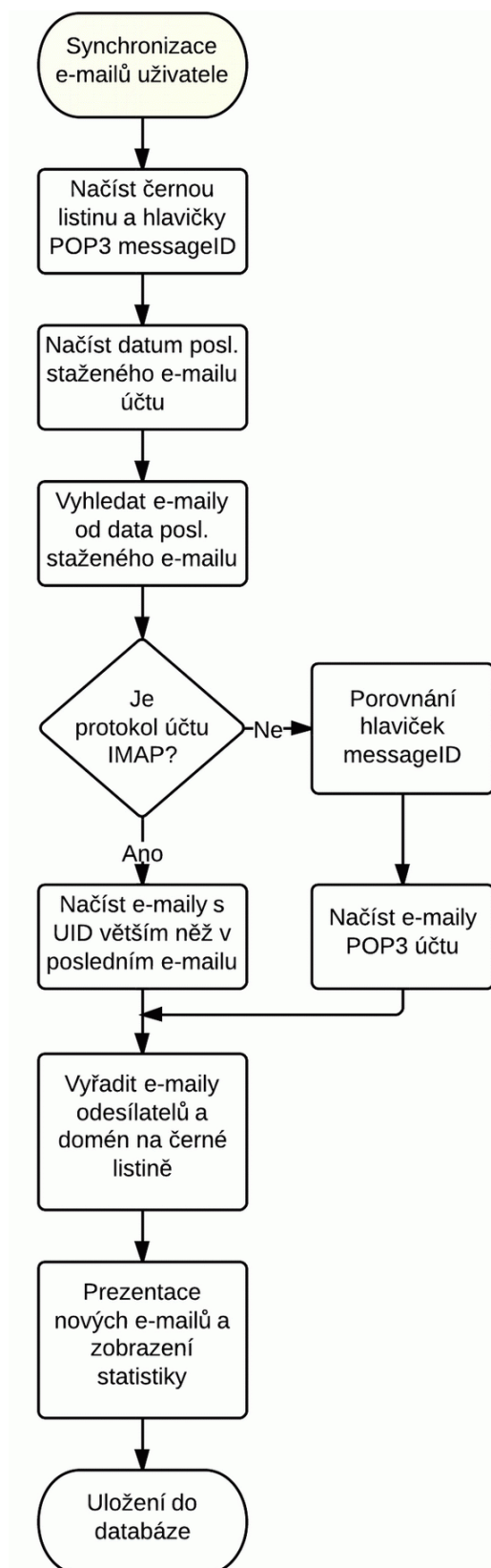
© 2013-2014 Bc. Jan Frkal, FEKT VUT

B POKROČILÁ STATISTIKA

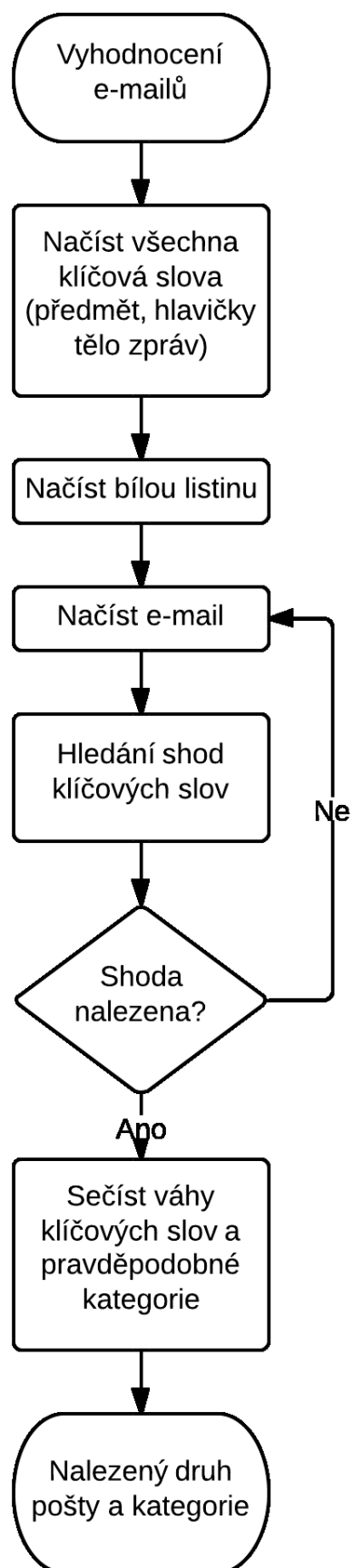


C VÝVOJOVÉ DIAGRAMY

C.1 Synchronizace e-mailů



C.2 Vyhodnocení (analýza) e-mailů



D OBSAH PŘILOŽENÉHO CD

Příložené CD obsahuje adresář s webovým systémem **DP** a druhý adresář **SQL_zdroj** s SQL soubory pro vytvoření databáze a tabulek MySQL. V kořenovém adresáři se nachází PDF soubor s textem diplomové práce.

V adresáři **DP** je nutné nastavit v podadresáři *inc* konfigurační soubor *config.inc.php* tak, aby se webová aplikace mohla připojit k MySQL serveru a dále s ním pracovat. Výchozím uživatelským jménem a heslem pro testování běhu systému jsou: **admin**, **admin** s právy oprávněného uživatele. Ve správě podřízených uživatelů v předvolbách lze přidat zkušebního uživatele **test** s UID **2**. Online verze aplikace je k dispozici na <http://kartovo.php5.cz/AN/index.php>.

Webový systém byl vyvíjen a testován v jazyce PHP 5.4.17 na serveru Apache 2.4.9 (obsluhováno programem USB Webserver 8.6).

Aplikace byla testována v prohlížeči Mozilla Firefox 29.0.1. MySQL databáze vytvořena a spravována aplikací Adminer 4.0.2. Použité formátování v PHP a MySQL je UTF-8.