VYSOKÁ ŠKOLA EKONOMICKÁ V PRAZE
FAKULTA INFORMATIKY A STATISTIKY

Obor: Informatika

Důvěryhodnost informací na Internetu

(bakalářská práce)

Autor: Miroslav Vlach

Vedoucí práce: Ing. Antonín Rosický, CSc.

Prohlášení:

Prohlašuji, že jsem celou bakalářskou práci vypracoval samostatně a vyznačil všechny citace z pramenů.

V Praze dne 9. srpna 2006 .....

podpis

Poděkování:

Na tomto místě bych rád poděkoval vedoucímu mé bakalářské práce panu Ing. Antonínu Rosickému, CSc. za cenné rady a odborné vedení. Můj vděk patří také všem, kteří mě při práci podporovali, hlavně mým rodičům, Ing. Petře Kadlecové a Andree Madárové.

Abstrakt:

V této práci jsem se zabýval otázkou informací dostupných na Internetu, jejich spolehlivostí a možným využitím v reálném životě. Hledáním na Internetu jsem si sám ověřil, že je na něm nepřeberné množství informací, které nejsou autorizované, v horším případě, jsou pak záměrně zfalšované. Způsob vyhodnocování důvěryhodnosti informací jsem rozdělil do několika kategorií: důvěryhodné, pochybné a matoucí. Řešil jsem postup ochrany těchto informací a potenciálnost jejich zneužití. Dále jsem věnoval velkou pozornost Internetu, jeho historii, možnostem ochrany dat na něm se vyskytujících a služeb, které jsou k dispozici.

Cílem mé práce bylo najít způsoby, jak rychle a pohodlně získat informace, které by byly zajištěny a jejichž obsah, by byl pro danou situaci kredibilní. Dále jsem se snažil čtenáře informovat o všech místech, kde jsou uloženy jakékoli informace a o nebezpečích, která jim mohou zkomplikovat hledání nebo je dovést do slepé uličky. Okrajově jsem se zmínil o službách, které Internet poskytuje, např. elektronická pošta, World Wide Web, FTP atd.

Hlavní problém, který jsem ve své práci řešil, bylo zjistit význam důvěryhodnosti ve spojení s Internetem a pojetí věrohodnosti z různých úhlů pohledu. Snažil jsem se stanovit, co ovlivňuje spolehlivost a jak tomuto působení zamezit. V závěru jsem pro dokreslení vnímání důvěryhodnosti Internetu použil data z průzkumů a grafů, které byly sestaveny na základě vypovědí uživatelů a neuživatelů Internetu různých věkových skupin.

Perspektiva do budoucna není špatná. S lepšími technologiemi bude i větší možnost zabezpečit informace proti zcizení nebo zneužití, a tím se zlepší ochrana identity uživatelů. S modernějšími vyhledávacími stroji se zkrátí čas k nalezení potřebných dokumentů a zároveň se sníží náklady a náročnost projektů z toho plynoucích. Účelnost Internetu by se zvětšila, kdyby k němu měl přístup každý a kdokoli mohl využívat služby, které nám nabízí.

Abstract:

The following text deals with issue of information available at Internet, its reliability and possible use in common life. Searching the Internet I realized there was a huge amount of information that was not authorized, in some cases even fixed.

Method of evaluation of information creditability can be divided into three categories: credible, doubtful and misleading. In the text, I also describe process of information protection and possibility of information abuse. I also pay attention to Internet, its history, possibility of data protection at Internet and also services available at Internet.

The aim of this text is to find ways how to get credible and guaranteed information quickly and easily. I try to inform about rooms where any information is saved, and dangers that can make looking for information more complicated. Marginally, I mention services provided at Internet, such as electronic mail, World Wide Web, FTP etc.

Main issue I deal with is to find the meaning of „credibility at Internet“and different approaches to credibility. I try to set what is credibility influenced by and how to prevent the negative factors from influencing the information. It the end I show data and charts from several researches on credibility of Internet.

Future perspective is not bad. With advanced technologies there is expectation that information will be better protected against alienation or abuse and the quality of protection of user identity will be higher. With advanced search machinery time to finding proper information will be shorter and the costs and complicated projects connected with it will be lower. Effectivity of Internet would be higher, if anybody were able to use it.

OBSAH

SEZNAM TABULEK

SEZNAM OBRÁZKŮ

1 ÚVOD

„Máte-li špatné informace, ale dokonalou logiku, pak jsou vaše závěry jistě mylné. Dopřejete-li si totiž sem tam nějakou trhlinu v logickém uvažování, můžete díky náhodě dospět ke správnému závěru.“
[Autor neznámý, str. 98, Matematika 1 pro ekonomické fakulty, Jan Coufal a Jindřich Klůfa]

Z tohoto citátu jasně vyplývá vážnost tématu, které jsem si vybral a zpracoval. „Máte-li špatné informace“ – a že není problém takové na Internetu najít, nelze z nich stanovit pravdivé závěry, což má za následek malou nebo žádnou kredibilitu. Proto jsem se také zaměřil na různá pojetí důvěryhodnosti. Zkoumal jsem, jaké faktory ovlivňují kvalitu dokumentů. Soustředil jsem se na hlavní problémy informací ve vztahu k uživatelům i k Internetu. Zabrousil jsem na citlivá místa anonymity a soukromí, která jsou pro nás jak přínosem, tak i problémem. Těmi právě jsou nedostatečné pokrytí vyhledávacími agenty, informací bez známky potvrzení pravosti aj. Dále mě zajímaly etické problémy dnešní společnosti, jako jsou nelegální činnosti, manipulace s informacemi nebo ovlivňování skupin lidí pomocí zfalšovaných údajů. Tvrzení, že Internet není bezpečný se pokouší vyvracet mechanismy a technologie, jež se snaží bránit nájezdům zločinců a vandalů.

Argumentů pro zpracování této základní myšlenky je několik. Určitě není pochyb, že hodnocení důvěryhodnosti informací je aktuální a ožehavé téma. V dnešním rozmachu informačních technologií a služeb, kdy jsme závislí na přínosech Internetu, na datech, která čerpáme a dále zpracováváme, se někdo, kdo s tímto světem nemá nejmenší zkušenosti, může mylně domnívat, že pokud bude chtít nějaký údaj vyhledat, nebude s tím mít sebemenší problémy. My zkušenější víme, že to je chybná představa. Kolikrát se nám stalo, že místo, abychom za krátký okamžik získali dokument s cennou a postačující informací, jsme se hodiny a hodiny potili u počítače ochotně vyzkoušet cokoli, abychom uspěli a pakliže se nám to nakonec podařilo, zjistili jsme, že si s nimi pohrál nějaký „vtipálek“.

Zejména kvůli osobním zkušenostem i zkušenostem nebližšího okolí, jsem se rozhodl zvýšit informovanost a představit techniky, které by předešly problémům s nekonečným hledáním zapříčiněným nekvalitními daty, která se dají těžko rozlišit od autentických, spolehlivých a rychle dostupných. Zdroje, ze kterých jsem čerpal, jsou sice rozmanité, ale přednost jsem dával odborným publikacím, které se zabývají tímto námětem. Pro jedince, kteří zatím velké zkušenosti s Internetem nemají, jsem obecně popsal, čím se Internet stal od počátku jeho vzniku. Kromě toho jsem charakterizoval, jaké přednosti přináší a které služby nabízí. Velkou pozornost jsem věnoval právě věrohodnosti informací, jejich hodnocení a zabezpečení této kvality. Tuto část jsem završil výzkumem, který proběhl minulý rok, a v němž se ptali občanů České republiky, zda-li věří informacím umístněných na Internetu.

Očekávám, že po přečtení této práce, budeme schopni odpovědět například na tyto otázky:
Proč se Informace musí hodnotit?
Proč se na Internetu nalézají nedůvěryhodné zdroje?
Jakým způsobem lze minimalizovat riziko zfalšování dat?
Jaké výhody přináší přes všechna úskalí Internet?

Právě tato práce by měla pomoci při minimalizování chyb, kterých se při hledání vhodných informací můžeme dopustit. Budeme hledat a analyzovat různé souvislosti, které by nám dopomohli odpovědět na naše otázky. Dále zjistíme jak nejlépe zajistit bezpečnost dat, poznat důvěryhodnost informací a správně se rozhodnout, které „dveře“ ve světě Internetu otevřeme.

2 INFORMACE, DATA, ZNALOSTI

Jaký bude předpokládaný vývoj akcií dané firmy? Jakou strategii zvolíme při zajišťování sanace firmy? Copak se to dnes stalo? Těmito i dalšími otázkami se zabýváme den co den. Ale aby si každý mohl na jakoukoli otázku odpovědět, potřebuje jednu velmi důležitou věc - **informace**. Informace potřebujeme pro provozování nejrůznějších aktivit každodenního života, počínaje základními biologickými potřebami a konče řešením náročných systémových projektů. Plno informací se nachází kolem nás, jsou okamžitě a snadno dostupné, např. dopravní značka přikazující směr jízdy; jiné informace nejsou tak snadno přístupné a vyžadují po jedinci, aby vyvinul určité úsilí a provedl rozsáhlé vyhledávání v různých zdrojích.

Přestože jsou informace tak staré jako lidstvo samo, nyní, v době „informační společnosti“, dochází k ještě většímu růstu. Nejen, že roste počet, ale zvyšuje se i význam. S rozvojem Internetu a jeho další expanzí nás zahlcují nejrůznější dokumenty, obrázky a soubory, nesoucí v sobě nějakou informaci (resp. data). Proto také větu – „Našel jsem to na internetu“, slýcháme čím dál častěji. Informační prostor internetu vzrůstá každým okamžikem, připojují se komerční organizace i jednotlivci, vznikají stále nově dostupné zdroje. Jiným slovem se dá říci, že lze na Internetu najít odpověď na jakoukoli otázku.

Zde se nachází hlavní kámen úrazu. Není všechno zlato, co se třpytí aneb ne každá informace, musí být nutně informací potřebnou nebo pravdivou… zvláště na Internetu. Jelikož do internetové sítě může dát kdokoliv cokoliv, často se stane, že narazíte na bezcenné bláboly, který je zcela vymyšlený, zastaralý nebo pochází z krajně nedůvěryhodného zdroje.

2.1 INFORMACE

Jak již bylo řečeno, pojem **informace** zaznamenáváme všude kolem nás. Existuje v mnoha disciplínách i oborech. Proto ji vyjádřit jednou definicí se jeví jako krajně nevhodné, neboť by to nezvážilo její mnohotvárnost. Nyní si uvedeme nejdůležitější příklady definic:

1. „laický“, každodenní“ pohled na informaci
   - V tomto pohledu můžeme informaci pojmenovat jako *zprávu* nebo *sdělení*.
   - „*Jazykový projev vybudovaný na principu informačního slohového postupu, ve kterém se co nejobjektivněji věcně a dokumentaristicky konstatují určitá fakta.*“ [viz. http://info.sks.cz/users/ku/UIS/inform1.htm]
2. filozofické pojetí
   - „*Vlastnost hmotné reality být uspořádán a její schopnost uspořádat (forma existence hmoty vedle prostoru, času a pohybu.*“ [ZEMAN; viz. http://info.sks.cz/users/ku/UIS/inform1.htm]
   - „*Vnímatelný obsah poznaného nebo předpokládaného obrazu skutečnosti, který je možno využít pro člověka.* “ [viz. http://info.sks.cz/users/ku/UIS/inform1.htm]
3. komunikační pojetí
   - Zde lze chápat jako obsah komunikace, přenos zpráv a oznámení osobním kontaktem, zvukem, signálem nebo prostředky masové komunikace
   - „Každý *znakový projev, který má smysl pro komunikátora i příjemce.*“ [LAMSER; viz. http://info.sks.cz/users/ku/UIS/inform1.htm]
   - „*Objektivní obsah komunikace mezi souvisejícími hmotnými objekty, projevující se změnou stavu těchto objektů.*“ [BRILLOUIN; viz. http://info.sks.cz/users/ku/UIS/inform1.htm]
4. kybernetické pojetí
   - „*Název pro obsah toho, co se vymění s vnějším světem, když se mu přizpůsobujeme a působíme na něj svým přizpůsobováním. Proces přijímání a využívání informace je procesem našeho přizpůsobování k nahodilostem vnějšího prostředí a aktivního života v tomto prostředí.*“ [WIENER, viz. http://info.sks.cz/users/ku/UIS/inform1.htm]

- Obecněji lze napsat že získání informace je část poznání, která slouží k orientaci, k rozhodování s cílem zachovat kvalitativní systém a ten pak dále zlepšovat a rozvíjet.
- Proces, kdy jeden určitý systém předává zprávu jinému, pomocí určených signálů, která nějakým způsobem mění stav přijímacího systému.

5. matematický přístup k informací
   - Zde se jedná o energetickou veličinu, jejíchž hodnota je úměrná zmenšení entropie[1] systému
   - „*Poznatek, který omezuje nebo odstraňuje nejistotu týkající se výskytu určitého jevu z dané množiny možných jevů.*“ [viz. http://info.sks.cz/users/ku/UIS/inform1.htm]
6. sémiotické pojetí
   - sémiotika je obor, zabývající se interpretací znaků a jejich teorií
   - „*Informaci můžeme chápat jako statistickou pravděpodobnost určitého signálu či znaku, který je na vstupu stanoveného systému. Čím menší je pravděpodobnost daného znaku, tím větší má takzvanou informační hodnotu.*“
     [SHANNON; SA_320, Informační systémy, přednáška č. 5]
   - Shannonovo sémiotické pojetí zasahuje k matematicko/statistické teorii.

Kromě vymezení podle předcházejících pojetí, byla stanovena další měřítka pro zhodnocení informace. Například podle České terminologické databáze knihovnictví a informační vědy (dále jen TDKIV), která informaci chápe jako údaj o reálném prostředí, o jeho stavu a procesech v něm probíhajících. Informace snižuje nebo odstraňuje neurčitost systému (např. příjemce informace). TDKIV dále popisuje, že množství informace je dáno rozdílem mezi stavem neurčitostí systému (entropie), kterou měl systém před přijetím a stavem neurčitosti, která se doručením informace odstranila.

V tomto smyslu můžeme informaci považovat jak za vlastnost organizované hmoty, vyjadřující její hloubkovou strukturu, tak za produkt poznání fixovaný ve znakové podobě v informačních nosičích.

Jak již bylo řečeno, s informacemi pracují lidé v mnoha disciplínách a oborech. V exaktní disciplíně se za informaci považuje sdělení, které vyhovuje přísným kriteriím logiky či příslušné vědy. Pro ekonomy zpravidla bývá informace zpráva nebo sdělení sloužící k dosažení zisku. Z informační koncepce, která je pro nás nejdůležitější, může být definice formulována takto: „*Informace je podmnožina poznatků, která je použita v konkrétní situaci pro řešení problémů*.“
[KUHLEN; SKLENÁK, Vilem, Data, informace, znalosti a Internet, Praha 2001, str. 3]

Z hlediska vyhledávacího procesu se dá říci, že informace jsou data (viz. dále), která obsahují souvislou, srozumitelnou a použitelnou formu. Jinými slovy to jsou data, jež v určitém kontextu začínají dávat smysl a jejíchž výsledkem jsou zodpovězené otázky. Kdybychom měli jít ještě dále a pokusit se zobecnit výše zmíněné podklady, dostaneme se k vymezení informace, která se stává jádrem základní myšlenky - „***Informace je význam, který člověk přisuzuje údajům.***“ [ČSN 369001; SA_320, Informační systémy, přednáška č. 5] a „***Význam, který člověk přiděluje datům prostřednictvím zvyklostí používaných na jejich prezentaci.***“ [BS 3527; SA_320, Informační systémy, přednáška č. 5].

2.2 HODNOTA INFORMACE

Každá informace může mít význam pro určitou osobu, ale pro osobu jinou toto sdělení nemusí znamenat žádný přínos. Čím to je? Hodnota informace má subjektivní charakter, neboť je součástí transformace dat. Proto je ve své podstatě závislá i na uživateli. Pro jednoho a téhož příjemce je zpráva informací pouze v určité situaci. Každá zpráva je tedy potencionální informace, ale informací se stává teprve ve vhodném kontextu. Tato hodnota je potom důležitá pro zajištění správného rozhodování a následné kredibility.

[1] neurčitost, nejistota, neuspořádanost; střední hodnota míry informace potřebné k odstranění neurčitosti, která je dána konečným počtem vzájemně se vylučujících jevů

2.3 DATA

Většina lidí se domnívá, že všude kolem nás jsou informace. I když jsem tuto myšlenku na začátku zmínil, je třeba ji trochu poopravit. Informace sice v našem prostředí existují, ale pouze pro subjekty, jejichž hodnota není nulová. Ostatní subjekty mohou „informaci" interpretovat jako data. Všechno, co vnímáme a nejsou to smysluplné informace, jsou data.

Data lze z hlediska computer science chápat jako čísla, text, zvuk, obraz, popř. jiné smyslové vjemy reprezentované ve vhodné podobě pro zpracování počítače. Z tohoto hlediska je rozdělujeme na strukturovaná a nestrukturovaná data.

Strukturovaná data, která se například nacházejí v relačních databázových systémech, explicitně vyjadřují fakta, objekty, atributy. Dělí se na textová, číselná, časová a na logická (zaznamenání toho, zda je či není splněna nějaká podmínka, případně zda má či nemá daný objekt jednu konkrétní vlastnost) nebo kategorické (hodnota vlastnosti, která je vybraná z nějaké škály). Výhodou je, že se strukturovanými daty je jednoduchá manipulace a snadno se vyberou jenom ta data, která jsou vhodná např. pro řešení informačního problému.

Do kategorie nestrukturovaných dat patří volný text, audio, video, grafika. Jejich velkou předností je, že poskytují více informací než strohé strukturované údaje. Jejich použití v IS však má jednu slabinu, kterou je pro efektivní využití nutno odstraňovat. Problém spočívá v těžkosti vyhledávání nejenom na Internetu. Jediné zatím dostupné řešení je v doplnění o data strukturovaná (vyhledávácího klíče).

2.4 ZNALOSTI

S informacemi neodmyslitelně souvisejí znalosti. Slouží nejenom k posuzování dat a přiřazování jim informační hodnoty, ale taky tímto způsobem sami sebe neustále upravují. Můžeme říci, že znalosti jsou zobecněním informací v paměti člověka, která slouží k posuzování a interpretaci dat budoucích. Pomocí interpretace člověk následně provádí kognitivní operace[2]. Na základně těchto operací dokáže člověk například předvídat, co se může v reálném světě stát a podle toho upravovat své cíle.

Chackland a Scholes dobře vyjadřují vzájemnou souvislost a podmínečnost dat, informací a znalostí: „*Technologie pracují s daty, lidé je interpretují jako informace nesoucí význam, které se stávají podnětem pro další jednání. Proces interpretace je kognitivní záležitost, ve kterém stěžejní roli hrají znalosti.*" [SKLENÁK, Vilem, Data, informace, znalosti a Internet, Praha 2001, str. 4]

Zajímavé je rozdělení znalostí na explicitní (vyjádřenou nebo vnější)/implicitní (vnitřní nebo skrytou) nebo na individuální (unikátní)/skupinovou (sdílenou).

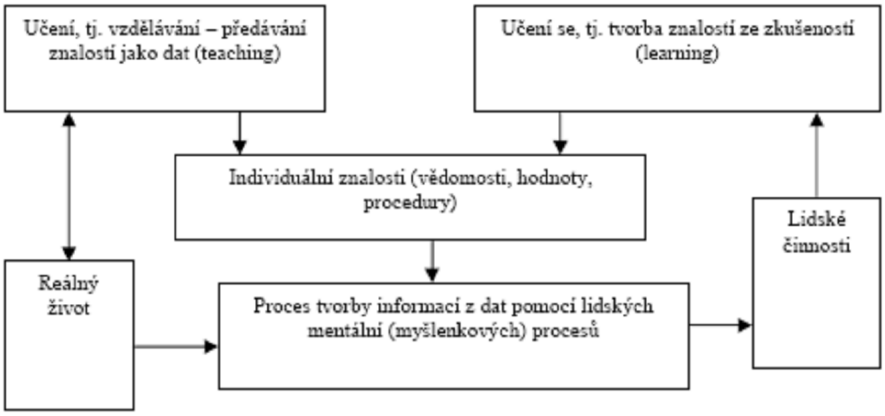
ZNALOST:	***individuální** / unikátní*	***skupinová** / sdílená*
*vyjádřená / vnější - **explicitní***	**Koncepty** (představy), vyjádřené vhodnou formou	Společně sdílené **modely a příběhy** (narativní modely)
*vnitřní / skrytá - **implicitní** (tacit)*	Vlastní **zkušenosti, schopnosti a hodnoty** (interakce s okolím)	Konkrétní **kultura** a/nebo **styl práce** a komunikace

Tab. 1: Základní rozdělení znalostí

Za zmínku dále stojí, že *implicitní* znalost je ze své biologické povahy vázána na konkrétního člověka. Je *individuální*, tzn. není ani objektivní, ani subjektivní, a *konceptuální*, tzn. že je úzce propojena

[2] Tím je řečeno „myšlení" a „vnímání". Znalost formuje a přetváří určité vzory a provazuje významné situace, vhodná řešení a váže je (i částečně) se slovy nebo symboly.

s jazykovým, resp. symbolickým vyjádřením a uvažováním. Bez přirozené schopnosti používat jazyk, se s informacemi a tedy i se znalostmi nedá pracovat.



Obr. 1: Tvorba znalosti

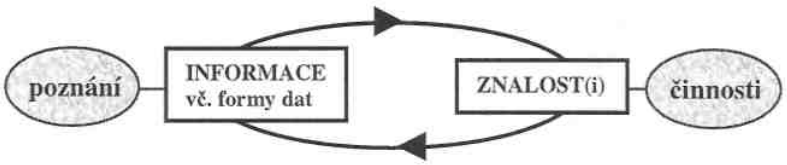
2.5 VZTAH DATA, ZNALOSTI, INFORMACE

Každému je určitě známý vztah mezi těmito třemi entitami. Jednoduše lze tento vztah napsat jako přímku, na které leží data, informace a znalosti. Velice podobný pohled mají na věc prof. Checkland, který do klasického modelu používá navíc capta (zachycená data), a prof. Ackoff, který dále zahrnuje pojem moudrost (wisdom). Všeobecně řečeno na Internetu se probíráme datovými zdroji, z nichž vybíráme, ty které mají velkou hodnotu. Z těch pak sestavujeme informace a výsledek „vložíme“ do paměti jako znalost pro budoucí potřebu. V moderní době se ale tato představa považuje za mylnou z důvodu dřívější neznalosti daných subjektů.

Obr. 2: Informace uvažována v lineárním (kauzálním) řetězci podle prof. Checklanda a prof. Ackoffa

V realitě je to však trochu složitější. Na následujících obrázcích je znázorněna asi nejužívanější vzájemná spojitost mezi objekty.

Obr. 3: Vztah mezi entitami



Obr. 4: Kruhový vztah informace a znalosti: Data jsou jednou z forem informace, která je výsledkem poznání a znalosti, kterou zároveň (re)formuje

3 INFORMAČNÍ ZDROJE

V předcházející kapitole jsme si řekli vše podstatné o informaci a její důležitosti, která pro nás znamená. Nyní si můžete myslet, že pokud je informací v moderním světě nadměrné množství, není žádný problém je sehnat. Omyl. Jsou dvě důležité věci, které by se neměly podceňovat. První: najít informaci ve správném kontextu a druhá: vědět, kde přesně hledat. Ono znát, jakou informaci potřebujeme, neznamená, že ji i získáme. V této kapitolce si ukážeme, kde se informace skrývají a jak se k nim lze dostat.

I když jsou informační zdroje velmi rozsáhlé, přesné definování je obtížné. Mohli bychom je chápat jako systém, který je reálným nebo potencionálním šiřitelem a zprostředkovatelem informací, což může být internet, televize, informační centra nebo knihovny.

Informační zdroje se mohou například dělit podle typu informace (zdroje). Na ekonomické zdroje, právnické zdroje, sociální zdroje apod.; nebo podle zdroje poskytující textové, obrazové, zvukové, a smíšené informace. Jiný pohled skrývá členění podle dostupnosti na veřejné, komerční a utajované.

Mezi nejdůležitější zdroje patří knihovny a specializované firmy zabývající se obchodní činností v oblasti poskytování informací. Dalším velkým zdrojem, který spojuje i předcházející subjekty je Internet. Je to dáno tím, že většina knihoven i specializovaných firem poskytují právě své služby přes Internet.

3.1 HODNOCENÍ INFORMAČNÍHO ZDROJE

Jelikož je hledání správného informačního zdroje náročná práce, je vhodné si nejdříve zjistit jeho kvalitu. Pokud hovoříme o hledání v rámci Internetu, je tím myšleno třeba právě kvalita knihoven, firem, databázových center, online rubrik apod. Uveďme příklad: Například jednoduché je zjistit kontakt na určitou firmu při hledání práce. Naopak již složitější je jako pracovník banky zjistit vhodnost dané firmy k poskytnutí úvěru. Zde se už nevypořádáme jenom s jednou informací, ale budeme muset využít více vzájemně nezávislé zdroje. Jinak řečeno, musíme si být jisti, že zdroj, odkud čerpáme informace je kvalitní. Ale pozor, i kvalitní zdroje občas mohou poskytovat nepoužitelné informace. I přesto je jeden z hlavních postupů, jak nabýt důvěryhodných informací. Při hodnocení informačního zdroje se berou v úvahu hlavně tyto charakteristiky:

- **Typ informací** udává, jestli jde o informaci bibliografickou nebo o úplný text dokumentu (sekundární nebo primární informace). Dále zdali je v obrazovém, zvukovém, textovém či jiném formátu.

- **Rozsah** uvádí, kolik záznamů zdroj obsahuje. Pokud je služba placena, tak čím dražší, tím obsáhlejší záznamy.

- **Úplnost** stanovuje, kolik ze všech dostupných informací, jimiž se zdroj zabývá, je ve něm uloženo.

- **Retrospektiva** ukazuje, jak daleko do minulosti uchovávané informace sahají. Při analýze časových řad velmi výhodné.

- **Perioda aktualizace** udává, v jakém intervalu se přidávají nové informace.

- **Producent** spoluurčuje důvěryhodnost zdroje. Největší důležitost se zajišťuje například u databází – čím známější, tím vyšší cena.

- **Dostupnost** říká, zda je zdroj volně dostupný nebo se za něj platí (komerční zdroj).

- **Cenu** udává kvalita informací nebo renomovanější producent. Platí to pouze u komerčních zdrojů. Způsob platby může být rozdílný - od paušální přes platby na časový úsek.

3.2 INTERNET

Toto slovo zná v dnešním moderním světě asi každý. Drtivá většina lidí, kteří potřebovali najít určitou informaci, zažili sílu Internetu. Často můžeme při jízdě metrem či tramvají zaslechnout útržek rozhovoru: „…perfektní stránky, člověče, úplně všechno jsem tam našel, asi jim zamejluju…“ Internet hýbe světem. To je bez diskuze. I když jeho informační prostor je obrovský, z průzkumů vyplývá, že v posledním půlroce se počet připojených stanic zdvojnásobuje každé dva měsíce. Ale co vůbec internet je? Kde a jak vznikl?

3.2.1 HISTORIE

Snad každá rubrika o historii Internetu začíná slovy: „…na počátku byl ARPANET.“ A má pravdu. ARPANET byl zárodečnou sítí, kterou v USA vytvořili vojáci. Přesnější bude, že ARPANET vytvořili lidé z akademické sféry za peníze vojáků, které tekly grantovou agenturou ARPA (Advanced Research Projects Agency). Cílem bylo ověřit, zda je možné vybudovat velmi robustní počítačovou síť, která by byla schopná přežít útok nepřítele. Kdysi jenom málokdo mohl tušit, v co vyroste ARPANET.

„Are you receiving this?“ byla první věta, která byla v srpnu 1969 poslána z University of Kalifornia v Los Angeles po síti složené ze čtyř uzlů: UCLA, Stanford Research Institute a UC Santa Barbara a University of Utah v Salt Lake City. Do roku 1971 bylo napojeno ještě dalších 11 univerzit.

Rozhodující význam pro sítě s přepojováním paketů měl vývoj sady protokolů TCP/IP, jehož tvůrci byli Vinton Cerf a Robert Kahn. Na tvorbě, která byla zahájena již v roce 1973 se podílela i řada jiných osobností. První testy se začali uskutečňovat v roce 1975. V roce 1977 proběhla demonstrace práce sítě ARPANET pod řízením internetových protokolů. Hlavní protokoly Internetu TCP/IP se začali s konečnou platností používat od roku 1983. Poté začala ARPANET využívat celá akademická sféra. Různé školské organizace, které měli své sítě, se pomalu připojovali k ARPANETu, který stále financovali vojáci. ARPANET se stal jakousi páteřní sítí, na kterou se nabalovaly další a další sítě. Výsledkem byla soustava vzájemně propojených sítí (v angličtině internetwork; zkráceně INTERNET).

Vývoj protokolů znamenal pro Internet veliký boom, když z přibližně tisíce připojených počítačů v roce 1983 vzrostl počet v roce 1992 na více než milion. Za následek měla expanzi mimo americký kontinent. Od roku 1986 začíná být Internet financován akademickou sférou skrze grantovou agenturu NSF (National Science Foundation), který následně zárodečný ARPANET nahrazuje páteřní sítí NSFNET. NSF určuje podmínky pro připojení jenom pro akademické instituce. Rychlost připojení je stanovena na 56 kb/s. Jelikož existuje mnoho propojených počítačů, musel se řešit problém efektivního přidělování jmen. Proto byl v roce 1984 zaveden doménový systém DNS (Domain Name Services).

Počátkem roku 1989 se na půdě ústavu částicové fyziky CERN objevil dokument HyperText a CERN, jenž popisoval možnosti vytvoření interního distribuovaného systému jako jednotné nadstavby nad mnoha různorodými informačními zdroji. Autorem dokumentu byl Tim Berners-Lee, který později v listopadu roku 1990 předvedl první prototyp WWW serveru. Operačním systémem byl zvolen NeXT.

Prudký nárůst obliby WWW se datuje k září 1993, kdy byla dostupná první funkční verze velmi populárního prohlížeče NCSA Mosaic. Do toho roku zůstával Internet doménou především vědeckých a akademických pracovišť. Situace se ale (již) začala měnit v roce 1991, kdy americký kongres přijal zákon High Performance Computing Act, jehož iniciátorem byl senátor Al Gore. Od roku 1994 se na Internetu začaly objevovat komerční organizace. Nejprve informační, posléze i jiných oborů lidské společnosti. V mnoha státech se Internet stával běžnou součástí každodenního života.

V druhé polovině 90. let dochází k dalšímu rozmachu Internetu a jeho komercializace. Firma, která sama sebe neprezentuje na Internetu, jakoby neexistovala. Od roku 1995 jsou postupně americkou vládou budované financované sítě v BNS, které jsou spojovány pěti superpočítačovými centry. Nejpoužívanější službou se stává elektronická pošta. Objevují se nové možnosti připojení k internetu přes satelit, wi-fi a kabel. Dále dochází k zavádění Internetu do domácností. Uživatel může z pohodlí domova číst denní tisk, prostřednictvím virtuálního obchodního domu nakupovat nebo provádět bankovní transakce.

3.2.2 DEFINICE

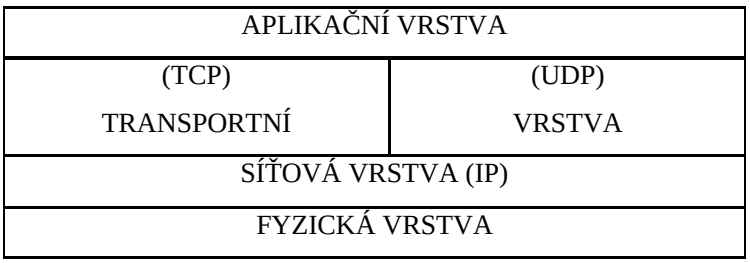
Nyní už známe cestu, kterou se Internet prodíral. Ale jak ho lze popsat nebo definovat? Internet můžeme charakterizovat jako celosvětový informační systém, který je logicky propojen do jednoho celku prostřednictvím globálního adresného prostoru, jenž je založený na protokolu IP (Internet Protocol) nebo jeho následných rozšíření. Dále je schopen podporovat komunikaci prostřednictvím rodiny protokolů TCP (Transmission Control Protocol) nebo jeho nástupcích, eventuálně jiných protokolů kompatibilních s IP. Rovněž nabízí veřejně nebo privátně dostupné služby vyšší úrovně, které jsou založeny na komunikační a další infrastruktuře.
[SKLENÁK, Vilem, Data, informace, znalosti a Internet, Praha 2001, str. 183]

3.2.3 ARCHITEKTURA A PROTOKOLY

Co se týče architektury, dnešní internet se skládá ze tří vrstev.

- síťová vrstva používá IP protokol, který „drží“ Internet
- transportní vrstva používající protokoly TCP a UDP se přizpůsobuje potřebám aplikace
- aplikační vrstva používající protokoly konkrétních aplikací, které se starají o chod programu

V mnoha publikací lze nalézt ještě fyzickou vrstvu. Ta je z hlediska Internetu zodpovědná za doručení v rámci lokální sítě a může být dvou bodová (např. telefonní spojení) nebo síť více účastníků (Ethernet).



Obr. 5: Architektura internetu

Co to jsou vlastně protokoly? Jsou to pravidla, která vymezují postupy a parametry používající se při vysílání a přijmu dat. Například jde o definici formátů, chybové kontroly aj.

Komunikace probíhá tak, že stanice a servery si mezi sebou posílají zvláštní zprávy, jejichž obsah může být různý (datové soubory, řídící informace). Tyto zprávy se vysílají po kratších oddílech, které jim počítač stanoví podle pevně dané délky. Dlouhé zprávy jsou tedy rozděleny na více částí, velmi krátké zprávy jsou naopak doplněny (padding). Jednotlivé úseky zpráv nazýváme pakety (packet). Paket, který ve své hlavičce obsahuje informace o zdroji a je před odesláním doplněn o údaje cílové a zdrojové adresy, posloupnosti a kontrolního součtu správnosti dat, se nazývá rámec (frame). Když pakety dosáhnou cílové stanice, jsou spojeny do původní podoby souboru.

Soustava protokolů TCP/IP je v současnosti nejpoužívanějším a nejrenomovanějším síťovým protokolem. Použití je všestranné. Nemá problémy v síti Internetu, domácích sítích nebo na platformách Novellu, Microsoftu či Linuxu.

- IP (Internet Protocol) pracuje na síťové vrstvě modelu ISO/OSI a je protokolem spojově neorientovaným. Přijímá datové segmenty a přidává do nich svoji hlavičku, čímž vytváří IP datagram. Dále odesílá data na adresy v něm obsažené. Ale zda data přišla v pořádku, nekontroluje.

- TCP (Transmission Control Protocol) doplňuje IP. neboť (Protože) umí kontrolovat, zda jsou došlá data správná. Pokud by tedy existoval pouze protokol IP, byla by komunikace velice nespolehlivá.

Služby protokolu TCP		
Port	Služba	Užití
20	FTP (File Transfer Protocol)	určen pro přenos dat, datový kanál
21	FTP (File Transfer Protocol)	určen pro přenos dat, řídící kanál
23	Telnet (Telecomunication Network)	terminálové relace, tedy ovládání serveru ze vzdáleného počítače
22	SSH (Secure Shell)	Sada programů pro vzdálené (terminálové) přihlašování k počítačům. Zabezpečená varianta k Telnetu.
25	SMTP (Simple Mail Transfer Protocol)	odesílání e-mailových zpráv přes e-mailový klient
53	DNS (Domain Name System)	stará se o překlad IP adres na jmenné adresy a naopak
80	HTTP (Hyper Text Transfer Protocol)	prohlížení webových stránek
110	POP3 (Post Office Protocol)	stahování e-mailových zpráv ze serveru na lokální PC
123	NTP (Network Time Protocol)	synchronizace času na lokální stanici s internetovým serverem
143	IMAP 4 (Internet Message Access Protocol 4)	umožňuje manipulaci s e-maily přímo na serveru poskytovatele, bez nutnosti stahování
220	IMAP 3 (Internet Message Access Protocol 3)	to samé jako IMAP 4
443	HTTPS (HTTP Secure)	Šifrovaná (SSL) verze protokolu HTTP
990 989	FTPS	zabezpečená verze protokolu FTP
993	IMAPS	zabezpečená verze protokolu IMAP
995	POP3S	zabezpečená verze protokolu POP3
1863	MSN Messenger	komunikační protokol
5190	ICQ (I Seek You)	dnes velmi používaný komunikační protokol

Tab. 2: Nejpoužívanější služby protokolu TCP

- UDP (User Datagram Protocol) má shodné vlastnosti jako protokol TCP. Liší se pouze tím, že nenavazuje relaci a nekontroluje správnost došlých datagramů. To má za následek větší rychlost a menší náročnost na hardware. Jeho slabinou je ovšem relativně vysoká nespolehlivost.

Služby protokolu UDP		
Port	Služba	Užití
53	DNS	stejný účel jako u protokolu TCP
69	TFTP (Trivial FTP)	určen pro přenos krátkých souborů
137	NetBIOS	v současnosti není příliš využíván a je doporučeno jej uzavírat ve firewallu
161	SNMP (Simple Network Management Protocol)	používán pro řízení rozlehlejších sítí

520	RIP (Routing Information Protocol)	používán pro řízení paketů
2840	Windows Messenger	komunikační program

Tab. 3: Nejčastější služby protokolu UDP

3.2.4 INFORMAČNÍ SLUŽBY INTERNETU

3.2.4.1 Telnet

Telnet je nejstarší základní službou Internetu. Vznik je datován před začátkem prosazování osobních počítačů. Službu používali jen střediskové počítače v rámci organizací, firem a škol, pro které byly vybudovány terminálové sítě. Terminál pak sloužil jako koncové zařízení, pomocí nichž mohli pracovat samotní uživatelé. Terminál tehdy představoval fyzické zařízení, které se skládalo z klávesnice a monitoru a nějakého prvku, který umožňoval uživateli vstup do systému.

V současné době se vstupním terminálem může stát jakýkoliv osobní počítač, který je ho schopný pomocí programu emulovat. Hlavním důvodem proč služba Telnet pracuje pouze v textovém režimu je minimalizace objemu dat přenášených mezi hostitelským počítačem a terminálem. Proto chybí prvky grafického uživatelského rozhraní, jako jsou okna, ikony apod. Uživatel se musí obejít i bez myši.

V minulosti představoval Telnet jedinou možnost, jak uživatelé mohli pracovat s informačními zdroji. U veřejných zdrojů se uživatel dozvěděl veřejné jméno pro přihlášení. Zatímco u komerčních služeb si musel založit účet, který byl jištěný například heslem. Dnes naprostá většina zdrojů přešla ke službě WWW, ale stále existují takové, které poskytují služby prostřednictvím Telnetu (archivy, knihovny, diskusní fóra aj.).

3.2.4.2 Gopher

Služba Gopher vznikla koncem 80. let na vysoké škole v americké Minnesotě. Šlo o nejpopulárnější službu před nástupem WWW pro svou snadnou lokalizaci informací v prostředí Internetu. Gopher představoval první službu navigačního typu, umožňující přistupovat k informačním zdrojům, aniž by uživatel musel předem znát jejich konkrétní umístnění. Jeho myšlenka znamenala, že místo výpisu jmen v adresáři bylo uživateli nabídnuto menu, kde se na jednom řádku slovně uvádělo, co daný soubor obsahuje, případně jaké je zaměření dalšího podadresáře. Uživatel pak získal soubor tak, že aktivizoval odkaz, který se za řádkem skrýval, případně se tímto způsobem dostal na další menu popisující soubory v adresáři. Nevýhoda hierarchických nabídek, na jejíž principech Gopher pracoval, byla komplexnost při mnoha úrovních. Pomocí odkazů z nabídek se uživatel mohl pohybovat jak v rámci jednoho serveru, tak i navštívit externí servery. Těmito externími odkazy se vlastně vytvářel Gopher-prostor, jehož hranice byly dány jen objemem Internetu. Gopher představoval velikou pomoc při potřebách uživatelů soustředit se na cíl svého vyhledávání.

3.2.4.3 FTP (File Transfer Program, File Transfer Protocol)

Na Internetu se nachází plno uzlů s programy, obrázky, dokumenty a jinými soubory. K těmto uzlům se lze připojit a pohybovat se v nich jako na lokálním disku. Tyto data se dají lehce stáhnout. K tomuto účelu se používá služba FTP. Ta je určena pro přenos dat mezi počítači, které mohou být provozovány pod různými operačními systémy. Služba FTP patří mezi nejstarší aplikační protokoly, ale stále je velice využívaný. Hlavní rozdělení služby se dělí na anonymní FTP a identifikovatelné FTP.

3.2.4.4 World Wide Web (WWW)

> *„World Wide Web je služba, která zpřístupňuje v prostředí Internetu hypertextové[3] dokumenty a je založena na architektuře klient/server[4].“*
> [SKLENÁK, Vilem, Data, informace, znalosti a Internet, Praha 2001, str. 188]

V roce 1989 se v rámci organizace CERN zrodil požadavek na kvalitní a snadno ovladatelný systém, který bude přístupný všem a s jehož pomocí se zrychlí práce s Internetem. O rok později poznal světlo světa systém, který dnes patří spolu s elektronickou poštou k nejrozšířenější službě. Nejenom, že systém nabízel možnost kombinace textu s obrázky, ale i vzájemné propojení jednotlivých textů nebo jejich částí pomocí hypertextových odkazů. První verze umožňovala kromě prohlížení i editaci dokumentů.

Velký rozmach WWW začal od roku 1993-5, když byly vypuštěny první grafické prohlížeče Netscape Navigator a Internet Explorer. Příjemné uživatelské rozhraní zjednodušilo nejen práci s hypertextovými dokumenty, ale také s dalšími službami. Prohlížeč umožňoval přistupovat i k serverům pomocí protokolů FTP nebo Gopher. Díky grafickému rozhraní mohli začít Internet používat i lidé bez technických znalostí.

Služba WWW má nekonečné obzory využití. Možná proto představuje pro mnohé uživatele synonymum k Internetu. Je to ale mylná domněnka. Důvodem, proč zůstává služba WWW stále atraktivní, je řada faktorů:

- Stálá možnost vkládání obrázků, tabulek, používání barev, vzorů, animací a jejich formátování má za následek neustálé obohacování prezentačních schopností.
- Služba WWW zaručuje zjednodušení a zvýšení rychlosti při vyhledávání a následné práce s informacemi.
- Služba se stala interaktivní a dynamickou knihovnou, která slouží jako oboustranná vazba. Tzn. že tok dat směřuje nejenom od zdroje k zákazníkovi, ale i naopak.
- Interaktivita služby vyvolala další trend, který využívá WWW prohlížečů jako univerzální klientské platformy. Prostřednictvím prohlížečů jde s dokumenty pracovat i mimo Internet.

Mezi základní kameny služby WWW patří hypertext, jazyk HTML, URL a protokol HTTP. Bližší popis si ale uvádět nebudeme.

3.2.4.5 Elektronická pošta (e-mail)

Elektronická pošta neboli e-mail poskytuje základní spojení ve světě Internetu. Každý, kdo je připojen k síti, má svou elektronickou adresu, na kterou mu lze posílat elektronické dopisy. Elektronická pošta je nejvyužívanější službou na Internetu i v lokální nebo WAN [5]sítích.

E-mail, je jedinou funkcí, kde hlavní přenos probíhá v pozadí. Uživatel obvykle jen přímo napíše text dopisu a ten je pak v pozadí v Internetu dopraven až k adresátovi. Ovšem interaktivní spolupráci bezprostředně dvou spolu komunikujících programů se nevyhneme. Vždy spolu pracují dva automatické programy. Způsob jejich spolupráce popisuje protokol SMTP[6], který se stará o přijímání a odesílání zpráv.

[3] Hypertext – Nelineární uspořádání textu, ve kterém jsou jednotlivé části navzájem propojeny hypertextovými vazbami.
[4] Model klient/server je určitá forma distribuovaného zpracování, kdy jeden program (klient) komunikuje s jiným programem (server) za účelem výměny informací.
[5] WAN (Wide Area Network) – Označení rozsáhlé počítačové sítě; na vzdálenost mezi jednotlivými uzly sítě nejsou kladena žádná omezení. V síti WAN patří mezi nejvýznamnějšími uzly páteřní linky.
[6] Simple Mail Transfer Protocol

Elektronická pošta je sice nejstarší, ale přesto stále zůstává nejpoužívanější službou Internetu určeno pro komunikaci. Vznikala ještě v době ARPANETu. Od té doby prošla mnoha úpravami a vylepšeními. Dnes lze e-mailem poslat skoro vše, od klasických zpráv, souborů, dokumentů počínaje, hotovými programy konče.

3.2.4.6 Elektronické konference a diskusní skupiny

Na rozdíl od elektronické pošty, kdy dochází s jiným účastníkem sítě ke komunikaci typu 1:1, u elektronické konference dochází k výměně názorů s více lidmi najednou – tedy M:N. Hlavní účelem bylo diskutovat s lidmi se stejnými názory, zájmy nebo problémy. Přesný počet se dá jen těžko odhadovat, protože existují jak celosvětové konference, tak konference s omezenou působností. Témata konferencí sahají od debat o obecných věcech až po obsahově zaměřené na využívání konkrétních programových produktů.

Z hlediska otevřenosti rozlišujeme konference otevřené nebo uzavřené, podle volnosti probíhající diskuse na konference nemoderované nebo moderované.

Jejich nevýhody vyplývají z nutnosti zjišťovaní adres, potřeby uskutečnit administrativní úkony pro přihlášení do konferencí, z řešení problémů odběru při dlouhodobé nepřítomnosti, z nebezpečí zahlcení uživatele při současném odebírání více konferencí nebo výskytů flamewaru[7], které uživatele obtěžují.

Některé z uvedených nedostatků se snaží odstranit služby, pro něž používáme název diskusní skupiny. Tato služba má naprosto stejný účel jako u elektronické konference, ulehčit komunikaci mezi více lidmi. Umožňuje uživateli přečíst reakce z mnoha diskusních skupin a následně se k nějakému problému vyjádřit. Základním kamenem je diskusní skupina (newsgroup), která je stanovena daným problémem nebo okruhem. V ní jsou pak soubory představující názory od jednotlivých uživatelů.

3.2.5 INTERNET JAKO ZDROJ INFORMACÍ

Na Internet se dá dívat z různých úhlů pohledu. Třeba jako gigantické seskupení počítačů, soubor programů, soubor online zdrojů nebo jej můžeme chápat jako celosvětovou elektronickou knihovnu. Každá z těchto možností je částečně pravdivá. Uživatel vždy bude chápat Internet podle aktuálního způsobu používání. Ale všechny tyto úhly jsou velmi důležité pro pomoc při hledání informací a uživatel by si měl mezi nimi umět vybírat, protože jednostranný pohled zbytečně ztěžuje vyhledávání. Někteří uživatelé tvrdohlavě chápou Internet jako knihovnu. Používají vyhledávací stroje jako ekvivalent katalogových lístků. Pokud ale dostali odpověď od vyhledávacího stroje „nothing found“, dojdou k mylnému závěru, že daná informace neexistuje a přestanou dále hledat. Jestliže uživatel převážně používá vyhledávací služby, může také nabýt dojmu, že Internet je organizován a že všechny dostupné informace jsou seřazeny jako ve škatulkách. Tak jednoduché to bohužel není.

3.2.5.1 Internet jako seskupení počítačů

Při zrodu Internetu se lidé dívali na něj hlavně jako na souhrn počítačů, které tvoří síť. Kdo chtěl najít informaci k požadovanému tématu, musel znát adresu počítače. Tento úhel má ale nevýhodu. Aby člověk mohl efektivně pracovat s informacemi, bylo zapotřebí si zapamatovat množství jmen počítačů. Proto se v této době začali vytvářet seznamy počítačů s uvedením jejich obsahového zaměření. Přes veškerou snahu však hledání zůstalo obtížné a zdlouhavé. Pro snadnější zapamatování se s úspěchem začala používat tzv. doménová jména. Uživatel se může lehce dovtípit, jakou adresu zadat, aby se dostal na požadované místo. Například na navštívení firmy Microsoft stačilo zadat www.microsoft.cz. Uvedená metoda není univerzální a občas se může stát, že z různých důvodů selže.

[7] Série zpráv, které mají charakter stupňující se hádky mezi dvěma nebo i více uživateli.

3.2.5.2 Internet jako soubor programů

Plno knížek, učebnic či encyklopedií nás učí, jak používat na Internetu různé programy, jak správně pracovat s prohlížečem WWW nebo jaké funkce má určitý software. Tento pohled je pro uživatele velmi užitečný, neboť se naučí správně pracovat s Internetem či WWW a tím zjednodušit práci při pracování s informacemi.

3.2.5.3 Internet jako soubor zdrojů

Na Internet se dá nahlížet i jako formu „Zlatých stránek“. Jestli uživatel potřebuje najít např. informace o filmech s Karlem Rodenem, pak se podívá v knize na heslo „filmy, Karel Roden“ a najde seznam URL adres, diskusních skupin či elektronických konferencích. Tyto knihy se snaží o maximální pokrytí šíře témat. Výběr zdrojů k jednotlivým položkám indexu není nahodilý, ale jsou zveřejňovány jenom kvalitní a hodnotné prameny. Na rozdíl od vyhledávacích strojů, kde se nalézají stovky zdrojů, ale bez záruky kvality. Jejich nevýhodou je, že Zlaté stránky Internetu rychle zastarávají. Dalším problémem je, že sice poskytují maximální šíři témat, ale je to vyplaceno malou podrobností. Konkrétních zdrojů nebývá mnoho. Poslední, ale pravděpodobně nejzávažnější nevýhodou je omezenost pohledu na jedno téma v jednom okamžiku. Jinými slovy řečeno, Zlaté stránky selhávají při hledání kombinací témat.

3.2.5.4 Internet jako knihovna

Knihovna a internet mají společné především to, že obsahují informace týkající se různorodých témat. Čtenář či uživatel si může většinou volně procházet a listovat dokumenty, které ho právě zajímají. Internet zde budeme chápat ne jako jednu obrovskou knihovnu, ale jako soubor určitých knihoven rozprostřených po celém světě. Bohužel vedení knihoven není centralizované, a tak se stává, že každá knihovna vlastní metodu organizování. Navíc neexistuje přesný výčet všech knihoven a jejich obsahu. Namísto toho, lidé vytvářejí desítky indexu, které obsahují rozličné informace, jinou strukturu a někdy i s rozdílnou aktualizací. Dalším kritickým faktorem je, že uživatel má ztíženou situaci při přecházení mezi jednotlivými knihovnami. Pokud ale například zná adresy některých knihoven, může se je pokusit prohlédnout. Naneštěstí není jisté, jestli se mu podaří najít dostatek knihoven, aby obsahovali pro něj užitečnou informaci. Na druhou stranu je knihoven takové množství, že je prakticky nemožné prohledat je všechny.

Typická práce uživatele je volit si vhodná klíčová slova do podoby dotazu, který je vyhledávacím systémem posouzen, a nakonec uživateli prezentuje výsledky a odkazy na možné relevantní informační zdroje. Velikou výhodou tohohle úhlu je jednoduché použití. Naopak fáze čekání může být nepříjemně dlouhá a většina výsledků jsou často nepřesná. Vše hlavně závisí na vhodném a přesně zadaném dotazu.

4 ZPŮSOBY ZÍSKÁVÁNÍ INFORMACÍ

V této části si řekneme, jakými způsoby se k datům nebo k informacím dostaneme. Už jsme se dozvěděli, že ve světě Internetu se jich vyskytuje nepřeberné množství. Pokud ale neznáme způsob a cestu, jakým bychom se k nim přiblížili, může nám hledání zabrat obrovské množství drahocenného času. Služby jako WWW nebo FTP nám sice urychlují práci, ale spíše v ohledu sběru dat. Cestě, jak ještě více zjednodušit hledání, vycházejí částečně vstříc služby jako předmětové katalogy nebo vyhledávací a metavyhledávací stroje aj. Přestože nám nemusejí stoprocentně zaručit úspěšné a rychle vyhledávání, ukáží nám oblasti, o kterých bychom jinak neměli ani ponětí.

Pro získávání informací na odborné úrovni se používá jeden termín, o kterém si něco podrobnějšího řekneme. Jedná se o **data mining**, neboli dolování dat. V dnešní informační době má tato metodologie nesmírný význam pro množství firem v nerůznějších odvětví. Co to tedy data mining je?

4.1 DATA MINING

Možnost a schopnost podniků dokázat využít informace na Internetu se stává v dnešní době jejich konkurenční výhodou. Internet představuje pro člověka miliardy a miliardy stran levných informací. Dříve se za získání podobných informací museli platit ohromné sumy peněz. Dnes je většina zdarma. Spoustu informací, pro dynamický vývoj firmy nezbytných, dříve nebylo možné vůbec získat. Přesto však značné kvantum firem tento fakt podceňuje nebo data mining u nich správně nefunguje. Data mining se používá v komerciální sféře, ve vědeckém výzkumu i v jiných oblastech. U marketingu slouží například při rozhodování, které klienty oslovit dopisem s nabídkou služeb nebo při analýze genetických informací u exaktních věd.

Tento pojem lze odvodit pomocí rozdílných definic. Některé obsahují postup kredibilních resp. užitečných vztahů v obrovských databázích. Jiné definice jsou založeny na specifických analytických technikách jako jsou například neuronové sítě, genetické algoritmy aj. Není dána žádná jedinečná metoda, která vyřeší všechny úlohy s libovolným typem dat. Jinými slovy se pro různé úlohy a data hodí odlišné metody. Nejlepší řešení, jak dosáhnout kvalitních výsledků, je použití kombinací rozdílných metod.

„*Data mining je analytická* metodologie, která se zabývá získáváním *netriviálních skrytých a potenciálně užitečných informací z dat. Někdy se chápe jako analytická součást dobývání znalostí z databází (Knowledge Discovery in Databases, KDD).*“
[BERKA, Petr, Dobývání znalostí z databází, Praha: Academia, 2003, str. 366]

„*Data mining je proces, který používá různé analytické nástroje pro odhalení ukrytých vzorů a závislostí v datech. Výsledkem je predikční model, který je podkladem pro rozhodování.*“
[Two Crows Corporation, viz. http://www.spss.cz/sl_datamining.html]

„*Data mining je proces objevování znalostí, pomocí získávaní dříve neznámých akčních informací z velmi rozsáhlých databází.*“
[Aaron Zornes, The META Group, viz. http://www.spss.cz/sl_datamining.html]

„*Data mining je proces objevování smysluplných nových korelací, struktur a trendů pronikáním velkým množstvím archivovaných dat, který využívá modelové poznávací technologie stejně jako statistické a matematické techniky.*“
[Gartner Group, viz. http://www.spss.cz/sl_datamining.html]

4.1.1 HISTORIE

První zmínky o data miningu se datují od 60. let minulého století spolu s rozvojem počítačové techniky. Prvotně se jednalo o využívání regresní analýzy s automatickým výběrem proměnných. První předvedení data miningové metodologie v praxi znamenal až rozvoj statistických metod,

databázových aplikací a umělé inteligence spolu s rychlou expanzí počítačové technologie. Kvůli velkým fluktuacím v korelaci ale neměl očekávaný úspěch.

Opravdového obratu jsme se dočkali počátkem 90. let. Pomohly tomu nově vybudované metody, které umožňovaly vyhnout se právě falešných korelacích. Navíc v té době vzrůstaly na Internetu počty komerčních organizací, který disponovaly velkými objemy dat. Organizace už nebyly schopny získávat podklady pro rozhodování klasickými metodami. Proto má v současnosti data mining velký význam pro komerční účely. Časté aplikace jsou především v oblastech přímého marketingu (výběr klientů pro oslovení), finančnictví (obchodovaní rizika), maloobchodního prodeje (analýza nákupních košíků), telekomunikací (prodej programů) nebo u internetových obchodů (efektivita reklamy). Nárůst aplikací byl zaznamenán i v oblasti softwarového a konzultačního trhu. Existuje poměrně široká nabídka specializovaných softwarů pro tento účel – SAS Enterprise Miner a SPSS Clementine patří mezi přední softwarové produkty.

4.1.2 METODOLOGIE

Jelikož data mining představuje nesmírný počet metod a způsobů práce, je složité popsat univerzální instrukci k postupu. Každá metoda má jiné vyžití a své klady a zápory. Avšak právě dvě největší organizace chtěly tuto propast překlenout. Během 90. let proto vznikly dvě obecné metodologie, které alespoň v hrubých rysech popisovaly jednotlivé kroky. První metodologie SEMMA vznikla pod hlavičkou firmy SAS a druhá CRISP-DM byla vyvinuta konsorciem spolu s firmou SPSS.

Obr. 6: Příklad metodologie data miningu (CRISP-DM)

Přestože jsou obě metodologie rozdílné, hrubé společné rysy, jak postupovat si teď představíme v několika krocích.

- **Obchodní/praktický** – Sem patří formulace úlohy a porozumění problému. Protože žádné automatické vyhledávání znalostí nelze provádět úplně naslepo.

- **Datový** – Nyní si připravíme data pro analýzu. Hlavním důvodem proč použít tento krok je, že statistické algoritmy potřebují mít data připravená v určité podobě, a ne surová.

- **Analytický** – V tomto kroku budeme hledat informace v datech, vytvářet statistické modely apod. Nejčastěji používanými metodami bývají logistická regrese, s automatickým výběrem proměnných, rozhodovací stromy nebo neuronové sítě.

- **Aplikační** – Zjištěné poznatky a modely se v této chvíli uvádějí do praxe, například spuštění reklamní kampaně.

- **Kontrolní** – Na závěr je ještě třeba zjistit zpětnou vazbu a v případě dlouhodobých projektů kontrolovat i aktuálnost.

Jak již bylo řečeno, obliba data miningu stoupá a šíře jeho používání se zvětšuje. Uveďme si tedy několik nejzákladnějších úloh.

- V bankovní sféře se nejčastěji využívá k detekci podvodů. Jde o posuzování žádostí o finanční prostředky či odhalování podvodného chování zákazníků.
- Vytvoření segmentů a následné zařazování zákazníků do určitých okruhů, patří v podnikatelské vrstvě k základním cílům.
- Stejně důležitá úloha je získání zákazníka a následně ho udržet.
- Pro stanovení správné diagnózy a podání vhodného léčiva na základě známých příznaků se používají data získaná pomocí minigového predikčního modelu, který dokáže zahrnout i anomálie a skryté závislosti.
- Výborně se hodí pro analýzu časových řad, kde klasické matematicko-statistické modely ztrácejí na úspěšnosti.
- Poslední neméně důležitou úlohou je analýza prohlížení stránek na Internetu, neboli tzv. web mining. Jelikož stránky mohou obsahovat skryté vazby, je obtížné hledat ten správný smysl. Díky data miningu můžeme z těchto dat získat informace o nejčastějších vzorech v prohlížení či zákazníky segmentovat podle jejich chování na Internetu.

4.1.3 MODELY A METODY

Velké rozšíření a víceúčelnost data miningu závisí na rozsáhlém spektru metod, které se využívají při budování modelu. Žádná metoda není naprosto universální a použitelná pro všechny výsledky. Nejlepšího resumé se dosáhne za použití kombinací různých přístupů. V seznamu metod najdeme nejen statistické procedury, ale i například rozhodovací stromy či neuronové sítě. Jsou pojetí závislá na testovacích hypotézách i dynamické modely, které mění svou strukturu na základě zpracovaných dat.

Predikční model pracuje na základě vstupních údajů, z kterých pak poskytuje data výstupní. Potom z nich sestavíme předpovědi. Před zavedením do praxe tyto data otestujeme a následně zkontrolujeme.

Genetické algoritmy řídí proces učení modelu. Zpravidla hledají nejvhodnější strukturu modelu pro konkrétní data. Tento model vychází ze vzoru biologické evoluce, kdy následující generace modelu je dokonalejší než rodič.

K vybudování predikčních modelů slouží například tyto ***metody***: Lineární a logistická regrese, diskriminační, seskupovaní nebo faktorová analýza, asociační pravidla, rozhodovací stromy, neuronové sítě apod.

4.1.4 POTENCIÁLNÍ NEBEZPEČÍ DATA MININGU

I když použití data miningu je v dnešní době skoro nutné, obsahuje pár nevýhod, na které někteří upozorňují. Nejde ale o technické selhání, nýbrž lidské. Vysvětleme si to. Komerční data mining představují masivní a inteligentní zpracování osobních údajů. Proto vznikají často obavy ze zneužití těchto informací. Kromě klasických negativ spojených se záměrným i nezáměrným únikem dat, které se potom využijí pro různé nečestné aktivity od spamu až po vydírání, zde hrozí také specifické zneužití statistických technik. Jedná se například o to, že zločinec si vytipuje své oběti pomocí analýzy dat.

Toto nebezpečí je podle mnoha zdrojů minimální. Důvodem je chybějící databáze „pozitivních příkladů“ úspěšných zločinců, na niž by mohli své modely postavit. Za mnohem větší hrozbu se ale považují technologie, k jejichž vzniku data mining přispívá v akademické sféře. Jedná se například o práci s genomy, které mohou být použity k selekcím osob na principu rasismu. Další ukázkou je metoda identifikace osob, která může být zneužita pro špehování občanů.

4.2 VYHLEDÁVACÍ STROJE A PŘEDMĚTOVÉ KATALOGY

Tyto služby patří mezi nejnavštěvovanější a nejpoužívanější na Internetu. Příčinou je jejich dostupnost, jak se k nim dostat a představují jednoduchý způsob, jak rychle získat data resp. informace. Vyhledávací služby můžeme rozdělit na komerční a nekomerční. Komerční služby se snaží třídit informace s cílem upoutání pozornosti nebo z reklamních důvodů. Nekomerční služby existují z několika důvodů. Ve světě Internetu jsou jich tisícovky. Pamatujte, že ne každá služba je vyhovující.

Vyhledávací služby lze z funkčního pojetí rozdělit na dva hlavní druhy:

- ***Vyhledávací stroje*** indexují slova nebo termíny, které se vyskytují v dokumentech na Internetu. Mezi nejznámější stroje patří Google nebo AltaVista
- ***Předmětové katalogy*** naopak klasifikují dokumenty nebo servery podle tématické klasifikace. U nás je to například Seznam, ve světě Yahoo.

Odlišnosti mezi zmiňovanými principy vyhledávání u jednotlivých služeb nejsou vždy úplně zřejmé. Takovým příkladem je služba Excite, která používá morfologickou analýzu pro testování shody klíčových slov. Vzhledem k těmto okolnostem můžeme Excite zařadit jak mezi vyhledávací stroje, tak i mezi předmětové katalogy. K tomuto vzoru se začínají přibližovat i jiné služby, takže v současnosti u většiny nelze jednoznačně říci, kam patří. Úspěšné vyhledávací služby, jako jsou Yahoo nebo AltaVista, se také stávají základem pro vytváření tzv. ***portálů*** (viz. dále), jejichž hlavním úkolem je integrace přístupu uživatele k základním informačním službám na Internetu.

Získávat informace pomocí vyhledávání není vůbec jednoduché, jak se možná někomu na první pohled může zdát. Existuje nesčetně množství faktorů, které stojí proti dosažení cíle. Uživatel může strávit dlouhou dobu vymýšlením perfektního dotazu, ale stejně nic nemusí najít. Důvodem může být buď to, že se požadovaná informace na internetu nenachází nebo spíše, že se osoba ptá na špatném místě či špatným způsobem. Další situace, která může nastat je, že se zobrazí obrovské množství dotazů, jenž ale nemají žádnou užitečnou hodnotu. Rovněž se setkáme s tím, že člověk potřebuje informaci okamžitě, ale vyhledávací služba reaguje velmi pomalu. Jiné hledisko spočívá v tom, že uživatel dostane odpovídající výsledky, ale server, na kterém se právě tyto informace nacházejí, je momentálně nedostupný. Za další faktor se dá označit, že neexistuje záruka trvalé dostupnosti. I když je člověk s informací spokojen a odkaz na ni si nechá na pozdější dobu, může být nemile potěšen. Taktéž zdroje komerčního typu nás občas překvapí. Problém je hlavně v tom, že uživatel není připraven platit. Na druhou stranu, informace z profesionálně budovaných databází mají zaručenou kvalitu.

Je zřejmé, že při hledání informací na Internetu se člověk nemůže spolehnout na náhodu. Nyní se pokusíme formulovat základní pravidla pro větší úspěch:

- Měli byste pečlivě volit vyhledávací systém. Různé vyhledávací služby mohou mít přes stejný dotaz rozdílné výsledky.
- Pokud si budete myslet, že v okamžiku budete disponovat uspokojivým výsledkem, jste na omylu. Správné hledání zabere nějaký ten čas. Není prozíravé, když věříte jen jednomu zdroji. Mějte vhodnější přístup a nebojte se experimentovat.
- Pokud při prvních pokusech narazíte na neúspěch, nevzdávejte se hned. Zkuste jiné varianty dotazů nebo úplně odlišnou techniku.
- Učte se vlastními chybami a postupně zdokonalujte svoje metody.
- Je dobré znát prostředky. Čtete manuály nebo se dívejte na online nápovědy.

4.2.1 VYHLEDÁVACÍ STROJE

Tato služba patří v poslední době k nejoblíbenějším. Pomocí vyhledávacího stroje je lokalizace informací rychlá a jednoduchá. Uživatel určí klíčové slovo, kterým se snaží vyjádřit svoji informační potřebu. Poté uživatel vloží dotaz do vyhledávacího stroje, který je složen z jednoho nebo více klíčových slov. Pak už stačí čekat, jestli vyhledávací stroj najde dokument vyhovující zadanému dotazu. Tento dokument se obecně označuje jako „hit“. Samozřejmostí je, že k jednomu dotazu může „vyhledávač“ najít více dokumentů (hitů).

Vyhledávací stroje jsou tvořeny třemi hlavními prvky:

První složkou je ***robot***, který pravidelně shromažďuje dokumenty, aby síť WWW byla stále aktuální a co neúplnější. Z těchto dokumentů se vytváří ***index***, který organizuje data, aby se daly rychle a efektivně vyhledat. Třetí složkou je samotný ***vyhledávací stroj*** představující uživatelské rozhraní, ve kterém se zadávají dotazy.

Pokud uživatel potřebuje nalézt na Internetu informace a hodlá využít služby vyhledávacích strojů, měl by se jeho první dotaz zaměřit na název, který je charakteristikou dokumenty. Neméně důležitý je stručný popis obsahu dokumentu, který se nemusí při normálním zobrazení objevit. V neposlední řadě to jsou již zmíněná klíčová slova, která jsou typem metainformace[8] a která vhodně identifikují dokument pro indexování a následné vyhledání. U některých textů se může poštěstit, že je lze najít dle vlastního obsahu, tzv. těla.

Na výsledku a velikosti úspěchu záleží hodně faktorů. Ve stručnosti můžeme například uvést pořadí klíčových slov v dokumentu, výskyt klíčových slov v názvu nebo nadpisech či relativní frekvence klíčového slova.

4.2.2 PŘEDMĚTOVÉ KATALOGY

V prvních letech fungování se předmětové katalogy udržovaly výhradně ručně. Šlo o údržbu z věcného hlediska. Sledovalo se dění ve WWW prostoru, kde vznikaly nové servery. Nový zdroj se opatřil krátkou charakteristikou, a potom se zařadil do odpovídající kategorie v hierarchii katalogu. Bylo možno zařadit zdroj do více kategorií, jestliže tomu charakterově odpovídal. Později toto břemeno bylo přenecháno na uživatelích, kteří sami pomocí registračních formulářů zaslali požadavek na zařazení do katalogu.

Kombinací vlastního zkoumání a oznamování nových zdrojů jsou katalogy postupně obohacovány. Omezujícím faktorem pro ještě větší růst katalogů je, že u každého nového zdroje není povinnost ho registrovat. Proto se může stát, že i zajímavé a informačně hodnotné zdroje nejsou v různých katalozích vůbec zaznamenány. Vysvětlení se nachází ve velkém počtu katalogů. Provozovatel nebo uživatel může zdroj zaregistrovat pouze u několika z nich.

Mezi základní výhody patří:

Obecné vyhledávání = Vyhledávací katalogy se výborně hodí v situaci, kdy potřebujeme takové informace, které jsou se svou povahou a zaměřením dobře klasifikované.

Záruka kvality = Tato velmi výborná vlastnost je dána způsobem, jakým jsou katalogy udržovány. Ruční práce při registrování nových odkazů do katalogu by měla zaručovat určitou minimální kvalitu zdrojů.

[8] Slouží k popisu informační jednotky. Účel spočívá ve snažení nakládání s danou informační jednotkou např. pro potřebu vyhledávání.

A zde je výčet hlavních nevýhod:

Rozsah katalogu = Způsob údržby má za následek, že jejich rozsah je velmi omezený

Obecnost popisů = Jelikož jsou popisy vytvářeny jenom na základě podrobného prozkoumání, pak nemusejí zcela přesně vypovídat obsahu zdroje a jejich informační hodnota může být nízká.

Struktura kategorií = Poněvadž se v každém předmětovém katalogu používá trochu jiné rozdělení, může dojít ke ztížení orientace uživatelů.

Časová náročnost = Hierarchické uspořádání vychází ze subjektivního pohledu správců daného katalogu. Pokud se uživateli jeví uspořádání zmatečné a neorientuje se v něm, může hledáním strávit hodně času.

Platnost odkazů = Přestože je prostor WWW dokumentů velmi dynamický, nemusí být katalog ani přes průběžné kontroly v aktuálním stavu. Je to způsobeno např. krátkodobým výpadkem příslušného serveru nebo již odstraněnými daty.

Se stoupající růstem kategorií a podkategorií v hierarchické struktuře, postupně klesá přehlednost předmětového katalogu. Pak se jednoduchost práce s katalogem může snadno zvrátit v překážku, kdy uživatel i přes správně zvolenou kategorii, nemůže najít relevantní odkazy. Aby se této situaci zabránilo, většina katalogů začala nabízet jako doplňkovou službu možnost vyhledávání v rámci katalogu. Toto vyhledávání se z uživatelského hlediska podobá stejnému principu jako ve vyhledávacích strojích. Na dotaz uživatel obdrží odpověď v podobě kategorií či rovnou konkrétních odkazů.

Analogie s vyhledávacími stroji je pouze povrchní. Každý pracuje na jiném principu. Hlavní rozdíl záleží na tom, kde se vlastně vyhledává. Zatímco u vyhledávacích strojů jde většinou o plné texty, u katalogů probíhá vyhledávání nad texty, které tvoří kategorie, podkategorie, případně odkazů s krátkou charakteristikou.

4.2.3 METAVYHLEDÁVACÍ STROJE

Použití metavyhledávacích strojů je elegantním řešením, jak při vyhledávaní dosáhnout vyšší úplnosti. Je to zapříčiněno tím, že metavyhledávací systémy provedou automaticky a simultánně dotazy do vyhledávacích systémů, které posléze zpracují výsledky vyhledávání a prezentují jej zpět uživateli.

K důvodům, proč používat metavyhledávací stroje, patří:

- Uživatel nemusí znát všechny způsoby zadávání dotazů u různých vyhledávacích strojů.
- Také se zjednodušuje i prezentace výsledků.
- Uživatel nemusí sledovat změny stávajících nebo vytvoření nových vyhledávacích systémů.
- Zkracuje se vyhledávání tím, že nemusíme zadávat stejné dotazy různým vyhledávacím strojům.
- Přispívají k vyšší úplnosti vyhledávání.

Ovšem setkáme se i s negativy:

- Jelikož jsou stroje řešeny jednotným přístupem, musí obcházet uživatelské rozhraní jednotlivých vyhledávacích systémů. Tato vlastnost je velmi cenově náročná. Navíc je uživatel ochuzen o speciální schopnosti jednotlivých vyhledávacích strojů.
- V metavyhledávacích systémech se nenachází informace reklamního charakteru.
- Stává se pravidlem, že s růstem objemu informací na Internetu a větší oblibou těchto systémů (vyhledávacích i metavyhledávacích) dochází ke snížení výkonnosti.

> Uveďme si příklad z knihy Data, informace, znalosti a Internet. Typický přístup metavyhledávacího systému spočívá v tom, že z výsledku každého dílčího vyhledávacího stroje převezme jen 10-50 hitů a z těch potom sestavuje celkový výsledek. Někdy může uživatel stanovit tento počet hitů přebíraných z každého vyhledávacího stroje, ale obvykle stejně nemůže překročit určitou horní mez. Například uveďme, že na dotaz „*unemployment rate*“ vyhodnocovaný jako fráze pomocí systému MetaCrawler s nastavením maximálního počtu 30 hitů z každého vyhledávacího stroje bylo nalezeno 82 hitů. Naproti tomu zpracování strojem Excite dalo pro stejnou frázi výsledek více než 43000 hitů.

Aby metavyhledávací systém mohl fungovat, je zapotřebí tří základních komponent:

- **Rozesílací mechanismus**, který určuje výběr vyhledávacích systémů, kterým bude dotaz zaslán
- **Agent rozhraní**, který pomáhá navazovat spojení s určitým vyhledávacím strojem
- **Zobrazovací mechanismus** prezentující výsledky uživateli

Nejznámějšími metavyhledávacími systémy jsou AskJeeves (http://www.askjeeves.com), Dogpile (http://www.dogpile.com), Metacrawler (http://www.metacrawler.com), Profusion (http://www.profusion.com) a Search (http://www.search.com).

4.2.4 PORTÁLY

Portály vznikly za účelem zpřístupnit uživatelům ještě větší komfort při vyhledávání informací. Základním principem je jakési spojení vyhledávaní informací katalogového a fulltextového typu. Dále portály poskytují uživateli komplexní služby jako je zpravodajství, zábavu, internetové obchodování, diskusní fóra nebo bezplatný email. Pro uživatele portál znamená místo, kde má všechny základní služby blízko sebe, a proto nemusí při hledání jednoduchých informací jako je například předpověď počasí spoléhat na svoje hledací dovednosti nebo na vzdálené servery.

5 VYHODNOCOVÁNÍ DŮVĚRYHODNOSTI INFORMACÍ

V dnešním světě informačních technologií a služeb se nevyhneme skutečnosti, že s velkou pravděpodobností narazíme na chybné nebo nepřesné informace. Avšak není třeba zoufat. Fakt, že informační zdroj může obsahovat množství nesprávných údajů, ještě neznamená pro uživatele těchto údajů pohromu. Zde je základním aspektem to, abychom byli schopni tyto informace rozlišit. Na druhou stranu musíme ale uznat, že poznat tyto informace v prostředí Internetu je velice těžká záležitost. Proto význam slova *„důvěryhodnost“* má tady velkou váhu.

Internet má neomezené možnosti pro každého, a tak jeho případná regulace není přijatelná a snad ani možná. Úplně každý si může vytvořit profesionálně vypadající webovou stránku, a to i zcela anonymně nebo v totožnosti jiného subjektu. Tento podvod byl například uskutečněn na společnost Bloomberg. Proč vytvářet další a další prezentace nebo stránky? Důvodů je několik. Jednak je prezentace na Internetu levnější než ji tisknout a propagovat běžnou cestou. Dalším aspektem je, že zde existují motivační faktory, např. zábava, komerční i nekomerční zájmy nebo třeba právě i ilegální činnost.

Další ztížení vyhodnocení přináší tzv. uzavřenost systému. Když člověk posuzuje důvěryhodnost daného webového dokumentu, většinou musí spoléhat na další odkazy (např. hypertextové), které se v něm objevují. Můžeme s určitou pravděpodobností říci, že dokument je spolehlivý, pokud odkazy na zdroje jsou důvěryhodné. Pokud ovšem nemáme o těchto odkazech žádné hodnocení, existují zde dvě nebezpečí. Za prvé, že se budeme pohybovat v bludném kruhu a za druhé, že jeden nedůvěryhodný pramen je hodnocen jiným nedůvěryhodným zdrojem.

Avšak shodují-li se spolehlivé zdroje, nemáme ještě zaručeno, že daný dokument je důvěryhodný. V dnešní době se stává pomalu trendem, že informace z jedné stránky jsou zkopírovány do několik dalších. Můžeme například uvést špatný postup odstranění určitého viru, který byl garantován na několika českých stránkách.

HLAVNÍ PROBLÉMY INFORMACÍ VE VZTAHU K INTERNETU	
Název problému	**Popis problému**
Anonymita zdrojů	Identifikace pouze URL adresou je nedostatečná. Právě napodobování adres stránky využívají někteří k „nekalostem“. Viz. www.whitehouse.gov (oficiální) a www.whitehouse.com (neoficiální)
Napadání zdrojů vandaly	Ani známé a prověřené firmy a instituce se neubrání útokům hackerů, kteří mohou změnit obsah stránky, aniž bychom to poznali.
Manipulace s informacemi	Každý si může založit svoji stránku a šířit např. neautorizované informace (viz. později)
Chybějící záruky	Vlastník nebo poskytovatel informací by měl zajistit ochranu proti napadení a zaručit se za jejich správnost. Např. u Internetového Obchodního rejstříku ČR záruky chybí.
Nemorální informace	Jedná se o informace, které jsou nežádoucí z hlediska obecné morálky.

Tab. 4: Příklady hlavních problémů důvěryhodnosti informací

Nejspíš se ale shodneme, že věrohodná, nebo-li **kvalitní informace by měla být spolehlivá, důvěryhodná a solidní.** Nyní si probereme jednotlivá přízviska:

- ***Spolehlivost*** je míra shody mezi skutečností a informací o této skutečnosti
- ***Důvěryhodnost*** je míra zabezpečení informace proti různým druhům možného napadení.
- ***Solidnost*** popsaná abstraktivními pojmy jako poctivost, spravedlivost, slušnost, mravnost,…

[Ing. DOUCEK P., Ing. BÉBR R., Manažerské informační systémy a jejich ekonomika, VŠE skripta, 2002, str.10]

Charakteristika kvality	**Pohled uživatele**
Správná data	Data, která potřebuji
Se správnou úplností	Všechna data, která potřebuji
Ve správném kontextu	Význam, kterému rozumím
Se správnou přesností	Mohu se na ně spolehnout
Ve správném formátu	Mohu je snadno použít
Ve správný okamžik	Když je potřebuji
Na správném místě	Kde je potřebuji

Tab. 5: Uživatelský pohled kvalitní informace

5.1 TYPOLOGIE DŮVĚRYHODNOSTI INFORMACÍ NA WWW

Nyní si odlišíme základní rozdíly mezi důvěryhodností. Podle Petra Boldiše z Ústavu informačních studií a knihovnictví z filosofické fakulty Univerzity Karlovy se důvěryhodnost rozděluje do tří kategorií.

a) matoucí a záměrně zfalšované informace
b) neautorizované informace
c) autorizované informace

Nyní si je blíže probereme.

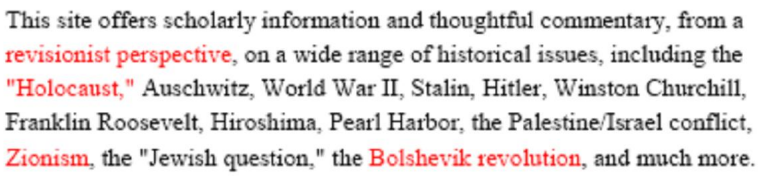
a) matoucí a záměrně zfalšované informace

Definovat falešné nebo zfalšované informace můžeme jako údaj, který se snaží dosáhnout účelového chování uživatele. Mají za cíl oklamat a podvést čtenáře. I když stránky mohou mít perfektní design a vypadat věrohodně, co se týče objektivity a přesnosti, vycházejí vniveč. k těmto stránkám patří například i propagandistické texty, které poskytují účelově zkreslené informace.

V bezpečí nejsme ani v různých důvěryhodných doménách, jako je například .org (organization). Jelikož zakládání webových stránek a přidávání dokumentu je na Internetu neomezené, může kdokoliv na tyto domény zaregistrovat svůj falešný materiál.

Příklad z <http://www.inforum.cz/inforum2003/prispevky/Boldis_Petr.pdf>

Z výňatku stránky Institute for Historical Review vidíme, že používání jazyka a výrazů může být velmi vypovídající. Červeně jsou uvedena slova, která samostatně nebo v použitém kontextu naznačují, že se nemusí jednat o odborný a nestranný materiál.


This site offers scholarly information and thoughtful commentary, from a revisionist perspective, on a wide range of historical issues, including the "Holocaust," Auschwitz, World War II, Stalin, Hitler, Winston Churchill, Franklin Roosevelt, Hiroshima, Pearl Harbor, the Palestine/Israel conflict, Zionism, the "Jewish question," the Bolshevik revolution, and much more.

Obr. 7: Výňatek stránky Institute for Historical Review

Textovou analýzou lze v tomto případě zjistit, že se pravděpodobně může jednat o stránku orientující se krajní pravicí a antisemitismem. Pro konečný závěr se ale musí tato hypotéza podpořit dalšími fakty o autorech, umístnění stránky apod.

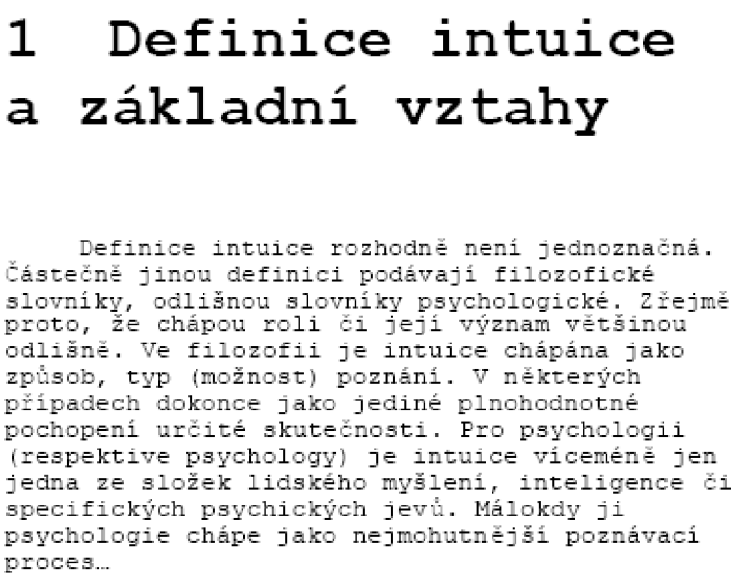
b) neautorizované informace

S tímto typem informací se můžeme na Internetu střetnout nejčastěji. Neautorizované informace se mohou vyskytovat ve dvou variantách.

1. varianta: Informace, u kterých není uveden zdroj, je porušením citační etiky.

Tento problém sužuje jak uživatele (resp. čtenáře), tak i autory jednotlivých stránek. Autoři se musí naučit vkládat správně použité zdroje a odkazy. V naprosté většině není jasné, zda autor předkládá svá vlastní tvrzení nebo se jedná o převzatý cizí text. Není třeba, mít přesně stanovenou strukturu, ale na druhou stranu zde existuje nepatrná hranice mezi necitováním a plagiátorstvím[9]

Příklad z <*http://www.inforum.cz/inforum2003/prispevky/Boldis_Petr.pdf*>


1 Definice intuice a základní vztahy

Definice intuice rozhodně není jednoznačná. Částečně jinou definici podávají filozofické slovníky, odlišnou slovníky psychologické. Zřejmě proto, že chápou roli či její význam většinou odlišně. Ve filozofii je intuice chápána jako způsob, typ (možnost) poznání. V některých případech dokonce jako jediné plnohodnotné pochopení určité skutečnosti. Pro psychologii (respektive psychology) je intuice víceméně jen jedna ze složek lidského myšlení, inteligence či specifických psychických jevů. Málokdy ji psychologie chápe jako nejmohutnější poznávací proces…

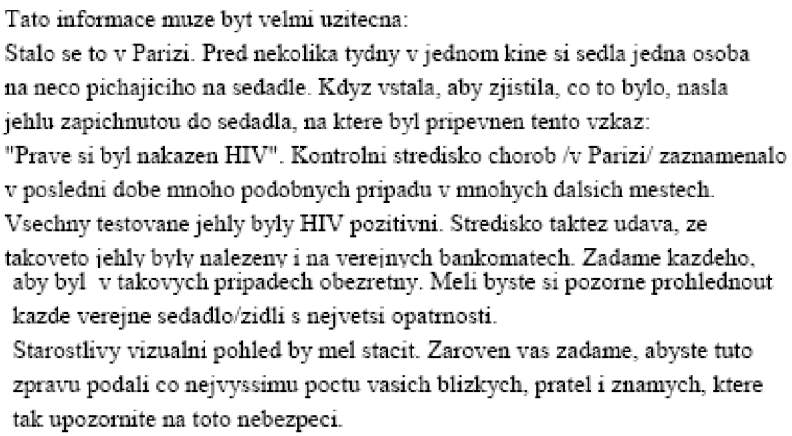
Obr. 8: Příklad neautorizované informace

2. varianta: Informace typu „pověsti“ (hoax)

Jedná se o neověřená tvrzení nebo informace, které jsou šířeny bez ověření pravdivosti. Tento druh informací je šířen převážně pomocí elektronické pošty a webových stránek. Jedná se o různé nepodložené zprávy (upozornění na nové viry), fámy (angl. hoax) apod.

[9] činnost, při které je napodobována originální umělecká díla a u níž je jako původce díla uveden napodobitel.

Příklad z <http://www.inforum.cz/inforum2003/prispevky/Boldis_Petr.pdf>



Tato informace muze byt velmi uzitecna:
Stalo se to v Parizi. Pred nekolika tydny v jednom kine si sedla jedna osoba na neco pichajiciho na sedadle. Kdyz vstala, aby zjistila, co to bylo, nasla jehlu zapichnutou do sedadla, na ktere byl pripevnen tento vzkaz:
"Prave si byl nakazen HIV". Kontrolni stredisko chorob /v Parizi/ zaznamenalo v posledni dobe mnoho podobnych pripadu v mnohych dalsich mestech. Vsechny testovane jehly byly HIV pozitivni. Stredisko taktez udava, ze takoveto jehly byly nalezeny i na verejnych bankomatech. Zadame kazdeho, aby byl v takovych pripadech obezretny. Meli byste si pozorne prohlednout kazde verejne sedadlo/zidli s nejvetsi opatrnosti.
Starostlivy vizualni pohled by mel stacit. Zaroven vas zadame, abyste tuto zpravu podali co nejvyssimu poctu vasich blizkych, pratel i znamych, ktere tak upozornite na toto nebezpeci.

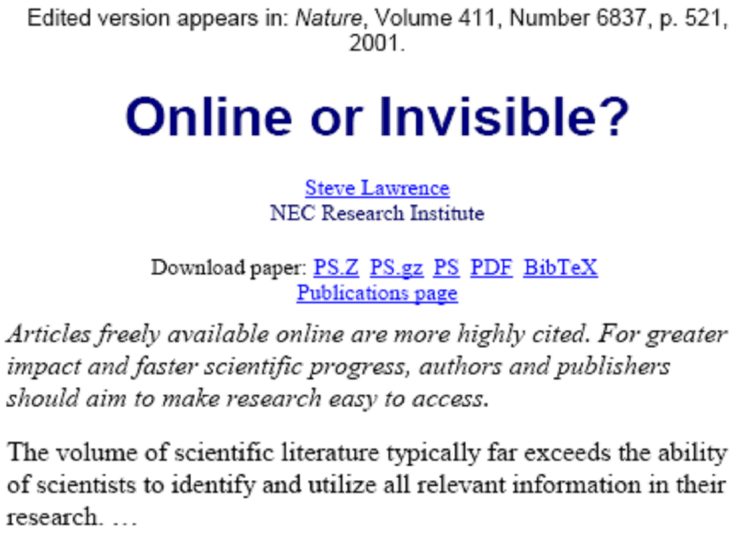
Obr. 9: Ukázka fámy

Jako ochranu proti těmto zprávám vznikly speciální stránky (http://www.hoax.cz), které sbírají nejrůznější zprávy, obíhající po Internetu, a následně vyvracují jejich tvrzení.

c) autorizované informace

Informace tohoto druhu uvítá každý velmi rád. Jedná o zdroj, kde se uvádí pramen informace, důvod nebo účel, proč byla publikována apod. Hlavní výhodou je, že v dokumentu je jasně viditelné autorství a další znaky, podle kterých se dá určit původ (kontakt, odkaz na vyšší sídlo, aktuálnost).

Příklad z <http://www.inforum.cz/inforum2003/prispevky/Boldis_Petr.pdf>



Edited version appears in: *Nature*, Volume 411, Number 6837, p. 521, 2001.

Online or Invisible?

Steve Lawrence
NEC Research Institute

Download paper: PS.Z PS.gz PS PDF BibTeX
Publications page

Articles freely available online are more highly cited. For greater impact and faster scientific progress, authors and publishers should aim to make research easy to access.

The volume of scientific literature typically far exceeds the ability of scientists to identify and utilize all relevant information in their research. …

Obr. 10: Ukázka autorizované informace

5.2 POJETÍ DŮVĚRYHODNOSTI INFORMACÍ

Definovat důvěryhodnost informací je velmi složité. Jedná se totiž o komplexní vztah charakteristik informačních zdrojů. Dalo by se říci, že úzce souvisí s kvalitou, spolehlivostí a přesností. Tímto

pojmem se zabývá celá řada různých disciplín, od psychologie, filosofie, přes sociologii, až po politologii či marketing.

Podle Fooga a Tsenga je důvěryhodnost vnímána jako kvalitativní vlastnost. Dále objevili, že většina odborníků uvádějí ***věrohodnost*** a ***odbornost*** jako dvě základní složky důvěryhodnosti. Dále uvádějí, že celková důvěryhodnost se skládá ze čtyř dílčích hledisek.

DVA FAKTORY VNÍMÁNÍ DŮVĚRYHODNOSTI	
Věrohodnosti	**Odbornosti**
nezaujatost nepředpojatost pravdivost správnost čestnost objektivnost spolehlivost platnost	zkušenost inteligence působivost informovanost praxe vliv význam

Tab. 6: Faktory vnímání dle Fogga a Tsenga

DÍLČÍ TYPY DŮVĚRYHODNOSTI	
Typy	**Příklady**
vznikající z obecných předpokladů vnímání nebo z domněnek a stereotypů v dané kultuře	typy domén počet zobrazení na počitadle
čerpající z informací získaných od třetí strany	rady od právníka, lékaře či přítele odkazy z autoritativních webových stránek
vycházející pouze ze vzhledu	profesionální design webové stránky
čerpané z přímé zkušenosti	navigační mapy na stránkách odkazy rychlý přístup v minulosti prověřený dokument

Tab. 7: Rozdělení důvěryhodnosti dle Fogga a Tsenga

VLIVY WEBOVÝCH SLOŽEK NA DÍLČÍ TYPY DŮVĚRYHODNOSTI				
	obecné předpoklady	**informace ze třetí strany**	**vzhled**	**přímá zkušenost**
provozovatel	neziskové organizace	experti	provozovatel známý v i mimo web	rychlá komunikace mezi provozovatelem a uživatelem
obsah	reklama společností	hodnocení obsahu nezávislou agenturou	poměr obsahu informace a reklamy	správné a objektivní informace
design	vytvořeno profesionální firmou	hodnocení a ocenění od jiných	atraktivní design, snadná čitelnost	rychlý pohyb po stránce a správné zdroje

Tab. 8: Ovlivňování důvěryhodnosti pomocí složek webu dle Fogga a Tsenga

Hodnocení důvěryhodnosti obvykle bývá chápáno z dvou hledisek. Jedná se o chyby z důvěřivosti a o chyby z nedůvěřivosti.

Důvěřivostní chybou rozumíme důvěryhodný dokument, který důvěryhodný ve skutečnosti není. Opakem je chyba z nedůvěřivosti, kdy určitý objekt je důvěryhodný, i když ho uživatel považuje za nespolehlivý.

DVĚ CHYBY PŘI HODNOCENÍ DŮVĚRYHODNOSTI		
	uživatel vnímá objekt jako důvěryhodný	**uživatel vnímá objekt jako důvěryhodný**
dokument je důvěryhodný	správné přijetí	chyba z nedůvěřivosti
dokument není důvěryhodný	chyba z důvěřivosti	správné odmítnutí

Tab. 9: Základní chyby při posuzování důvěrohodnosti

Další pohled teorie přináší B. J. Fogg. Jeho pojetí „výraznost – interpretace“ se zakládá na uživateli, jenž si při hodnocení důvěryhodnosti nejdříve všímá nápadných prvků webové prezentace, a potom si o nich udělá úsudek. K úspěšnému hodnocení důvěryhodnosti nedojde, pokud nejsou přítomny všechny z klíčových složek (viz. tab. 10).

KLÍČOVÉ SLOŽKY TEORIE „VÝRAZNOST – INTERPRETACE“	
Výraznost jednotlivých elementů	**Interpretace jednotlivých elementů**
Faktory ovlivňující výraznost 1. motivace a schopnost prozkoumat obsah prezentace (*uživatel s vysokou úrovní motivace si všimne více detailů)* 2. téma prezentace 3. cíl uživatele 4. zkušenosti uživatele 5. individuální rozdíly	***Faktory ovlivňující interpretaci*** 1. předpoklady a domněnky uživatele (kultura, minulé zkušenosti) 2. dovednosti / znalosti uživatele 3. kontext (prostředí , očekávání uživatele, situace, úkol)
↓	↓
Vliv na důvěryhodnost *(vliv, který element má na hodnocení důvěryhodnosti)*	

Tab. 10: Klíčové složky teorie dle B. J. Fogga

5.3 INDIKÁTORY DŮVĚRYHODNOSTI

Jelikož určit správnost informace je problém, byla na Internetu publikována pravidla pro hodnocení některých dokumentů (např. webových prezentací nebo stránek). Ta se točí okolo základních otázek, které si uživatel musí zodpovědět.

- **Téma a obsah** (Co?) – Co je obsahem prezentace? Co je tématem?

  Obsah je pravděpodobně rozhodujícím a hlavním faktorem pro hledání tématických celků a jejich následné hodnocení. Ostatní faktory jako věcné zaměření, šíře pokrytí, míra podrobnosti a úplnosti jsou podpůrné – pouze ovlivňují hlavní tok obsahové a tématické orientace.

  Z hlediska šíře pokrytí může být zdroj úzce spjat nebo naopak může pokrývat více tematických celků. Z pohledu hloubky jde o míru podrobnosti zdroje. Šíře pokrytí a míra podrobnosti však nejsou v nepřímé úměrnosti, jak si hodně lidí myslí. Hodnotit obsahovou stránku může s úspěšností jenom ten, u kterého se předpokládá určitá informovanost zdroje, a tím i vysoká míra znalosti.

  Na hodnocení obsahu zdroje má vliv i povaha. Poskytnuté informace mohou mít formu úvah, názorů či zpráv. Taktéž se oceňuje, je-li informace původní nebo jedná-li se o výtah či zkrácení.

  V neposlední řadě obsah ovlivňuje tzv. kvalita psaného jazyka. Jde o srozumitelnost a jednoznačnost daného textu.

- **Umístnění a dostupnost** (Kde?) – Kde se stánky nacházejí? Lze zjistit majitele?

  Hlavním hlediskem umístnění je jeho dostupnost a snadnost k jeho nalezení. Pro uživatele je nejsnadnější, pokud je dokument volně dostupný a rychle přístupný na Internetu. Právě rychlost, cena a jednoduchost jsou hlavní faktory, jaké lidé očekávají od přístupu k informacím.

- **Přesnost a aktuálnost** (Kdy?) – Kdy vznikla? Kdy se stránky aktualizují? Jsou stránky aktuální?

  Přesností rozlišujeme informace od skutečnosti. Zpracováváme informace právě z důvodu, aby vyjadřovaly fakta a nelišily se od skutečnosti. Snadnost posouzení je závislá na povaze informací a na zkušenosti hodnotitele. Někdy je velmi těžké zjistit, jak moc je informace odlišná. Některá fakta, např. teorie, mohou mít subjektivní charakter.

  Při vyhodnocování přesnosti se bere v úvahu několik faktorů:
  - U časopisů se kontrolují zdroje ještě před publikováním.
  - Záleží také na tom, jestli je zdroj založen na provedeném výzkumu.
  - Nutná je objektivnost a neutralita informací.
  - Samozřejmě, že u kvalitního zdroje nesmí chybět odkazy, odkud autor čerpal.

  Aktuálnost vyjadřuje stav informace vzhledem k času. Pro lidi je důležité, aby dokumenty vždy informovaly o současném stavu, který stále platí. Proto je vhodné sledovat, jestli se například webové stránky aktualizují a poskytuji návštěvníkovi vždy aktuální a nová sdělení.

- **Pokrytí a objektivita** (Jak?) – Jakým způsobem je psáno o daném tématu? Je objekt posuzován subjektivně? Jak je prezentace spojena s odkazy na zdroje, resp. s dalšími stránkami?

  Objektivita je úzce spjata s autorstvím. Autor musí psát objektivně, bez zaujetí a s nadhledem. Právě hodně dokumentů znehodnocuje fakt, že po zevrubném prozkoumávání zjistíme, že se autor úmyslně nebo neúmyslně přiklonil k té či oné straně. Na druhou stranu plno informací o veřejných věcech (hodnocení zápasů ve sportu, odůvodnění změny zákonů) se stěží posuzuje a může se stát, že dva kvalifikovaní autoři dvou dokumentů, mohou mít odlišné pohledy.

- **Původ a autorství** (Kdo?) – Kdo je autorem stránky nebo prezentace? Jaká je jeho kvalifikace? Má zkušenosti nebo kvalifikaci psát o daném tématu?

Autor je člověk, který napsal danou informaci a který je za ni odpovědní. U každého dokumentu, prezentace, stránky, by měl být zobrazen. Posouzení autora vychází z několika faktorů. Vodítkem můžou být renomované firmy nebo instituce, které zveřejňují jeho jméno. O autoritě může svědčit také URL adresa. Zde si ale musíme dát pozor na freewebové služby (např. http://www.mujweb.cz), jejichž autory je nutno brát s rezervou. Pověst a tedy věrohodnost autora lze nepřímo ověřit pomocí odkazů, které na tento pramen směřují z prostředí WWW. Dalším dobrým měřítkem je jeho zařazení do databáze zdrojů buď služeb, které jsou zaměřeny na hodnocení, nebo do virtuálních knihoven.

- **Důvod** (Proč?) – Proč dokument vznikl? Jaký je cíl publikace?

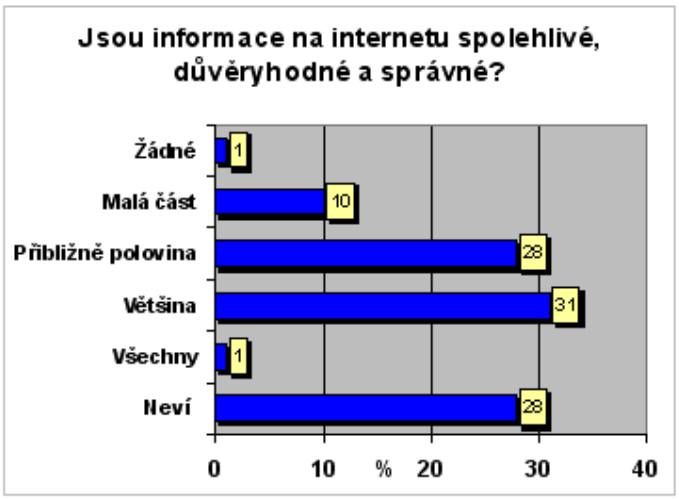
  Účel zdroje poukazuje na jeho záměry a cíle, se kterými byl vytvořen. Určuje rozsah působnosti a cílovou skupinu, pro kterou daný dokument vznikl. Buď je dokument určen pro hromadnou veřejnost a jeho pole působnosti je široké, nebo je určen určitému typu lidí (např. studentům, fanouškům Star Treku, atd.)

- **Design a grafika** má podle B. J. Fogga pro uživatele podstatný význam. Pokud vypadá stránka seriozně a profesionálně, působí velmi důvěryhodně. Naopak se uživateli jeví podezřelé stránky, které mají divoký nebo překombinovaný styl. Na druhou stranu, i profesionálně vypadající stránka může v sobě skrývat falešné informace. Podle mého názoru dokonce podvodníci schválně nechávají poutat osoby na perfektní design svých stránek, aby vzbudili dojem věrohodnosti.

- **Spojitost s dalšími zdroji** bývá v souvislosti s citacemi a odkazy velmi významná. Vylepšit prezentaci mohou právě citace z podložených zdrojů nebo se rovnou na některé odkázat.

5.4 PRŮZKŮM DŮVĚRYHODNOSTI INFORMACÍ

V září roku 2005 proběhl výzkum „Word Internet Project – the Czech Republic“, který se zaměřil na hodnocení informací na Internetu. Cílem bylo zjistit, zda-li a kolik lidí věří Internetu. Jestli považují informace z Internetu za spolehlivé, důvěryhodné a správné. Respondenti odpovídali na následující otázku: ***„Kolik informací na Internetu je podle Vašeho názoru skutečně spolehlivých, důvěryhodných a správných?“***

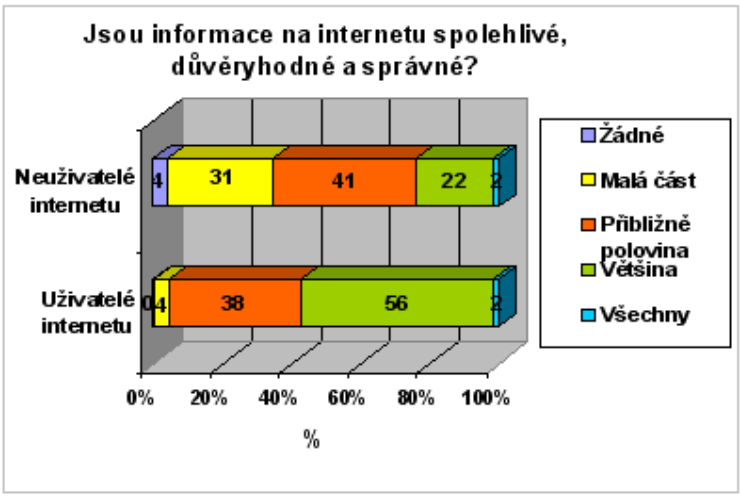


Obr. 11: Graf průzkumu spolehlivosti informací

Výzkum přinesl zajímavé výsledky. Celkem 28 % dotazovaných uvedlo, že neví, kolik informací je na Internetu spolehlivých. Po 1 % respondentů řeklo, že buď všechny nebo žádné informace nejsou

důvěryhodné. 31 % odpovědělo, že většina a 28 % že polovina. Závěrem k této otázce lze říci, že 60 % dotázaný projevila důvěru minimálně k polovině informacím.

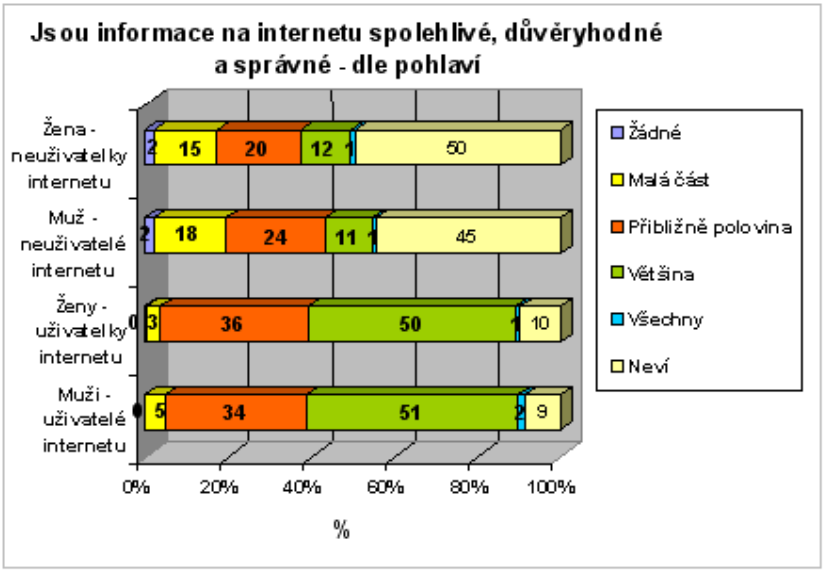
Následující graf ukazuje členění lidí na uživatele a neuživatele Internetu.



Obr. 12: Graf spolehlivosti informací podle uživatelů

Na grafu je vidět, že neuživatelé internetu věří důvěryhodnosti informací výrazně méně než jeho uživatele. Celých 35 % uvádí, že informace buď nejsou spolehlivé nebo jenom malá část. Na druhou stranu u uživatelů Internetu bylo zjištěno, že 96% dotazovaných věří alespoň polovině informací.

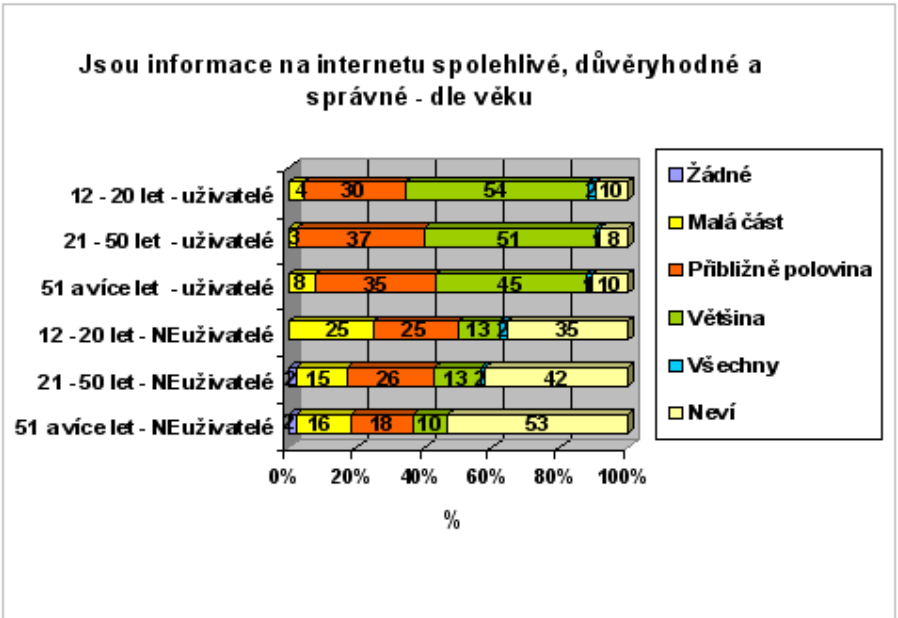
Další graf ukazuje rozdíly mezi muži a ženami – uživateli, neuživateli.



Obr. 13: Graf spolehlivosti informací podle pohlaví

Je patrné, že rozdíly mezi muži a ženami jsou poměrně malé. Za povšimnutí stojí, že u neuživatelů (můžu i žen) panuje větší nejistota než už u uživatelů.

Poslední graf představuje hodnocení informací podle věku dotázaného.



Obr. 14: Graf spolehlivosti informací podle věku

I tady jsou výsledky přibližně stejné. Je ale vidět, že s vyšším počtem let, klesá názor spolehlivosti u většiny informací – 54 % respondentů u 12 – 20 letých, 51 % u 21 – 50 letých a 45 % u více než 51 letých. Dále se dá zjistit, že mladší generace má tendenci více hodnotit a vyjádřit svůj názor. Za zvláštnost můžeme považovat, že nejmladší „neuživatelé“ Internetu poukázali, že jsou k hodnocení informacím velmi kritičtí.

6 ZAJIŠŤOVÁNÍ DŮVĚRYHODNOSTI A BEZPEČNOST INFORMACÍ

Samotná existence věrohodných informací na Internetu bohužel není reálná. Abychom si stále mohli být jisti jejich užitkovou hodnotou, musíme je zabezpečit proti zneužití. Lidi očekávají od dnešního Internetu čím dál tím více. Čerpají informace na nejrůznější účely jako je korespondence mezi lidmi, placení, investování, styk s úřady nebo právě jen analyzování důvěryhodnosti. Většina uživatelů si možná neuvědomuje, jak moc Internet není bezpečný. Proto je ochrana dat prvořadým úkolem, jak zabránit, aby se „tajné“ informace nedostali do nepovolaných rukou.

Pravděpodobně se někdo zeptá, proč tvůrci Internetu nezakomponovali více bezpečnostních procedur? Příčinou je, že v době vzniku Internetu nikdo ani v nejmenší netušil, jakou obrovskou popularitu a jaké využití bude mít dnešní Internet. Nikdo nepředpokládal, že na Internetu bude k dispozici hromadné množství osobních údajů nebo tajných informací, které budou pod neustálým útokem nelegálních činností. Původní požadavek nebyl soustředěn na bezpečnost, ale souvisel se spolehlivostí a robustností. Autoři TCP/IP chtěli vytvořit takové protokoly, které by byly odolné proti nejrůznějším poruchám, výpadkům a současně by pracovali rychle a efektivně.

Díky jednoduchosti architektury Internetu a robustnosti protokolů TCP/IP je v současné době provoz rychlý a stabilní. Lze jen předpokládat, že pokud by Internet stavěli na principu bezpečnosti, vznikali by složitější protokoly, čímž by se zpomalil jeho vývoj. Náklady na zabezpečení, údržbu a vývoj by byl několikanásobně vyšší a přenos dat by nedosáhl takové rychlosti jako nyní.

Na druhou stranu říci, že Internet není bezpečný, by bylo příliš přehnané. Naštěstí dnes existují nejrůznější hardware a software, které obstarávají dodatečné zabezpečující mechanismy tam, kde je to nejvíce potřeba.

6.1 SLABINY INTERNETU

V tomto odstavci si přiblížíme, v jakém ohledu není Internet bezpečný. Za hlavní dva důvody můžeme považovat chování přenosových mechanismů u protokolu IP a chování aplikačních služeb.

6.1.1 CHOVÁNÍ IP PROTOKOLU

Internet Protocol je hlavním přenosovým protokolem síťové vrstvy. Principem fungování je přenášení dat, které se dělí na bloky nebo-li pakety (paketový přenos). Dále pracuje nespojovaně, tzn. že jednotlivé pakety přenáší nezávisle na sobě. Tento způsob připomíná cestu dopisu poštou. V každém uzlu se totiž rozhodne, jak a kudy bude datagram dále přenášen. V praxi ovšem předem nevíme, jakou cestou se bude paket přesně ubírat. To znamená, že odlišné pakety mohou putovat cíli různou cestou. Z hlediska stability to je ideální nápad, ale protože předem neznáme přenosovou trasu, nelze tyto pakety nějak zabezpečit. Navíc spolehlivost IP protokolu pokulhává. Přestože protokol kontroluje, zda přenesená data nejsou po cestě poškozena, pakliže nějaké pakety zmizí nebo se poškodí, nepostará se o jejich náhradu. Počítá s tím, že o tuto nápravu se postará vyšší vrstva.

Další nevýhoda spočívá v neexistenci šifrujícího nebo kódujícího zařízení, které by zabezpečovalo data svěřená k přenosu. Takže není těžké získat tyto data. Naštěstí cestu zabezpečit a udělat ji více spolehlivou lze pomocí kombinací řešení na transportní a aplikační úrovní. Na transportní vrstvě existuje protokol TCP, který transformuje přenos na spolehlivý a spojitý. Na aplikační vrstvě potom můžeme bezpečnost zajistit pomocí šifrování, kódování, resp. jiných vhodných prostředků.

6.1.2 CHOVÁNÍ APLIKAČNÍCH SLUŽEB

Dalším důsledkem nízké bezpečnosti Internetu se projevuje v chování některých aplikací. Většina z nich totiž vychází z předpokladu, že uživatelé se chovají slušně, ohleduplně a morálně, a proto nepodnikají žádné významné kroky ke kontrole a vynucení tohoto etického chování. Jako příklad

jmenujme služby FTP nebo Telnet, které potřebují znát identifikace uživatele (nejčastěji jméno a heslo). Ty pro připojení přenášejí identifikační údaje v nezakódované podobě, tedy jako čistý text. Potom není těžké tyto údaje získat. Dále pod tuto skupinu spadají služby, které se neubrání nekorektnosti uživatelů. Například u služeb elektronické pošty, kdy lze jednoduše zfalšovat hlavičky a poslat zprávu jménem někoho jiného.

6.2 FAKTORY PODPORUJÍCÍ NELEGÁLNÍ ČINNOST

Nyní si objasníme, proč je Internet tak lákavou pochoutkou pro zločince. Tyto faktory mají významný vliv na chování člověka, který by jinak v jiném kontextu zločin nespáchal:

1. **Rychlost přenosu dat**
   Jak již bylo řečeno, rychlost Internetu je obrovská. Stejně rychlá ale může být krádež informace. Když srovnáme krádež konvenčními způsoby (fyzické loupání, otevření trezoru, hledání ve stohu dokumentů) a loupež provedenou pomocí počítače a Internetu, jistě dáte za pravdu, že z časového hlediska máte informaci dostupnou během několika vteřin.

2. **Soukromí a anonymita**
   Provádění ilegálních operací ze sítě nebo samostatného počítače, kdy je „zločinec“ v soukromí a anonymitě, vede k iluzi, že se mu nemůže nic stát. Naproti fyzické vloupání, kdy stačí k nalezení stopy malá chybička, vede k pocitu odtažitosti. V době, kdy je na Internetu uchováno plno osobních údajů, stoupá představa jejich snadného získání a to v důsledku pocitu menšího rizika přistižení.

3. **Digitalizace informací**
   Tento trend je nepochybně užitečný při zálohování, kdy se obáváme, že o svá data přijdeme. Naopak zloděj si je přesně vědom, že originál je jako kopie. Kopírování originálů v libovolném počtu může majiteli způsobit velkou škodu. Navíc snadnost šíření a zachování kvality je dneska velkým problémem.

4. **Elegance a tvořivost**
   Tyto dva faktory patří mezi nejzákeřnějšími příčinami nezákonného chování. Jsou spjaté s určitou mírou dovedností, schopností a znalostí daného „zločince“. Navíc mohou být doprovázeny pocitem intelektuálního uspokojení z nalezení řešení při překonávání překážek. Mnozí útočníci (např. hackeři) se nabourávají do různých systémů, aby předvedli svoji obratnost. Vesměs jde o inteligentní osoby s vysokými tvůrčími schopnostmi, které si chtějí něco dokázat. V posledním období ale u těchto lidí začíná převažovat kriminální podtext nad hravostí.

5. **Snadná dostupnost obětí**
   Jelikož jsou počítače propojovány do sítí, není problém si vybrat potencionální cíl pro svůj vpád. Již neplatí, že cílem útoku se stávali servery velkých společností a firem. Ve skutečnosti se cílem útoku může stát každý počítač připojený na Internet.

6.3 ETICKÉ PROBLÉMY

S rozvojem Internetu vznikla celá řada oblastí, které mohou nabádat k nemorálnímu chování. Mnoho lidí si neuvědomuje, co taková internetová etika představuje. Plno z nás ignoruje nebo zlehčuje chování, jakým působíme na svět Internetu.

6.3.1 SOUKROMÍ

Tvrzení, že svět Internetu je otevřen komukoliv, vede k iluzi, že si můžeme dělat vše, co chceme. Internet možná navazuje pocit anonymity a soukromí, ale není to tak. Lze například použitím volně dostupných informačních zdrojů a vyhledávacích služeb zjistit osobní profil uživatele. Pomocí diskusních skupin a konferencí i díky otevřenosti elektronických pošt, plateb atd. se dá zjistit o

člověku plno informací. Není problém si zjistit, jaké WWW servery nejčastěji navštěvujete, jaké jsou vaše názory na různé tématické okruhy, co rádi kupujete a za kolik. Není potom těžké, tyto informace zneužít.

Důvodem nedostatku soukromí na Internetu je jeho volné pojetí. Právě protože nejsou stanovena žádná přesná pravidla, která by definovala informace soukromé povahy. Proto se nedivme, že dennodenně dochází k získávání osobních informací třetími osobami. V zemích s více prosperujícím Internetem dokonce vznikají různé tajné firmy, které se zabývají sbíráním a následným prodejem osobních profilů.

Aby se uživatel vyhnul těmto potížím, musí si své informace pečlivě hlídat a dbát na to, aby používal správné prostředky.

6.3.1.1 Elektronická pošta

- *Bezpečný přenos* – Přenos dopisů standardním způsobem je velmi nespolehlivý. Pomoci k ochraně může šifrování, které se řeší samostatně.

- *Anonymizovaný přístup* – Ve světě internetu existují anonymní remailery, které umožní zatajit identitu uživatele. Jsou vhodné proti neoprávněnému získání informací, které špehové mohou získat v diskusních skupinách, a nebo před demograficky orientovanými marketingovými firmami. Dále anonymní remailery dávají možnost vyjádřit svobodně svůj názor, aniž by hrozilo riziko následků.

  Anonymní remailer pracuje na principu zvláštního poštovního serveru, jehož pomocí uživatel zašle zprávu se žádostí o její předání na určitou adresu. Remailer ji svým jménem odešle.

  Jak je již patrné i tento způsob se dá lehce zneužít k posíláním anonymních dopisů. Na adrese http://www.cs.berkeley.edu/~raph/remailer-list.html se nachází aktuální seznam anonymních remailerů.

6.3.1.2 Konference a diskusní skupiny

Zde by si měl uživatel dávat pozor, co píše do diskusních skupin nebo konferencí. Musí si uvědomit, že tyto služby jsou veřejné a že si je může přečíst každý. Dalším nebezpečím se mohou stát archivy zpráv, ve kterých se pak dá vyhledat daná položka. Takto se může naplnit úsloví „co je psáno, to je dáno“.

6.3.1.3 WWW

- *Cookies* – Cookies jsou krátké textové informace (soubory) o velikosti kolem 100b, které vysílá server klientovi a ukládá je na náš počítač. Jelikož je http protokol bezstavový, cookies sbírají informace a zjišťují, jestli se jedná stále o nás (např. když dnes nakupujeme zboží do košíku a zítra se rozhodneme pokračovat). Do cookie může například server zaznamenávat, jak často jej uživatel navštěvuje, může ukládat osobní informace, co dělá, jeho oblíbené odkazy atd. Jinými slovy cookies představují veškeré informace o uživateli.

  Existují potom metody, kdy jeden server se vydává za jiný a tím vám přistoupí k cookies. Uživatelé se mohou proti cookies bránit. Buď stačí soubor vymazat nebo cookies v prohlížeči nadobro zakázat.

- *Anonymizovaný přístup* – další možností, jak ochránit svoje soukromí v prostředí WWW je využití anonymizačního serveru na adrese http://www.anonymizer.com. Server Anonymizer používá techniku zprostředkovaného spojení, aby umožnil uživateli obracet se na WWW servery anonymně.

6.3.2 VLASTNICTVÍ

Mezi základní lidská práva patří samozřejmě i právo na vlastnictví. Jestliže si toto právo spojíme s Internetem, jedná se o duševní vlastnictví. Právě v souvislosti se světem Internetu si musíme uvědomit, že objektem ochrany není konkrétní forma (papír, dokument), ale intelektuální obsah (informace). V současné době, s rozvojem digitalizace informací, je porušování autorských práv velkým problémem. K ochraně duševního vlastnictví jsou určeny dva instituty – copyright nebo patent.

Copyright je typem právní ochrany duševního vlastnictví. Vlastník si může s informací nakládat libovolně podle své vůle. On rozhoduje, zda-li bude informace mít privátní nebo veřejný charakter. Problémem duševního vlastnictví je, že předmět tohoto vlastnictví může používat více lidí, aniž by původnímu vlastníkovi něco ubylo. Zde se nemluví o odcizení, ale o kopírování. K vyznačení této ochrany se využívá značka ©. Obvykle pod tuto značku spadají literární, umělecká nebo vědecká díla. Dále pod ochranu copyrightu přibývá i software.

Jistou výjimku z ochrany copyrightem tvoří tzv. **Fair Use**. Jedná se vytváření kopií jinak chráněného díla za specifickým účelem, jako jsou např. zpravodajství nebo kritika.

Nyní si probereme dopady copyrightu na práci ve světě Internetu:

6.3.2.1 Software

Vlastník softwaru a majitel copyrightu si může definovat podmínky pro kopírování:

- *Copyboarding* – Situace, kdy autor zakazuje jakékoliv kopírovaní nebo šíření.

- *Licensing* – V současné době nejčastější způsob šíření prodávaného softwaru, kdy kupující musí dodržet určitě licenční podmínky.

- *Shareware* – Jedná se o zdarma distribuovaný software, u kterého stále platí autorská práva. Například o časově limitovaný software na vyzkoušení, kdy potom musí uživatel zaplatit poplatek.

- *Copylefting* – Jiné označení pro General Public License (GLP, GNU). Volně šiřitelný software, který lze modifikovat, ale nelze bránit jeho dalšímu šíření.

- *Freeware* – Označení pro volně šiřitelný software. Tento program je zdarma distribuován autorem po Internetu. Každý si ho může volně zkopírovat nebo ho distribuovat dále.

- *Public Domain* – Název pro software, který je určen pro veřejné použití. Tento program lze bezplatně šířit, kopírovat a používat. Na public domain se nevztahuje ochrana pomocí copyrightu.

Problémy s porušováním copyrightu u softwaru jsou značné. Vytváření nelegálních kopií je po celém světě značně rozšířeno. Pro neautorizované kopírování programů se vžilo označení softwarové pirátství. Přestože to je činnost nezákonná a ve většině zemí trestná, boj s ní je velmi nevyrovnaný. Podle průzkumů se odhaduje, že ve střední Evropě více než 60 % softwaru je používáno nelegálně. V asijských zemích toto číslo dosahuje až 90 %. Hlavním důvodem vytváření a používání nelegálních kopií je jednoduchost získání a cena jednotlivých programů.

6.3.2.2 WWW dokumenty

Ochrana vlastních, publikovaných WWW dokumentů má zvláštní charakter. Každý autor má určitý styl, který se zobrazí ve zdrojovém tvaru, jenž vytváří jakýsi design dokumentu. Aby byl problém ochrany vidět, je nutné se na dokument podívat ve více rovinách:

- *Obsah* – Jde o pohled na WWW stránku jako na tradiční autorské dílo typu kniha apod.

- *Design* – HTML kód, určující umístnění a vzájemnou interakci jednotlivých obsahových prvků dokumentu, za předpokladu originality a netriviálnosti je chráněn copyrightem. Pokud by vznikla stránka s nápadně stejným designem, jednalo by se o podvod.

  Podobný případ se stal v roce 1997 v České republice, kdy druhým rokem fungoval předmětový katalog Seznam a v květnu se objevil nový katalog HotList. Problémem bylo, že se prakticky jednalo o dokonalou kopii služby Seznam. HotList byl po týdnu na nátlak stažen.

- *Odkazy* – Na hypertextové odkazy v textu lze nahlížet jako nedílnou součástí obsahu dokumentu.

6.3.2.3 Diskusní skupiny a elektronické konference

Každý příspěvek zaslaný do diskusní skupiny nebo konference je chráněn copyrightem. Musí se však jednat o tvůrčí a originální dílo svého autora. Toto dílo může autor posílat i do jiné diskusní skupiny či konference.

6.3.3 SVOBODA

Člověk má mít maximální nezávislost, omezenou pouze oprávněnými nároky na svobodu druhých. V prostředí Internetu má každý uživatel pocit, že jeho volnost je nedozírná. Z hlediska obecné etické teorie ji lze rozdělovat na svobodu rozhodování a jednání.

6.3.3.1 Svoboda rozhodování

Pod tímto pojmem chápeme část vnitřní svobody člověka, který se rozhoduje mezi jednotlivými variantami. Výhoda spočívá v tom, že zde Internet působí jako globální komunikační platforma, která poskytuje uživateli neustálý přísun informací. Ty nám pak pomáhají při správném rozhodování. Nesmí ovšem dojít k zahlcení. Jestliže si jedinec vypěstuje závislost, jeho svoboda rozhodování v reálném světě je tím narušena.

6.3.3.2 Svoboda jednání

Zde se jedná o míru jednání, která je určena hranicemi jistých pravidel. Z hlediska Internetu jsou zejména důležité

- *Anonymita* – Již zde byla několikrát představena. Zde je původce jednání zejména pro zločince, neboť mají pocit, že neexistuje možnost jejich odhalení.

- *Svoboda projevu* – Jinými slovy také můžeme říci cenzura nebo omezování svobody tvorby. Internet, jak sami dobře víme, je prostorem, kde můžeme volně šířit informace, názory, pokud nekolidují se zákonnými nebo morálními pravidly. Problém je v možnosti volně publikovat jakékoliv informace. I když nikdo nechce nikoho omezovat, je pravda, že v posledních pár letech byli několikrát občané pobouřeny určitou zprávou, která byla eticky diskutabilní (návod na sestavení bomby, interrupce, …).

6.3.4 HACKING, CRACKING

Obě tyto činnosti patří mezi velmi nebezpečné zločiny. Napadání webových dokumentů a měnění jejich obsah může mít dalekosáhlý dopad jak u majitelů stránek, tak u návštěvníků. Mezi nejčastější argumenty hackerů patří výmluvy, že firmy mají slabou obranu, že v určitých kulturách není hacking trestný nebo že tím odhalují chyby a zranitelná místa informačních systémů. Dopadení těchto zločinců není jednoduché, ale v současné době existují spolehlivé metody, jak pachatele dopadnout.

Škody způsobené hackery mohou být různého druhu. Ze strany zpracovávání informací lze jmenovat neoprávněné použití informací, jejich únik nebo narušení integrity dat. Činnost hackerů může být považována za *vandalství* v případě, když dojde ke změně designu nebo za *manipulaci* v případě změny obsahu. Může se ovšem stát, že změna designu (formy) ovlivní i obsah dokumentu.

Crackingem zase rozumíme postup úpravy či zkoumání programového kódu určitého programu. Cracker dále upravuje zdrojový kód, aby změnil běh programu nebo přerušil jeho zabezpečení. Pro autory je cracking stejně nebezpečný jako hacking.

6.3.5 TECHNOLOGIE AGENTŮ

Jelikož se od Internetu čeká rozšiřování možností a kapacity, bude stále obtížnější nalézt požadované informace, které jsou roztroušeny v mnoha formách. Řešením této situace může být navržením speciálních WWW agentů a robotů, kteří by vyhledávali informace na základě umělé inteligence nebo složitých algoritmů. V tomto směru snadno narazíme na morální problémy. Jestliže agent filtruje veškeré informace, které pak daná osoba obdrží, existuje riziko zúženého pohledu uživatele na stanovený problém. Lehce se totiž stane, že agent vyřadí informaci, která mohla být klíčovou.

Druhým sporným bodem se jeví naprogramování agentů. Agenti jsou naprogramováni, aby nám usnadňovali práci. Na druhé straně, stejně jednoduše může někdo naprogramovat vyhledávacího agenta tak, aby získával informace např. finančního charakteru, jako jsou třeba údaje o kreditní kartě.

6.3.6 INFORMAČNÍ NEROVNOST

Tvrdit, že Internet je dostupný každé osobě a na jakémkoliv místě, není úplně správné. Z obecného hlediska se sice tento výrok blíží pravdě, ale díky různým ekonomickým, technickým, socio-kulturním i interkulturním omezením se zatím jeví jako nedosažitelný. Právě díky těmto omezením má k Internetu přístup jen malé procento obyvatelstva této planety. Problémy najdeme nejenom mezi jednotlivými zeměmi navzájem, ale často i v rámci jednoho státu. Co se týče vztahů mezi zeměmi, větší podíl přístupu k Internetu mají velmoci a rozvinuté státy, naopak bez možnosti připojení jsou chudé země. V rámci samostatného státu je přístup k Internetu otázkou ekonomických a politických podmínek, to platí v makroekonomickém i mikroekonomickém měřítku.

6.3.7 INFORMAČNÍ OBSAH

Právě obsah je jedním ze závažných etických problémů, které podvodníci rádi využívají k ovlivňování názorů uživatelů. Zařazujeme sem pravdivost a věrohodnost informací, které jsou na Internetu k dispozici. Jejich posuzování a hodnocení je velmi obtížné, protože chybí spolehlivé indikátory, jak je tomu například u konvenčních masových médií. Grafická forma vyjádření může v uživateli probudit iluzi spolehlivosti a autentičnosti. Zvláště snadnost publikování a anonymita v prostředí Internetu je ideální líhní pro šíření pověstí, fám, lží a pomluv. Dále se na Internetu objevují teorie spiknutí, či různé propagandistické dokumenty. Navíc, informace v původním kontextu, která se jeví jako věrohodná a spolehlivá, bývá v jiné souvislosti použita k protikladným účelům. Tímto způsobem vzniká k obtížně rozpoznatelnému směšování normální a virtuální reality.

S pomocí Internetu se setkáváme s dokumenty, jejichž obsah je velmi eticky pochybný. Lze nalézt dokumenty s rasistickým podtextem, náboženskou nenávistí, data napomáhající terorismu nebo

přinášející dětskou pornografii. Avšak tyto díla nespadají pod jednu „kupičku“. Obecně se dají dělit na několik podskupin:

1. Eticky sporný obsah zaměřený na poškozování menšin nebo příslušníků jiných kultur

2. Obsah s extremisticky propagandistickým tématem, kdy jsou šířeny netolerantní a radikálně vypadajícími představami o světě (např. politický extremismus, náboženský fundamentalismus)

3. Obsah, který prezentuje násilí nebo je zavrženíhodný (např. dětská pornografie, brutalita), což je skoro ve všech zemích považováno nejen za nemorální, ale i nezákonné.

Pokud chceme, aby se Internet stal bezpečným a morálním prostředkem k získávání důvěryhodných informací, musíme rozšířit a začít používat metody o bezpečnosti dat, právních úpravách a rychlé zjištění nezákonného chování. Víme, že svět Internetu velmi ulehčuje práci a zkracuje dobu přístupu k informacím, ale na druhé straně je také snadné získat data pochybného charakteru nebo údaje, které nám nepatří.

Podrobněji o tomto problému a hodnocení informací v kapitole 5.

6.4 OCHRANA INFORMACÍ

6.4.1 DRUHY INFORMACÍ DLE ZABEZPEČENÍ

Informace členěné podle přístupu mají velký význam pro zabezpečování Internetu i systémů. Typy přístupů vychází ze zákona č. 148/1998 Sb. – O ochraně utajovaných skutečností.

- *Všeobecné* – Tyto informace jsou určené pro nejširší veřejnost, patří sem např. internetové knihovny.

- *Vyhrazené* – Informace, které jsou vyhrazené určitým osobám. Jedná se o informace, představující předmět hospodářského nebo obchodního tajemství. Údaje tohoto typu se chrání před zneužitím v konkurenčním boji. Jsou to např. údaje o firmách, zákaznících, prodeji, výrobě, projektech, plánech, obchodní politice firmy apod.

- *Důvěrné* – Do této skupiny náleží osobní a zdravotnické údaje, elektronická korespondence, data z právní oblasti atd. Nahlédnout do nich může jenom majitel nebo příslušná osoba (pověřený policista, právník).

- *Tajné* – Informace důvěrného typu, jejichž zveřejnění by mohlo způsobit nevyčíslitelné škody.

- *Přísně tajné* – Informace, které podléhají ještě většímu utajení.

Zákon 101/2000 Sb. – O ochraně osobních údajů definuje ještě pojem *„citlivé údaje“*. Jde o informace národnostního a rasového původu, politického postoje, členství v odborových organizacích, náboženství, trestné činnosti, zdravotního stavu nebo sexuálního života. Tyto údaje jsou chráněny více než ostatní osobní data.

6.4.2 PŘÍSTUP K INFORMACÍM

Existují tři základní postupy přístupu k neveřejným informacím:

1. ***Autorizace*** – Jedná se o činnost, kdy stanovujeme přístupová práva k daným informacím.
2. ***Identifikace*** – Proces, kdy systém zjišťuje osobu, která v tomto okamžiku přistupuje k informacím.

3. ***Autentizace*** – Druh potvrzení, které oznamuje, že uživatel byl úspěšně ověřen a je autorizován pro přístup k vyžadovaným informacím.

6.4.3 DRUHY OCHRANY

V tomto odstavci si obecně povíme, před jakými hrozbami se informace dají zabezpečit. Především jde o ochranu proti ztrátě nebo zničení, před zneužitím a proti nežádoucím změnám.

6.4.3.1 Proti ztrátě a zničení

Zabezpečuje dostupnost informací. Bezpečnost se týká všech informací od všeobecných až po přísně tajné. Náleží k primárnímu druhu ochrany. Data mohou být napadena neúmyslně (havárie či porucha serveru) nebo úmyslně (viry, teroristický útok atd.)

6.4.3.2 Proti zneužití

Zde chráníme důvěrnost informací. Ve většině případů se zabezpečení týká všech druhů informací kromě všeobecných. Principem je pasivní přístup, což znamená, že k informacím mohou přistupovat jen oprávněné osoby.

6.4.3.3 Proti nežádoucím změnám

Ochrana proti nežádoucím změnám zajišťuje integritu informací. Vztahuje se ke všem druhům informací včetně všeobecných. Jde o aktivní přístup, tzn. že data smí zpracovávat (vkládat, měnit, rušit) jenom určité osoby (např. pracovníci serveru, na kterém je daná informace umístněna).

6.4.4 MOŽNOSTI OCHRANY

V dnešní době již existuje nespočet způsobů zabezpečení, nejen proti zneužití, ale i proti ztrátám, zcizení nebo změně obsahu. Za každou cestu se snažíme udělat v prostředí Internetu klidnou a bezpečnou zónu, která by se pro všechny stala oázou jistoty. Tento proces však není dokončen. Máme před sebou ještě dlouhou a strastiplnou cestu. Bude otázkou, jestli někdy vynalezeme takovou ochranu, a kdy žádný vir, hacker ani jiný záškodník nebude schopen zneužít věrohodnosti informací. Zatím to bohužel zní jako utopie. Možná se časem podaří snížit rizika na přijatelnou úroveň.

Nyní se podívejme, jaké jsou hlavní technologie k zajištění požadavků na identifikaci, autentizaci, integrity a důvěrnosti informací.

6.4.4.1 Symetrické šifrování

Jedná se o druh techniky, která slouží k šifrování se dvěma identickými klíči. Oba se dají využít jak pro zašifrování, tak i pro odšifrování. Důvěrnost dat je zajištěna, když odesílatel i příjemce jsou vybaveni těmito klíči. Současně může být splněn i požadavek na identifikaci a autentizaci. Výhoda symetrického šifrování je jeho výpočetní nenáročnost. Naopak mezi nevýhodu patří manipulace s identickými klíči. Zde hrozí nebezpečí, že se klíče dostanou do nepovolaných rukou. V praxi se často používá fyzické předání tajného klíče na vhodném nosiči.

6.4.4.2 Asymetrické šifrování

Již podle názvu se dá usoudit, že se při této technice pracuje se dvěma různými klíči, z nichž výhradně jeden musí zůstat utajen jako tzv. privátní klíč. Druhý klíč není kvůli snazší dostupnosti utajovaný. Proto je označován jako klíč veřejný. Podstatné pro výměnu dat je

důležité, že co je zašifrováno jedním z klíčů, lze pouze odšifrovat druhým. Asymetrické šifrování lze tedy využít dvěma rozdílnými způsoby.

Prvním způsobem je ***použití privátního klíče*** k zašifrování a veřejného klíče k odšifrování. Protože privátní klíč má pouze autor dokumentu, může libovolně odšifrovat, změnit obsah a znovu zašifrovat data, což splňuje požadavek na integritu. Díky veřejnému klíči může informaci v zásadě odšifrovat každý. Výsledkem požadavku je splnění identifikace a autorizace. Jelikož data jsou veřejně přístupná, zabezpečení důvěrnosti není naplněno.

Použití veřejného klíče k zašifrování je druhým způsobem asymetrického šifrování. Zde naopak privátní klíč slouží k odšifrování dat. Prostřednictvím veřejného klíče má přístup ke změnám obsahu kdokoli. Nicméně odšifrovat takto zašifrovaná data smí pouze vlastník privátního klíče. V praxi to znamená naplnění požadavku na důvěrnost. Ostatní požadavky ale uskutečněny nejsou.

6.4.4.3 Přístupová hesla

Hesla tradičně slouží k autentizaci a pro potřeby autorizace. Fungují jako ochrana proti zneužití a proti změnám. Dnes je velmi používaným artiklem, jak zabezpečit důvěrně informace na Internetu. Jelikož jsou tyto hesla přenášeny nezabezpečeným způsobem po nezabezpečené síti, není pro profesionála těžké je získat. Ztráta přístupových hesel k elektronické poště, či do on-line obchodu, může způsobit vážné problémy.

Základní pravidla pro výběr a údržbu hesel jsou, že nesmí obsahovat přihlašovací jméno, ani být odvozeno ze žádné informace, která má vztah k uživateli, mělo by být delší než šest znaků, snadno zapamatovatelné a pravidelně ho měnit. Samozřejmostí je neprozrazovat ho žádné osobě.

6.4.4.4 Jednorázová hesla

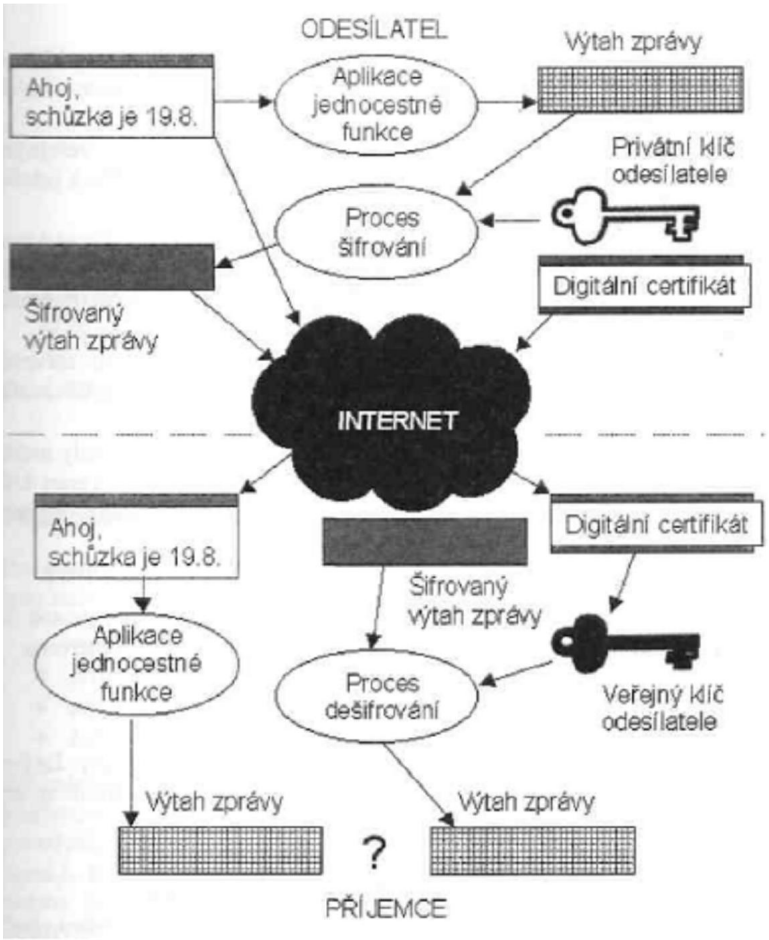
Zvláštní skupinou zabezpečení jsou hesla, která jsou platná jednorázově pro jedinou činnost, takže jejich případné vyzrazení není nebezpečné, protože je nelze znovu použít. Nevýhodou je, jakým způsobem mohou být generovány. Pro tyto účely musí mít uživatel k dispozici vhodný generátor.

6.4.4.5 Elektronické podpisy

Toto elegantní řešení dokáže spolehlivě vyhovět požadavkům na identifikaci, autentizaci i na integritu zpráv. Elektronický podpis funguje na principu asymetrického šifrování s použitím privátního klíče, který patří podepisující se osobě. Jelikož je privátní, nikdo jiný je nemá možnost jej použít. U veřejného klíče se každý může přesvědčit, že podpis je nezfalšovaný.

V praxi ale není v rámci elektronického podpisu privátní klíč použit na zašifrování podepisovaných dat. Doopravdy se šifruje jenom reprezentativní vzorek (tzv. hash), který byl z těchto dat vytvořen. Datový údaj, vzniklým šifrováním, je přiložen ke zprávě a poslán k příjemci. Příjemce, jenž si chce ověřit pravost podpisu, použije veřejný klíč odesílatele k tomu, aby mohl odšifrovat hash. Pak si příjemce vytvoří vlastní vzorek přijaté zprávy a oba je porovná. Jestliže se budou shodovat, má příjemce jistotu, že zprávu podepsal majitel příslušeného privátního klíče. Čím byl splněn požadavek na autenticitu a integritu. Pakliže se vzorky neshodují, znamená to dvě věci: Buď byla zpráva pozměněna nebo se za autora vydává někdo cizí.

Pro zajištění důvěrnosti zprávy lze využít současně s el. podpisem stejný mechanismus asymetrického šifrování, pouze s použitím veřejného klíče.



Obr. 15: Schéma el. podpisu

6.4.4.6 Certifikáty

U asymetrického šifrování existuje kritické místo, které ovlivňuje důvěryhodnost a celkovou spolehlivost mechanismu. Příčinou je, že s veřejným klíčem musí být dostatečně spojena identita osoby, která tento klíč vlastní. Tento konkrétní způsob spojení se nazývá certifikát. Jestliže si příjemce stáhne zprávu spolu s podpisem například z internetového serveru, může se stát, že autor není skutečně tím, za koho se vydává.

Pokud by s použitím veřejného klíče byla spojena identita autora, měl by si být adresát jist, že se skutečně jedná o tvůrce zprávy. Proto je použití veřejného klíče velmi důležité. Jakákoli možnost podvodu či zneužití souvislosti mezi veřejným klíčem a identitou vlastníka by mohla mít obrovské důsledky.

Neboť certifikáty mají být šířeny volně, je zde problém v jeho autentičnosti. V praxi tomu lze předejít dvěma způsoby. První spočívá v tom, že si ho vlastník vytvoří sám. Dalším postupem, který je sice nákladnější, ale spolehlivější je, že certifikát bude stvořen i ručen třetí stranou. Jde o komerční službu poskytovanou tzv. certifikační autoritou (CA).

7 ZÁVĚR

S rozvojem informační společnosti a stále větší oblibou Internetu si musíme přiznat, že náš život je závislý na každodenním přísunu informací. Pravděpodobně tento fakt lidé obvykle nepříliš vnímají, ale důležitá informace může mít neocenitelnou hodnotu. Informace, získaná v prostředí webu, může zachránit život, předejít materiální škodě, ulehčit životní styl. Pro firmy znamená správné rozhodnutí peníze nebo jiný užitek. Říká se, že zpráva se stává cennou, jestliže přijde v pravý čas, rychle a její obsah odpovídá realitě. Osobně bych přidal ještě jednu skutečnost – dostupnost. Je neobvyklé, že informace narazí na nás. Je na každém, aby ji našel. A je rozdíl, pokud požadovanou informaci budeme hledat hodinu, nebo týden. Internet se jeví jako ideální důl, kde můžeme najít hodnotné věci. V této práci jsou popsány možnosti, kde a jakými způsoby získat informace. Na druhé straně se v ní nedočtete návod, jak nad prostředím WWW vyhrát a stát se např. milionářem. Doufám, že se nepletu, když vám sdělím, že tento recept patrně neexistuje. Přesto všechno Internet poskytuje neomezené možnosti všem, co něco potřebují.

Jak už to však bývá, není nic jednostranné. I když se Internet zdá jako ideální zdroj informací, není to tak. Hlavní nedostatek je v obsahu informací… tedy spíše v jejich tvůrcích. Stejně jako je iluzorní dokonalost Internetu, je iluzorní i představa, že jakákoli informace je pravdivá a spolehlivá. Že každý autor nebo původce informace má ty nejčestnější úmysly pravdivě informovat společnosti. Existuje velký počet zpráv, jejichž obsah má negativní vliv. Jedná se o informace klamné, záměrně pozměněné, nepotvrzené. Jejich cílem je vyvolat v lidech paniku, způsobit škody, nebo jen tak si udělat z někoho legraci. Z vlastní zkušenosti mohu říci, že jedna klamavá informace přivodí více ujmy než tucet pravdivých zpráv. Pokud dojde díky těmto informacím k porušení soukromí, k zneuctění, může se stát, že žádná skutečnost už nenapraví tuto hanebnost.

Tvůrci Internetu se příliš nezaměřovali na bezpečnost, nicméně s jeho rozvojem musely být provedeny kroky, které by zamezily úniku informací a zneužití soukromí. Ne nejlepší, ale nejvíc rozšířená metoda ochrany soukromí jsou hesla. Používají se skoro všude, u elektronické pošty, diskusních skupin, při obchodování přes Internet atd. Mezi další pomocníky k zajištění ochrany se řadí speciální algoritmy, šifrování a protokoly, které zabezpečují plynulý přenos a ukládání dat serverech. Pro důležitost zachování věrohodnosti dat však samotné mechanismy nestačí. Hlavní úlohou zde hraje člověk, respektive morální a etické aspekty. Pokud bude fungovat právní legislativa spolu s prevencí, je šance, že se jednou sníží počty kriminálních činů v prostředí počítačů. Nezbytností je přimět vandaly, aby si uvědomovali následky svých jednání.

Autoři dokumentů, kteří mají svá díla k dispozici na Internetu, by měli také přispět. Psát jenom potvrzené zprávy, zveřejňovat odkazy na kvalitní prameny, brát odpovědnost za obsah, to jsou jedny z významných faktorů při publikování dokumentů.

S ohledem na výše uvedené, je v budoucnu velká šance, že se ještě ztenčí časová propast mezi informací na Internetu a uživatelem, který ji hledá. Nové technologie, metodiky a integrace všech prvků tohoto koloběhu nám ulehčí práci, jak při vyhledávání, tak i při snížení rizika při kontrole autentičnosti. Nesmíme však zapomenout na to, že při zlepšování kvality zpracovávání informací je nutné se dívat zpět do minulosti. Jenom s odstupem času korektně zhodnotíme chyby, kterých jsme se v minulosti dopustili.

Současný uživatel dává přednost informacím v elektronické podobě, a tak si nevšímá mnoha papírových dokumentů v nedigitalizované formě, které nejsou zpracovány do databází informačních zdrojů. Proto je prioritou digitalizovat dokumenty, a zvětšit jejich dostupnost pro každého. Problém spočívá ve finanční náročnosti tohoto procesu. Dalším oříškem, který se nás v budoucnosti dotkne, jsou výpočetní a telekomunikační techniky pro uchovávání veškerých informací. První hledisko se týká doby životnosti nosičů dat (diskety, CD, DVD, HDD, …). S ohledem na ohromné množství informací, které nás obklopují, je velká spotřeba těchto nosičů. Navíc životnost není nějak závratná. Podle odhadů má např. CD životnost okolo 6 let. Navíc stačí málo a data uložená na něm se poškodí nebo nenávratně ztratí. Druhé hledisko se týká důsledků akcelerovaného vývoje těchto technik. Naše společnost prožívá velkou informační evoluci. Z toho plyne, že se nástroje na uchovávání dat neustále

mění. Minulostí jsou magnetické pásky, diskety a pomalu již stávající CD. Vyvíjejí se stále nová a nová media s větší kapacitou a životností. Aby informace v dokumentech byly stále dostupné, je nutné je přenášet na nové nosiče. Bylo by tedy ku prospěchu, kdyby se vyvinula taková media, která by usnadnila tento postup.

Další vývoj se očekává i v oblasti zpracování a vyhledávání informací. Už nyní je příslibem do budoucnosti formát XML[10], kterým začínají být zpracovávaná data. Tento jazyk významně usnadňuje automatizaci zpracování obsahu informace a tím otevírá cestu k webu nové generace. Tím je sémantický web. Rozvíjí se nové standardy pro zpracování informačních zdrojů do databází. Vyvíjejí se nové služby, které uživatelům ulehčí práci při vyhledávání a zpracovávání. Jmenujme například vyhledávací funkce integrované do prohlížečů, vylepšování vyhledávacích služeb inteligentními agenty, filtrováním a vizualizací. Ještě stojí za zmínku služby hybridního typu, expanzivnější indexování nebo rozšiřování přirozeného jazyka.

V neposlední řadě bude důležitá integrace informačních zdrojů. Konečná fáze tohoto řešení zatím v dohlednu není, ale představa, že všechny důvěryhodné a důležité informace budou v jednotné databázi (nebo ve více, ale vzájemně propojených) ve standardu, jehož vlastností bude kompatibilita, univerzálnost a jednoduchost.

S rychlostí, jakou se ubírá vývoj Internetu a počítačů, si lze jen těžko představit, jak dalece se zlepší práce s informacemi. Můžeme jen odhadovat, že s novými technologiemi a normami bude proces zpracování a hodnocení několikanásobně rychlejší, využití se rozšíří do všech sfér. Předpokládá se, že v té době nás Internet obklopí úplně, a tak se stane každodenní záležitostí. Jestliže se navíc podaří zabránit kriminalitě, pomůže to k lepší stabilitě důvěrnosti informací. Tento cíl je ale dlouhodobý. V nejbližších letech se dá pouze očekávat postupné zlepšení situace.

[10] eXtensible Markup Language

8 POUŽITÁ LITERATURA

8.1 LITERATURA KE KAPITOLE 2

[1] KUČEROVÁ, Helena, Definice informace. Data – informace – znalosti, 5. 11. 2005, dostupné z WWW: <http://info.sks.cz/users/ku/UIS/inform1.htm>

[2] KUNSTOVÁ, Renáta, Informatika pro ekonomy, skripta VŠE 2003, str. 23-45

[3] ROSICKÝ, Antonín, SA_320 Informační systémy, přednášky č. 5 a 6

[4] SKLENÁK, Vilem, Data, informace, znalosti a Internet, Praha 2001, kap. 1

8.2 LITERATURA KE KAPITOLE 3

[1] KUNSTOVÁ, Renáta, Informatika pro ekonomy, skripta VŠE 2003, str. 53-68

[2] PALOVSKÝ, Radomír, Informace a Internet (včetně úvodu do protokolu Internetu), skripta VŠE 1998, kap. 1,3,4,5,6,7,8,10

[3] PetrS, Komunikační protokoly, 19. 2. 2006, dostupné z WWW: < http://www.pcsvet.cz/art/article.php?id=5242>

[4] SKLENÁK, Vilem, Data, informace, znalosti a Internet, Praha 2001, kap. 6-8

8.3 LITERATURA KE KAPITOLE 4

[1] BERKA, Petr, Dobývání znalostí z databází, Praha: Academia, 2003

[2] PLAČEK, Petr, Dolování informací na internetu, Marketingové noviny, 23. 2. 2005 dostupné z WWW: <http://www.marketingovenoviny.cz/index.php3?Action=View&ARTICLE_ID=2917>

[3] SKLENÁK, Vilem, Data, informace, znalosti a Internet, Praha 2001, kap. 9-12

[4] SPSS Inc., Služby - data mining, dostupné z WWW: <http://www.spss.cz/sl_datamining.html>

8.4 LITERATURA KE KAPITOLE 5

[1] BOLDIŠ, PETR, Jak oddělit zrno od plev: ověřování informací v prostředí internetu, INFORUM 2003, dostupné z WWW: <http://www.inforum.cz/inforum2003/prispevky/Boldis_Petr.pdf>

[2] DOUCEK P., BÉBR R., Manažerské informační systémy a jejich ekonomika, VŠE skripta, 2002, str. 10

[3] FOGG, B. J., Prominence-interpretation theory: explaining how people assess credibility online, New York, ACM Press 2003, str. 722-723., dostupné z WWW: <http://delivery.acm.org/10.1145/770000/765951/p722-fogg.pdf>

[4] FOGG, B. J., What makes web sites credible? A report on a large quantitative study, New York, ACM Press, 2001, str. 61-68., dostupné z WWW: <http://captology.stanford.edu/pdf/p61-fogg.pdf>

[5] FOGG, B. J., TSENG H., The elements of computer credibility, New York, ACM PRESS 1999, str. 80-87.,
dostupné z WWW: <http://captology.stanford.edu/pdf/p80-fogg.pdf>

[6] SKLENÁK, Vilem, Data, informace, znalosti a Internet, Praha 2001, kap. 14

[7] ŠMAHEL, DAVID, Věří lidi Internetu?, 18. 1. 2006,
dostupné z WWW: <http://www.isdn.cz/clanek.php?cid=7501>

[8] TSENG, S., FOGG, B. J., Credibility and Computing Technology, Communications of the ACM, 1999, vol. 42/5, str. 39-43,
dostupné z WWW: <http://captology.stanford.edu/pdf/p39-tseng.pdf>

8.5 LITERATURA KE KAPITOLE 6

[1] DOUCEK P., BÉBR R., Manažerské informační systémy a jejich ekonomika,
VŠE skripta, 2002, str. 53-54

[2] KUNSTOVÁ, Renáta, Informatika pro ekonomy, skripta VŠE 2003, str. 107

[3] PETERKA, J., Bezpečnost na Internetu, Softwarové noviny, leden 2001,
dostupné z WWW: < http://www.earchiv.cz/b01/b0100010.php3>

[4] ROTHFEDER, J., Nedostatek soukromí při práci na Internetu, Computerworld 1997, č. 19, str. 21-23

[5] SKLENÁK, Vilem, Data, informace, znalosti a Internet, Praha 2001, kap. 15

[6] SMEJKAL, V., Jak hlídat informace, CHIP, 1996, č.7, str. 38 – 41

Centrum informačních a knihovnických služeb

KNIHOVNA VŠE

ZÁZNAM O BAKALÁŘSKÉ PRÁCI

AUTOR	Miroslav Vlach	
NÁZEV BP	Důvěryhodnost informací na Internetu	
FAKULTA	Informatiky a statistiky	
OBOR	Informatika	
ROK OBHAJOBY	2006	
POČET STRAN	53	
POČET PŘÍLOH	0	
VEDOUCÍ BP	Ing. Antonín Rosický, CSc.	
ANOTACE	Práce pojednává o problému důvěryhodnosti informací vyskytujících se na Internetu. Cílem je poskytnout čtenáři pohled na vzájemné propojení informací, dat a znalostí. Dále se věnuje Internetu, kde a jakým způsobem se informace čerpají. V neposlední řadě se zabývá problematikou hodnocení a ochrany informací.	
KLÍČOVÁ SLOVA	informace, data, znalosti, Internet, důvěryhodnost, věrohodnost, hodnocení informací, ochrana, zabezpečení, informační zdroje	
	MÍSTO ULOŽENÍ	SIGNATURA