

ED-67/06

**Univerzita Karlova v Praze  
Filozofická fakulta  
Ústav informačních studií a knihovnictví**

Studijní program: informační studia a knihovnictví  
Studijní obor: informační studia a knihovnictví

**Martina Tatranská**

**Webová agregace zpráv : současné světové trendy**

Diplomová práce

Praha 2006

Vedoucí diplomové práce:

PhDr. Richard Papík, Ph.D.

Oponent diplomové práce:

Ing. Martin Švec

Datum obhajoby:

22 ledna 2007

Hodnocení:

vyhověl(1)

## **ZADÁNÍ DIPLOMOVÉ PRÁCE** (PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

pro Martina Tatranská

obor Informační studia a knihovnictví

Název tématu: Webová agregace zpráv : současné světové trendy

### Zásady pro vypracování:

Cílem diplomové práce je definice a základní analýza způsobů agregace webových zpráv.

Předběžná struktura práce:

1. Kvantitativní nárůst informací ve webovém prostředí
2. Příležitosti a metody agregace zpráv
3. Agregace zpráv konkrétními produkty (Google news, Topix.net, Newsbot, Jyxo, Prave dnes aj.)
4. Porovnání systémů agregace zpráv
5. Trendy vývoje agregačních systémů

Diplomová práce bude připravena a upravena v souladu s platnými vnitřními předpisy FF UK a dalšími metodickými pokyny a normativními dokumenty.

Rozsah grafických prací:

Rozsah průvodní zprávy:

Seznam odborné literatury:

1. KENNEDY, Dennis M. Beating Information Overload with News Aggregators. *Law Practice* [online]. 2003, vol. 29, is. 8 [cit. 2006-05-30]. Dostupný z WWW: <<http://www.abanet.org/lpm/magazine/articles/v29is8an4.shtml>>.
2. MADNICK, Stuart, et al. *Surviving and Thriving in the New World of Web Aggregators* [online]. Cambridge (MA) : Sloan School of Management, Massachusetts Institute of Technology, 2000 [cit. 2006-05-30]. 21 s. Dostupný z WWW: <<http://ebusiness.mit.edu/research/papers/Aggregator%20paper%2010-22-00%20SEM-v20%20FINAL.pdf>>.
3. ZHU, Hongwei; SIEGEL, Michael D.; MADNICK, Stuart E. *Information Aggregation – A Value-added E-Service*. Cambridge (MA) : Sloan School of Management, Massachusetts Institute of Technology, 2001 [cit. 2006-05-30]. 12 s. Dostupný z WWW: <<http://ebusiness.mit.edu/research/papers/106%20SMadnick,%20Siegel%20Information%20Aggregation.pdf>>.

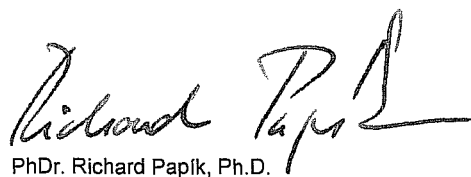


Vedoucí diplomové práce: PhDr. Richard Papík, Ph.D.

Datum zadání diplomové práce: 30.6.2006

Termín odevzdání diplomové práce:

L.S.



PhDr. Richard Papík, Ph.D.

.....  
Vedoucí součásti-ředitel ÚISK FF UK

.....  
Děkan FF UK

V Praze dne 30.6.2006

**Prohlášení:**

Prohlašuji, že jsem diplomovou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje.

V Praze, 15. prosince 2006

A handwritten signature in black ink, appearing to read 'Martina Václavová', written over a horizontal dotted line.

podpis diplomanta

## **Identifikační záznam**

TATRANSKÁ, Martina. *Webová agregace zpráv : současné světové trendy [Web News Aggregation : current world trends]*, Praha, 2006. 81 s. Diplomová práce. Univerzita Karlova v Praze, Filozofická fakulta, Ústav informačních studií a knihovnictví 2005. Vedoucí práce PhDr. Richard Papík, Ph.D.

## **Abstrakt**

Diplomová práce se zabývá agregací zpráv neboli monitoringem médií od jejího vzniku až po současnost. Cílem práce je uvedení do problematiky agregace zpráv a analýza nejvýznamnějších systémů zabývajících se elektronickým monitoringem médií.

V první části práce jsou vymezeny pojmy týkající se této oblasti a následně nastínění problematiky kvantitativního nárůstu informací, který se odráží i při tvorbě zpravodajských informací, jejichž základní vlastnosti jsou v práci popsány.

Druhá část práce více definuje činnost agregace zpráv, její historický vývoj vycházející z výstřižkových služeb, jak ve světě tak i v českých podmínkách, typologii a způsoby užití. Nejsou opomenuty formáty pro poskytování obsahu, jejich stručná historie a význam v agregaci zpráv, největší důraz je zde kladen na RSS kanály.

Další část práce je věnována analýze vybraných významných systémů.

V závěru jsou popsány trendy vývoje agregace zpráv a jejich možné perspektivy.

## **Klíčová slova**

agregace zpráv, webová agregace, monitoring médií, agregátory zpráv, RSS, výstřižková služba

PŘEDMLUVA .....	8
<b>1 ÚVOD.....</b>	<b>10</b>
1.1 VYMEZENÍ POJMŮ .....	11
<b>2 KVANTITATIVNÍ NÁRŮST INFORMACÍ VE WEBOVÉM PROSTŘEDÍ .....</b>	<b>12</b>
2.1 INTERNET JAKO PROSTŘEDEK ŽURNALISTIKY .....	13
2.2 TŘI ZÁKLADNÍ VLASTNOSTI INTERNETOVÉ ŽURNALISTIKY .....	13
2.2.1 <i>Hypertextualita</i> .....	14
2.2.2 <i>Multimedialita</i> .....	14
2.2.3 <i>Interaktivita</i> .....	15
2.3 ORIENTACE V INFORMAČNÍCH ZDROJÍCH .....	15
2.4 KONTROLA NAD INFORMAČNÍMI ZDROJI.....	18
<b>3 AGREGACE ZPRÁV.....</b>	<b>20</b>
3.1 DEFINICE AGREGACE ZPRÁV.....	20
3.2 HISTORIE AGREGAČNÍCH SLUŽEB NEBOLI OD VÝSTŘIŽKOVÉ SLUŽBY PO SOUČASNOST.....	20
3.2.1 <i>Počátky agregace zpráv ve světě</i> .....	21
3.2.2 <i>Počátky agregace zpráv v České republice</i> .....	21
3.3 RSS A JEHO VÝZNAM V AGREGACI ZPRÁV.....	24
3.3.2 <i>Význam RSS v agregaci zpráv</i> .....	27
3.4 TYPOLOGIE SYSTÉMŮ AGREGACE ZPRÁV.....	28
3.4.1 <i>Automaticky generovaný obsah</i> .....	28
3.4.2 <i>Obsah vytvářený uživateli</i> .....	29
3.5 TYPOLOGIE ZPŮSOBŮ UŽITÍ AGREGACE ZPRÁV .....	31
3.5.1 <i>Sledování přehledu článků</i> .....	31
3.5.2 <i>Volba kategorie zájmů</i> .....	31
3.5.3 <i>Definice vlastních kategorií (témat) neboli personalizace</i> .....	32
3.5.4 <i>Emailové zasilání zpráv neboli alertní služby</i> .....	32
3.5.5 <i>RSS čtečky</i> .....	32
3.6 METODOLOGIE TVORBY A ARCHITEKTURA WEBOVÝCH AGREGÁTORŮ ZPRÁV.....	33
3.6.1 <i>Základní metodologie webové agregace zpráv u jednodušších systémů</i> .....	33
3.6.2 <i>Základní metodologie webové agregace zpráv u propracovanějších systémů</i> .....	34
3.7 SOUČASNÉ SVĚTOVÉ TRENDY AGREGACE ZPRÁV .....	36
<b>4 VÝBĚR JEDNOTLIVÝCH SYSTÉMŮ A JEJICH CHARAKTERISTIKA.....</b>	<b>38</b>
4.1 SYSTÉMY ZDARMA PŘÍSTUPNÉ VE SVĚTĚ .....	38
4.1.1 <i>Google News</i> .....	38
4.1.1.1 <i>Newsmap</i> .....	40
4.1.2 <i>Topix.net</i> .....	41
4.1.3 <i>Inform.com</i> .....	44
4.1.4 <i>NewsNow</i> .....	46
4.2 KOMERČNÍ SYSTÉMY VE SVĚTĚ .....	49
4.2.1 <i>Factiva</i> .....	49
4.2.2 <i>Thomson Business Intelligence</i> .....	54
4.2.3 <i>Moreover</i> .....	56
4.3 SYSTÉMY ZDARMA PŘÍSTUPNÉ V ČESKÉ REPUBLICE .....	57
4.3.1 <i>Články.cz</i> .....	57
4.3.2 <i>Nový den</i> .....	58
4.3.3 <i>Krátce.cz</i> .....	60
4.4 KOMERČNÍ SYSTÉMY V ČESKÉ REPUBLICE.....	61
4.4.1 <i>NewsTin.com</i> .....	61
4.4.3 <i>Newton I.T.</i> .....	67
<b>5 ZÁVĚR .....</b>	<b>76</b>
<b>SEZNAM POUŽITÉ LITERATURY .....</b>	<b>78</b>

## PŘEDMLUVA

Předkládaná diplomová práce je zaměřena na teoretické a praktické aspekty webové agregace zpráv.

Pro výběr tohoto tématu jsem se rozhodla díky možnosti pracovat ve společnosti *I2S*, která se mimo jiné zabývá tvorbou agregátora zpráv *NewsTin.com*. Tematika organizace informací prostřednictvím agregátorů zpráv mě zaujala hned od počátku a na základě rozhovoru s ředitelem *NewsTin Technologies* D. Čaňkem jsem se rozhodla ji realizovat ve své diplomové práci. Mé rozhodnutí umocnila skutečnost, že žádná podobná práce, která by se zabývala vývojem agregace zpráv a jednotlivými systémy, ještě na *ÚISK* nebyla zpracována a možná ani v ČR, přestože technologie je na vzestupu.

Vzhledem k šíři tématu nebylo možné provést analýzu všech systémů poskytujících agregační služby, proto byly vybrány pouze nejvýznamnější a nejvyužívanější zdroje.

Při zpracování jsem vycházela zejména z volně dostupných zdrojů na internetu (zvláště v rámci vyhledávače *Google Scholar*) a z výběru dokumentů z polytematických a oborových databází Univerzity Karlovy.

Neméně důležitým okruhem zdrojů byly mé vlastní zkušenosti z praxe a možnost konzultace s odborníky z oboru. Některé informace byly získány na základě dotazování prostřednictvím elektronické pošty a v jednom případě i formou osobní schůzky. Jednalo se o dodatečné informace k některým českým agregátorům zpráv a k historii výstřižkové služby.

Použitými materiály jsou zejména časopisecké články, výzkumné zprávy, dokumenty typu *white paper* a samotné webové stránky jednotlivých systémů.

Použité informační zdroje jsou citovány v souladu s mezinárodními normami ČSN ISO 690 a ČSN ISO 690-2. Citování v textu je vytvořeno podle Harvardského systému, který je doporučován normou ČSN ISO 7144:1986 vytvořenou pro dokumentaci a formální úpravu disertací a podobných dokumentů. Citace v textu jsou uvedeny v hranatých závorkách a obsahují příjmení autora nebo první slova jména korporace, popřípadě úvodní slova názvu dokumentu a dále rok vydání a stránkování, pokud je uvedeno.



Za poskytnutí cenných připomínek a konzultací bych ráda poděkovala vedoucímu mé práce řediteli ÚISK PhDr. Richardu Papíkovi, Ph.D a za úvodní podněty rovněž řediteli *I2S/NewsTin Technologies* Davidu Čaňkovi.

# 1 ÚVOD

Cílem této práce je seznámení s problematikou agregace zpráv a analýza nejvýznamnějších systémů zabývajících se elektronickým monitoringem médií.

Problematiku kvantitativního nárůstu informací, jež nabyla na významu ve 2. polovině 20. století s rozvojem publikačních aktivit, logicky následovaly pokusy o organizaci dat a reprezentaci znalostí. Systémy agregace zpráv jsou jedněmi z nich.

Diplomová práce je rozdělena do pěti kapitol. Následující kapitola se zabývá nárůstem informací, který se odráží i při tvorbě zpravodajských informací, jejichž základní vlastnosti jsou též popsány.

Ve třetí kapitole je více definována činnost agregace zpráv, její historický vývoj vycházející z výstřižkových služeb, jak ve světě tak i v českých podmínkách, typologie a způsoby užití. Nejsou opomenuty formáty pro poskytování obsahu, jejich stručná historie a význam v agregaci zpráv, největší důraz je zde kladen na RSS kanály.

Čtvrtá kapitola předkládá charakteristiku vybraných světových a českých systémů, jak komerčních tak i zdarma přístupných.

Poslední část uzavírá práci s popisem perspektiv vývoje agregačních systémů zpráv.

## 1.1 Vymezení pojmů

V anglosaské odborné literatuře existuje celá řada pojmů vztahujících se k agregaci zpráv. Užívá se zde mnoho synonymních výrazů, kterými se pojmenovávají systémy této činnosti nebo samotná činnost jako taková.

Co se týká názvů pro tuto činnost, za nejrozšířenější termín lze považovat:

➤ **monitoring médií (*media monitoring*)**, méně už **monitoring zpráv** či **monitoring zpravodajství (*news monitoring*)**

Dalším hojně užívaným termínem je:

➤ **agregace zpráv (*news aggregation*)** – český termín není ještě velmi zažitý; zprávami (*news*) jsou zde míněny zpravodajské články

➤ **webová agregace (*web aggregation*)** – je také užívaným termínem, který ale může být zaměňován s jinými typy agregace zpráv, než jsou právě zpravodajské obsahy

➤ **elektronická výstřižková služba (*electronic clipping service* ještě více se užívá pouze *clipping service*)** – tento termín vychází z klasické výstřižkové služby, která byla předchůdcem dnešní elektronické agregace zpráv, v angličtině je tento termín dodnes hojně užíván v kontextu moderní elektronické výstřižkové služby

### Další pojmy

Pro samotné označení systému agregace zpráv se nejvíce užívá, opět více v anglosaské literatuře než v české, pojem:

➤ **agregátor zpráv** či ještě méně užívaný **agregátor zpravodajství (*news aggregator*)**

➤ **RSS kanály (*RSS feeds*)** – technologie užívaná pro agregaci zpráv

## 2 KVANTITATIVNÍ NÁRŮST INFORMACÍ VE WEBOVÉM PROSTŘEDÍ

„Getting information off the Internet is like taking a drink from a fire hydrant“<sup>1</sup>  
Mitchell Kapor

Zhruba od poloviny minulého století se začaly používat termíny informační exploze, informační přetížení jako reakce na čím dál rychlejší nárůst informačních zdrojů, které nebylo možné pojmout, zpracovat a využívat.

V průběhu 2. světové války dostaly výzkumné aktivity mnoha oborů značné impulsy, které byly důsledkem získání strategické a vojenské převahy nad protivníkem, a tento stav byl promítnut i do studené války. Dynamický rozvoj výzkumných aktivit se zákonitě projevil kvantitativním nárůstem informací a znalostí. Zároveň se modernizovaly prostředky zpracování, ukládání a vyhledávání informací. Tento fakt ovlivnil i publikační chování vědeckých komunit.

Výrazný posun byl zaznamenán až se vznikem celosvětového síťového propojení *world wide web*, resp. Internetu.

Se vznikem tohoto nového komunikačního média a jeho masivním rozšířením v 90. letech vznikl nový prostor pro publikační aktivity, které se tímto dostávají na stejnou úroveň ostatních médií masové komunikace.

Internet mění paradigma tradičních masových médií minimálně v tom, že zeslabuje moc organizovaného a profesionálního komunikátora, který až do příchodu Internetu bez problémů mohl dohlížet na obsah i způsob distribuce komunikovaného sdělení. Prostředí internetu je ovlivňováno chováním uživatele. O tom, zda se určitá stránka stane sdělením masové komunikace, nerozhoduje pouze komunikátor, který ji vytváří, nebo systém, který mediální produkt (noviny, časopisy, televizní a rozhlasové pořady) distribuuje, ale zejména příjemce sdělení, který si danou webovou stránku dohledá a začne ji pravidelně navštěvovat [ŠMÍD, 2006, s.7].

---

<sup>1</sup> Získávat informace z internetu je jako pít z požárního hydrantu.

## 2.1 Internet jako prostředek žurnalistiky

Technologický rozvoj v oblasti počítačů a informační vědy ovlivňoval žurnalistiku již v 80. letech a to zejména ve vyspělých průmyslových zemích, kde novináři používali textové editory a využívali online přístup k placeným databázím. Změna nastala v polovině 90. let rozšířením webového rozhraní *world wide web* zejména díky uživatelsky přívětivým aplikacím. Tímto se nejen zpřístupnila internetová technologie počítačovým laikům, ale zejména se výrazným způsobem začal šířit počet informačních zdrojů umístěných v tomto kyberprostoru [KOCH, 1991].

Tato událost ovlivnila žurnalistiku ve dvou směrech. Na straně jedné se internet stal nezbytným nástrojem žurnalistické práce. Novináři získali snadnější přístup k nejrůznějším archivům a databázím, ale také k volně přístupným webovým stránkám, kde byla orientace usnadněna vznikem katalogových a fulltextových vyhledávačů.

Na straně druhé komunikace prostřednictvím internetu zajistila konektivitu mezi žurnalistou a zdrojem, mezi žurnalistou a jeho redakcí a také mezi žurnalistou a čtenáři v případě, že jeho žurnalistický produkt byl na internetu umístěn. Tento nový model žurnalistické práce anglosaská jazyková oblast nazývá **CAR** – *Computer-Assisted Reporting*, počítačem asistovaná žurnalistika.

## 2.2 Tři základní vlastnosti internetové žurnalistiky

Publikování na internetu je ovlivněno třemi základními vlastnostmi internetové žurnalistiky, které vytvářejí přidanou hodnotu online žurnalistiky. Jsou jimi:

- **hypertextualita**
- **multimedialita**
- **interaktivita**

[ŠMÍD, 2006, s.10-11]

### 2.2.1 *Hypertextualita*

Hypertextualita se může v prostředí informačních sítí využívat např. pro **glosované zpravodajství** (*annotative reporting*) uplatňující odkazy k dalším informačním zdrojům či poznámkám [DEUZE, 2001].

Nebo může sloužit k **archivování materiálů** (*archiving ability*), které lze díky vyhledávacím strojům a hypertextu v případě potřeby zpětně vyvolat a dodatečně se s nimi seznámit [WARD, 2002]. I když zde vyvstává otázka problematiky proměnlivosti *URL* odkazů (*Uniform Resource Locator*). Touto problematikou se zabývala řada odborníků a výsledkem jejich snažení je vznik různých dalších identifikátorů, které by mohly mít trvalejší charakter, jedním z nich je systém *DOI* (*Digital Object Identifier*).

Na rozdíl od tradičního textu je hypertext proměnlivou veličinou, kterou si aktivuje každý uživatel sám způsobem, jakým ho řadí a „konzumuje“. Každé jednotlivé sdělení na internetu užívající hypertext je pak originální v té konfiguraci, kterou si uspořádal uživatel sítě. Tento způsob omezuje někdejší absolutní autorskou kontrolu nad textem, neboť autorovo sdělení je možné rozšiřovat o další obsahy, jež nejsou jeho. Některé z nich se samozřejmě mohou stavět i opačně proti jeho záměru a smyslu, jak se lze ostatně přesvědčit při sledování internetových diskusí k publikovaným článkům.

Hypertext je rovněž příkladem informačního systému přehlaceného podněty. Je proto třeba k němu přistupovat s jistou vyhledávací strategií a nenechat se dezorientovat [FOSS, 1989, s. 407-411]. Chytře odpovídá přirozenosti lidského myšlení asociovat další myšlenky, může však způsobit psychické problémy při technice zvané *browsing* (prohlížení) proti cílenému strategickému vyhledávání informací [PAPÍK, 2001, s. 81-90].

### 2.2.2 *Multimedialita*

Multimedialita umožňuje rozdělit zprávu do souborů různých formátů, tedy nejen textových, ale i zvukových, obrazových a audiovizuálních, čímž zpráva získává na své lákavosti a věrohodnosti. Jako příklad můžeme zmínit nedávno vzniklý

ambiciózní server *Aktuálně.cz* (jeden z mála, který není napojen na zpravodajství tradičního média), který se snaží využívat všechny zmíněné datové soubory.

Na internetu existují ale i jiné informační portály, u jejichž základu nestojí tisková média, ale rozhlasová nebo televizní. Například populární a hodně navštěvované servery *BBC* a *CNN*.

### 2.2.3 *Interaktivita*

Interaktivitou můžeme chápat **navigační interaktivitu**, kterou si uživatel vybírá a stanovuje pořadí webových stránek. Zároveň to může být **funkční interaktivita** vycházející z aktivní účasti v diskusních skupinách a fórech k publikovaným zprávám. Ale také **interaktivita adaptivní**, při níž si uživatel personalizuje svůj prohlížeč, počítač, software tak, jak to vyhovuje jeho specifickým potřebám a zájmům (oblíbené položky, *RSS* kanály aj.) [DEUZE, 2005, s.203-230].

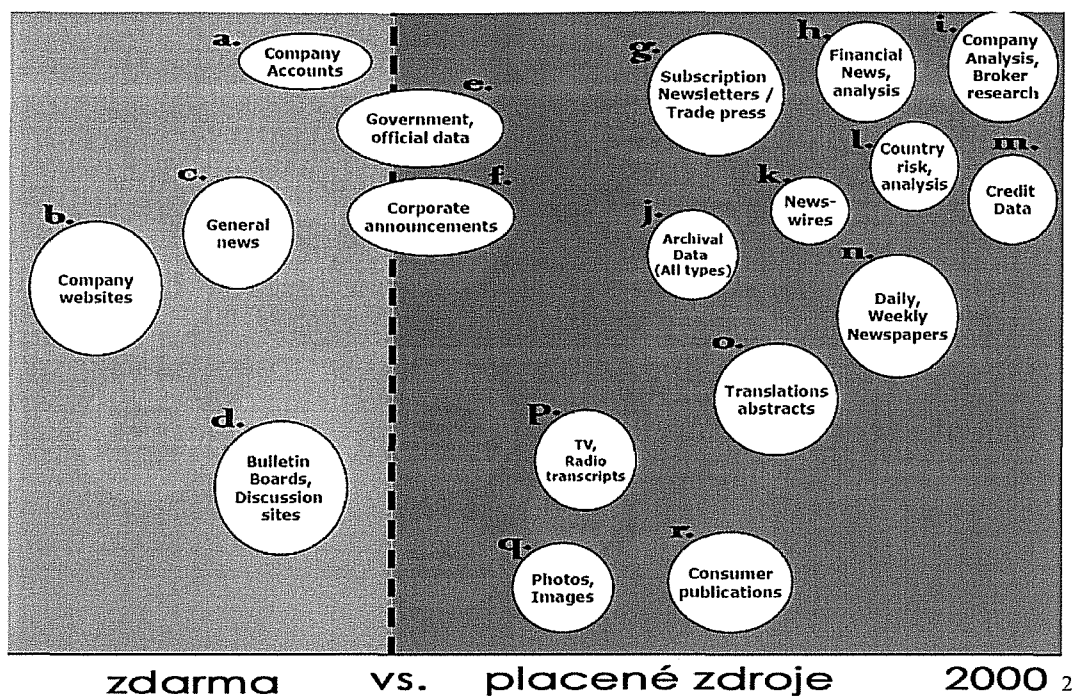
Důležité je také zmínit bezprostřednost – jako další z hlavních charakteristik publikování na internetu, která umožňuje okamžitou reakci na aktuální událost a jejíž sdělení je možné umístit bezprostředně na internet, aniž by bylo třeba využívat výrobní nebo distribuční organizaci.

## 2.3 **Orientace v informačních zdrojích**

Exponenciální nárůst informačních zdrojů zapříčinil nečekané zvýšení počtu informací, vycházejících z množství informačních kanálů, včetně internetu. Dřívější potřeba zpráv zprostředkovaných tiskem, typická pro průmyslové společnosti 19. a 20. století, je dnes přesycena kvantitou rozhlasových a televizních programů, elektronických informačních zdrojů, bezplatných inzertních novin, elektronické inzerce či nevyžádané pošty (*spam*).

Dříve byl rovněž větší kontrast placených webových služeb vůči zdarma přístupným. Nejlépe to je možné vidět na obrázcích 1 a 2, které srovnávají tento poměr v rozdílu pouhých 5 let, v roce 2000 a 2005.

Obr.1: Zdarma versus placené informační zdroje na internetu v roce 2000



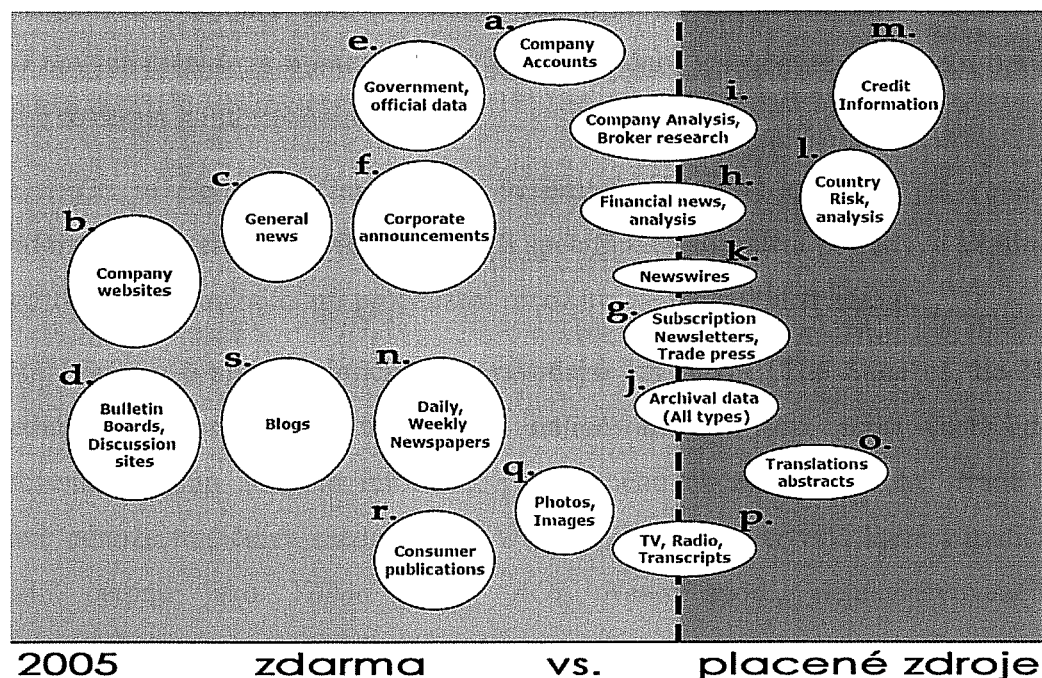
[SEFTON-WILLIAMS, 2005, s.4]

2

- a. firemní účetnictví
- b. firemní stránky
- c. obecné zpravodajství
- d. elektronické nástěnky, diskusní skupiny
- e. vládní informace
- f. firemní oznámení
- g. předplacená data, newslettery, obchodní tisk
- h. finanční zpravodajství, analýzy
- i. firemní analýzy, brokerské zprávy
- j. archivní data
- k. zpravodajské agentury
- l. analýzy rizika země
- m. kreditní informace
- n. deníky, týdeníky
- o. překlady, abstrakta
- p. televizní a rozhlasové záznamy
- q. fotografie, obrázky
- r. publikace pro spotřebitele
- s. blogy



Obr. 2: Zdarma versus placené informační zdroje na internetu v roce 2005



[SEFTON-WILLIAMS, 2005, s.4]<sup>2</sup>

Již tedy není problém se k informacím dostat, jak tomu bylo dříve, ale problematické začalo být se v nich orientovat.

Dle posledních výzkumů amerického projektu *Pew Internet and American Life Project*, 29 % uživatelů internetu – kolem 43 milionů lidí čte online zpravodajství minimálně třikrát za týden [HORRIGAN, 2006].

Výzkum byl proveden v americkém prostředí, ale nemá daleko k realitě ani v našich podmínkách. Na internetu existují miliony zpravodajských zdrojů, které mají své sourozence v tištěné verzi, ale také, které byly zrozeny přímo na internetu. Největší kvantitativní nárůst lze pozorovat u blogů. Blogová sféra se dokázala zdvojnásobit za pouhých 6 měsíců (od května 2005 do listopadu 2005) a je odhadováno, že denně je na světě vytvořeno 100 000 nových blogů [MOTZ, 2006]. Pro příklad jenom server *Technorati*<sup>3</sup> v současné době agreguje články z 27,2 milionů blogů.

Dnes opravdu není složité se dostat k plným textům světových periodik a jiným informačním zdrojům k nimž je rychlý a okamžitý přístup.

<sup>3</sup> <http://www.technorati.com>

Pro další příklad je možné navštívit zdarma přístupný agregátor zpráv *Pubsub.com*<sup>4</sup> čerpající zprávy z téměř 6,7 milionů zdrojů, které poskytují přes 800 nových článků za minutu. Jak už bylo zmíněno, v současné době není opravdu těžké se dostat k informačním zdrojům.

Co se ale stává čím dál tím větším problémem, je vědět, jak se dostat k informacím, které se nejvíce týkají naší specifické informační potřeby, a eliminovat tak ztrátu času vznikající při procházení množství nerelevantních zdrojů.

Zpráva o stavu amerických zpravodajských médií [The State of the News Media, 2006] dnes konstatuje, že online média jsou stále častějším a frekventovanějším zdrojem informací pro běžného občana. Na druhé straně jsou právě odnože tradičních médií jako *CNN*, *MSNBC* nebo servery, které z tradičních médií čerpají jako *Yahoo News*, *AOL News* apod., nejnavštěvovanějšími informačními servery. A to zejména proto, že tyto systémy výrazně usnadňují orientaci ve zpravodajství a dovolují uživatelům kontrolovat odkud a jak dostanou své vybrané zprávy, což také činí ve stále větší míře.

## 2.4 Kontrola nad informačními zdroji

Hodnota informačního systému je dána kvalitou a kvantitou obsažených informací, ale zároveň je stejně důležité nalézt jednoduchým postupem (uživatelsky přívětivém) relevantní informaci

Každý člověk má zažité návyky pro získání informací. Může to být například čtení:

- emailové korespondence (osobních a pracovních zpráv)
- nevyžádaných zpráv (spamů)
- konferenčních, informačních bulletinů
- oblíbených webových stran
- elektronických diskusí
- zpravodajských zdrojů

---

<sup>4</sup> <http://www.pubsub.com>

- dalších periodik (magazíny, brožury) atd.

Tímto vším může být uživatel zahlcen (*information overload*). Přemírou informací u něj může nastat pocit frustrace či stresu. Informační exploze či přetížení, nejsou už jen frázemi, ale čistou realitou. Zdá se paradoxem, že množství informací, které jsou denně vnímány uživateli, kvantitativně vzrostlo, ale kontrola nad nimi výrazně klesla.

Otázka, jak efektivně nakládat s masivním nárůstem informací, se stává výzvou jedinců i organizací.

### 3 AGREGACE ZPRÁV

„Don't search for your news, just read it.“<sup>5</sup>

František Vrábek, zakladatel NewsTin.com

Existují však způsoby řešení. Jedním z nich jsou **agregátory zpráv**. Mnoho lidí dnes nachází útočiště před neúměrným nárůstem informací právě v agregačních službách. Ve světě za několik posledních let vzniklo již několik systémů, které tyto služby komerčně či zdarma poskytují. Pomáhají uživatelům nacházet požadované informační kanály na jednom místě, aniž by uživatelé museli vynakládat příliš mnoho úsilí a času. Takové systémy se nazývají webovými agregátory.

#### 3.1 Definice agregace zpráv

Systémy agregace zpráv, neboli agregátory zpráv, nevytvářejí a nejsou autory vlastního obsahu.

Nabízejí relevantní informace seřazené do kategorií excerpovaných z více zdrojů na základě jejich analýzy a přidané hodnoty. Pomáhají tak jednotlivcům či korporacím efektivně využívat rostoucí množství informací na internetu.

Některé agregátory zpráv také poskytují služby s přidanou hodnotou, jako zasílání zpráv, dle předem nadefinovaných kategorií či témat elektronickou poštou nebo jsou uživateli zobrazovány v rámci jeho personalizované stránky.

#### 3.2 Historie agregačních služeb neboli od výstřižkové služby po současnost

Historie agregace zpráv sahá do doby ještě před vznikem Internetu. Prvopočátky této služby můžeme nalézt už u výstřižkových služeb nabízené knihovnami či tiskovými agenturami, které pro své zákazníky fyzicky vystřihávaly konkrétní články z novinových a časopiseckých titulů na základě zadaného tématu nebo sestavovaly jejich přehledy či připravovaly jednorázové rešerše.

---

<sup>5</sup> Nehledej své zprávy, pouze je čti.

### 3.2.1 Počátky agregace zpráv ve světě

Na konci 80. let 19. století knihovníci na *Clemson University* v Jižní Karolíně začali s využíváním metody výstřižkové služby na ochranu a organizaci informací o státu Jižní Karolína. Vybrané články byly vystřižené z místních novin, oštitkované ozdobným písmem, datem a názvem novin, nalepené na list papíru a rozdělené do složek dle předmětového třídění. Nejstarší články jsou dnes zažloutlé a rozpadají se. Indikují tak, kolik času od té doby již uplynulo [HENDERSON, 2001, s. 17-18].

V uvedeném případě jedné z prvních výstřižkových služeb se mapovalo pouze jedno specifické téma. Vzniklo ale i množství dalších systémů, které byly většinou zaměřeny polytematicky. V případě větších organizací bylo zapotřebí na tisíc zaměstnanců, aby mohly být uspokojeny požadavky klientů po rychlém a spolehlivém dodání dokumentů, které se vztahovaly k jejich firmě, konkurentům nebo k vybraným relevantním trhům. Přítomnost internetu v dnešní době tyto postupy mnohem usnadnila. Téměř všechny procesy, na které bylo potřeba velké množství pracovníků, se automatizovaly.

Začaly postupně vznikat také nové formáty sloužící pro interakci mezi daty, které výrazně napomohly webové agregaci zpráv. Více viz kapitola o RSS (3.3), která podrobně rozebírá historii vzniku těchto formátů.

### 3.2.2 Počátky agregace zpráv v České republice

I v České republice si lidé uvědomili, že není možné pojmout veškerý objem tištěných médií tak, aby neunikla žádná důležitá zpráva. A tak i zde historicky nejstarší podobou agregačních služeb byly služby výstřižkového typu, které jako první začaly zprostředkovávat pravidelné dodávání relevantních zpráv dle přání uživatele. Pro své klienty „vystřihávaly“ tematické články z novinových a časopiseckých titulů, nebo sestavovaly jejich přehledy, či jednorázové rešerše. Výstřižkové služby byly nejčastěji nabízeny tiskovými agenturami nebo knihovnami.

Výstřižková služba má v České republice dlouhou tradici. Již v roce 1929 vznikla v Praze filiálka středoevropské výstřižkové kanceláře *ARGUS*, která byla

zrušena „pro nepřátelský postoj k lidově demokratickému zřízení“ [Služba vystřihuje..., 1997].

V roce 1945 zřídilo tehdejší *Ministerstvo informací a osvěty* informační výstřižkovou službu **ORBIS**. Nová výstřižková služba v konkurenci s *ARGUSEm* obstála, a tak o tři roky později mohla převzít jeho funkci. Tak byl položen základ **Pražské informační služby**<sup>6</sup> (**PIS**), která dodnes poskytuje klasickou výstřižkovou službu.

Stejně tak jako dal rozvoj tištěných médií podnět ke vzniku klasické výstřižkové služby, tak si žádal i rozvoj elektronických médií vznik obdobného systému v digitální podobě.

**Začátky elektronické výstřižkové služby v České republice** sahají do roku 1996, kdy se na trhu objevila společnost *Newton IT*. O dva roky později byla založena společnost *Anopress* (nyní *Anopress IT*), která působí jako dodavatel monitoringu médií, a je navíc obohacena o retrospektivní data od společnosti *Credittax*, která se dříve monitoringem zabývala, ale v té době od něho upustila. *Anopress IT* je nyní součástí společnosti *Creditinfo Czech Republic*<sup>7</sup>.

Na počátku internetu byl stále předmětem monitorování především klasický „tištěný“ tisk, kterým se zabývaly zejména dvě zmíněné hlavní společnosti *Anopress* a *Newton IT*, jež v té době ještě neměly žádnou webovou platformu.

Ve stejném roce vzniku prvotní monitorovací služby (*Anopress*) bylo možné na internetu spatřit první online systém *Trafika*<sup>8</sup>, který nabízel jednotný přístup k více zpravodajským zdrojům (retrospektivně až do roku 1994) dokonce ještě před vznikem samotných zpravodajských serverů, jak je známe dnes. *Trafika* většinu papírových novin převáděla do digitální podoby a prezentovala na svých stránkách.

Ve své době byla *Trafika* opravdu revoluční, poskytovala plné a včasné obsahy všech hlavních českých deníků a rychlý přístup, kterému okolní státy či dokonce USA mohly jen těžko konkurovat. Bylo možné v ní vyhledávat, její vyhledávání bylo na velmi slušné úrovni, či také personalizovat [HLAVENKA, 2000].

---

<sup>6</sup> <http://www.pis.cz>

<sup>7</sup> <http://www.cicr.cz>

<sup>8</sup> Produkovala internetová společnost M.I.A na adrese [www.trafika.cz](http://www.trafika.cz), dnes již neexistuje.

Tento „internetový stánek“ umožnil zrod elektronické verze *Mladé Fronty Dnes* či *Lidových novin*. Vydavatelé těchto novin měli ale pochopitelně své záměry, proto elektronické verze později přesunuly na vlastní webové prostory.

*Trafika* měla naději stát se jedním z nevýznamnějších českých portálů, ale bohužel se po svém „velkém“ startu již dále nevyvíjela [PUMERA, 1999].

Obr.3 : Podoba dnes již neexistujícího portálu *Trafika*.



Teoreticky prvním agregátorem mohl být, dnes již rovněž neexistující, portál *Sféra.cz*. Více než agregátorem se *Sféra* snažila být originální (už jenom svojí 3D průvodkyní Anetou), o čemž svědčily její webové stránky (obr.4), které svým designem připomínaly spíše bulvární zpravodajský server. Články byly na server vybírány manuálně a rovněž tak byla ke každému z nich vytvořena žertovná anotace (nezávisle na serióznosti tématu). Ty pak odkazovaly na původní články. Uživatel mohl rovněž využít možnost automatické personalizace. Ve své době byla *Sféra* jistě průkopníkem mezi modely českých internetových serverů. [BARTOŇ, 2000]

Obr. 4: Sfera.cz – také již neexistuje



Nástup skutečných agregátorů, tak jak je známe dnes, přišel až s prvním využitím RSS kanálů. Jeden z prvních, kdo přivedl do České republiky formát RSS, byl v roce 2000 Daniel Dočekal, autor portálu *Svět namodro*. O rok později již vznikla první opravdová a dodnes využívaná webová agregace zpráv, *Právédnes.cz* (viz kap. 4.4.2).

### 3.3 RSS a jeho význam v agregaci zpráv

#### 3.3.1 Definice a vznik RSS



**RSS (Really Simple Syndication)** je formát pro poskytování obsahu založený na **XML schématu**. RSS umožňuje **sledování informací z jednoho místa** zveřejňovaných na různých serverech, a tím uživatele ušetří pravidelného navštěvování množství webových zdrojů. Všechny nejnovější zprávy je tak možné mít pohodlně a rychle na jediném místě.

Nedlouho po vzniku Internetu začali tvůrci prvních webových stránek řešit jak zajistit vzájemné provázání stránek na různých serverech. Uvědomili si, že z hlediska propagace stránek je výhodné spolupracovat s tvůrci podobných obsahů a vzájemně na sebe odkazovat. Jeden z prvních nápadů, a také technicky nejjednodušších, byla výměna log či textových odkazů, jak například můžeme vidět na stránkách



elektronického časopisu *Ikaros*<sup>9</sup>, který odkazuje na svého slovenského kolegu *ITlib*<sup>10</sup> zveřejněním jeho ikony na hlavní stránce a stejně tak činí na svých stránkách i slovenský časopis. Tento klasický způsob nám však nepodává aktuální informaci o tom, co se na daném serveru právě nachází.

Tuto nedokonalost vyřešily tzv. dynamické reklamy, které měly tu vlastnost, že se měnily v závislosti na aktuálním obsahu propagovaného serveru.

Toto řešení nebylo však natolik invenční jako myšlenka, se kterou přišel v roce 1997 **Dave Winer** a pojmenoval ji *Scripting News*. Poměrně jednoduchým způsobem použil *XML* formát pro přenos dat, do té doby se používal pouze text zabalený do *HTML* nebo obrázky ve formátu *GIF*. Oba tyto formáty nahradil stále se rozšiřujícím *XML*. Systém byl vytvořen zejména pro distribuci zpravodajských titulků a nebyť počínou společnosti *Netscape*, dnes by se o něm asi již nevědělo.

*Netscape* o dva roky později uvedl systém *RSS (0.9)*, ve kterém využil i Winerův *Scripting News*. *RSS* bylo původně vytvořeno pro integraci zpráv do portálu *my.netscape.com*, jak se však později ukázalo, jeho potenciál byl mnohem větší [BEDNÁŘ, 2004].

Podobnou myšlenku měla svého času i společnost *Microsoft* se svou technologií *CDF (Channel Definition Format)*, která se ale zásadním způsobem neprosadila. Obě technologie umožňují přebírat obsah zdrojů na internetu a v přehledné formě je nabízet uživatelům.

Prosadil se tedy *RSS*, soubor typu *XML*, který obsahuje seznam článků na určitém serveru. Pro každý článek je uveden název, *URL* adresa a stručná anotace. Některé verze *RSS* podporují další položky jako například datum vložení nebo aktualizace.

*RSS* se stalo důležitou technologií pro vzájemnou propagaci webových stránek a hraje důležitou roli při tvorbě zpravodajských portálů. V současné době je specifikace *RSS 2.0* nabízená pod licenci *Creative Commons*<sup>11</sup> a spravuje ji *Berkman Center for Internet & Society*<sup>12</sup> na Harvardské právnické škole.

<sup>9</sup> <http://www.ikaros.cz>

<sup>10</sup> <http://www.cvtisr.sk/itlib/itlib.htm>

<sup>11</sup> <http://creativecommons.org/>

<sup>12</sup> <http://cyber.law.harvard.edu/>

Zběžné nahlédnutí do zdrojového kódu *RSS 2.0* (obr.5) prozradí, že jde o poměrně jednoduchý formát zejména díky samopopisným názvům elementů a atributů.

Obr. č. 3. Ukázka *RSS 2.0*

```
<?xml version="1.0"?>
<rss version="2.0">
  <channel>
    <title>Zelená moucha</title>
    <link>http://www.exmaple.com/moucha/</link>
    <description>Ze života Zelené mouchy.</description>
    <language>cs</language>
    <pubDate>Tue, 10 Jun 2003 04:00:00 GMT</pubDate>
    <lastBuildDate>Tue, 10 Jun 2003 09:41:01 GMT</lastBuildDate>
    <docs>http://blogs.law.harvard.edu/tech/rss</docs>
    <generator>MujSystem 1.0</generator>
    <managingEditor>editor@example.com</managingEditor>
    <webMaster>webmaster@example.com</webMaster>
    <item>
      <title>Druhý den</title>
      <link>http://example.com/moucha/druhy-den.htm</link>
      <description>Druhý den života Zelené mouchy.</description>
      <pubDate>Tue, 03 Jun 2003 09:39:21 GMT</pubDate>
      <guid>http://www.example.com/moucha/archiv.htm#item3</guid>
    </item>
    <item>
      <description>A večer šla Zelená moucha spát.</description>
      <pubDate>Fri, 30 May 2003 11:06:42 GMT</pubDate>
      <guid>http://www.example.com/moucha/archiv.htm#item2</guid>
    </item>
    <item>
      <title>První den</title>
      <link>http://example.com/moucha/prvni-den.htm</link>
      <description>Zelená moucha je tady. Snad se bude libit.</description>
      <pubDate>Fri, 30 May 2003 9:16:28 GMT</pubDate>
      <guid>http://www.example.com/moucha/archiv.htm#item1</guid>
    </item>
  </channel>
</rss>
```

[BUREŠ, 2004]

*RSS* je dnes nedílnou součástí téměř každého zpravodajského serveru či weblogu. V poslední době také některé servery využívají formát *RSS* i pro čtení diskusních příspěvků.

Nepřehlednost různých specifikací znesnadňuje poskytovatelům informací orientaci ve verzích *RSS*. *RSS 1.0* například, jak by se mohlo zdát, není novější verzí *0.91*, *0.92*, *0.93* nebo *0.94*, stejně jako není *RSS 2.0* pokračovatelem *RSS 1.0*. Navzdory tomu se ale ve všech případech jedná o potomky originální verze *0.90 Netscapu*.

Proto byla v roce 2003, ve snaze dostat se ze současné situace, zformována **pracovní skupina IETF**, která měla za úkol vytvořit nový weblogový a syndikační formát, který je označován jako *ATOM* (dříve *Echo*).

Cílem autorů formátu *ATOM* bylo nahrazení stávajících *RSS* formátů. Prostředkem k dosažení tohoto cíle se stal návrh vycházející ze zkušeností s *RSS*. V létě 2005 *ATOM* dospěl k „dospělé“ verzi 1.0. Užívá se tedy jak *RSS* tak i *ATOM*, i když specifikace *RSS* je mnohem rozšířenější.

### 3.3.2 Význam *RSS* v agregaci zpráv

Přítomnost *RSS* kanálů na zpravodajských serverech přináší možnost získat informace bez nutnosti jeho návštěvy. Oproti např. dokumentům typu *newsletter*, dalším způsobům propagace webové stránky, *RSS* kanály nevyžadují registraci návštěvníka a odpadají tak problémy s neochotou sdělovat svá osobní data. Každá webová stránka si může nadefinovat více než jeden *RSS* kanál. Zpravodajské zdroje mívají pro každou rubriku vlastní kanál, a tím umožňují uživateli výběr dle preferencí.

Uživatel má hned dvě možnosti jak výhody *RSS* využít. Může tak učinit buď prostřednictvím pevného klienta v počítači nejčastěji nazývaným ***RSS* čtečka** či *RSS* agregátor (*RSS feed reader*, *RSS aggregator*), nebo přes webové rozhraní, **agregátora zpráv** (*news aggregator*).

***RSS* čtečky** jsou specializované formáty, které umí načíst zdroje z jednotlivých zdrojů na základě nadefinování *RSS* zdrojů (*feeds*). Softwarové programy, které čtou *RSS* formáty užívají vlastní prohlížeč webových stránek nebo využívají výchozí prohlížeče jako *MS Internet Explorer*, *Firefox*, *Opera* aj.

**Webové agregátory** jsou výhodné mimo jiné také proto, že není potřeba instalace žádného softwaru. Poskytují zdroje prostřednictvím svých webových stránek, do kterých pravidelně agregují články z předem nadefinovaných zdrojů. Většina z nich nabízí personalizaci, pomocí níž lze vybrat ty zdroje nebo obory zpráv, které jsou pro uživatele atraktivní. Systémy tohoto typu navíc často poskytují služby s přidanou hodnotou jako např. mediální analýzy.

(viz např. kap. 4.4.3 – systém *Newton IT*)

### 3.4 Typologie systémů agregace zpráv

Před vznikem webové agregace zpráv existovaly dva typy hlavních informačních služeb, které poskytovaly zpravodajství a firemní informace:

- **databáze** poskytující zpravodajské informace v reálném čase i retrospektivně (*Factiva, LexisNexis...*)
- **zpravodajské agentury** (*Reuters, Bloomberg...*)

Vedle těchto služeb se dnes objevily i informace nabízené **webovými agregátory zpráv**, které narozdíl od tradičních služeb poskytovaný obsah nevytváří, pouze zprostředkovávají.

Další z výrazných odlišností je, že tradiční modely platí za část svých výnosů vydavatelům, od kterých dané informace přebírají. Většinou se jedná o 20–25 %, ale může to být až 50 % z příjmu [SEFTON-WILLIAMS, 2005, s.5].

Naproti tomu webové agregační služby většinou žádné takové ujednání s vydavatelem nemají a informace z jejich serverů přebírají bez jakékoliv předběžné dohody.

Webové agregační služby můžeme dále rozdělit podle technologie agregace:

- **automaticky generovaný obsah** – obsah generován na základě filtrů (většinou RSS) a nadefinovaných vzorců, např. *Topix.net, NewsTin.com, Google News* (viz analýza v kap. 4.1 a 4.2).
- **obsah vytvářený uživateli** (*User Generated Content*) – obsah vytvářený více uživateli než vlastníky médií nebo vydavatelem, např. *Digg.com, Reddit.com*

#### 3.4.1 Automaticky generovaný obsah

Je možné jej dále členit:

- **jednoduchá agregace obsahu** (*Simple Content Aggregation*) – reprezentuje nejjednodušší způsob agregování zpráv
- **agregace podle zdrojů** (*Aggregation By Source*) – první logický krok k základní organizaci obsahu

- **obsahové shlukování** (*Content Clustering*) – organizuje informace do významových skupin bez dalšího pokročilého definování
- **obsahová kategorizace** (*Content Categorization*) – více propracovanější metodika, vyžadující předdefinovaný seznam klíčových slov v hierarchické struktuře (taxonomie), který dodá větší hodnotu organizaci informací

### 3.4.2 Obsah vytvářený uživateli

„Osobností roku jste Vy“  
časopis Time, prosinec 2006<sup>13</sup>

Neboli *User Generated Content* (dále jen *UGC*). Místo aby obsah vytvářela profesionální redakce, stránky založené na metodě *UGC* nabízejí svým uživatelům nástroje pro vytváření, sdílení či hostování vlastního obsahu.

*UGC* je klíčovou charakteristikou tzv. *web 2.0*, který identifikuje uživatele jako hlavního činitele v dalším vývoji internetu. Uživatelům neumožňuje spolupracovat na tvorbě webu jen proto, aby se ušetřilo na tvorbě jeho obsahu, ale zejména aby uživatel mohl službu používat po svém. U klasických zpravodajských portálů nemá běžný čtenář prakticky sebemenší možnost ovlivnit to, čemu se zpravodajství věnuje.

Existuje několik druhů systémů, které uplatňují koncepci *UGC*:

- **komentáře** – populární zejména na zpravodajských serverech
- **blogy** – uživatelé jako tvůrci obsahu, píšou zprávy a úvahy, které byly dříve jen v centrálních médiích

<sup>13</sup> Prestižní časopis *Time* jmenoval osobností roku všechny, kteří se podílejí na vytváření nového informačního věku. K volbě vedl redaktory časopisu obsah internetu, který dnes v obrovské míře ovlivňují sami uživatelé (jako příklad uvedli blogy a portály *YouTube* a *MySpace*). „Za to, že jste se chopili otěží globálních médií, založili a budovali novou digitální demokracii, za to, že jste pracovali zadarmo a předčili profesionály v jejich vlastní hře, se osobností roku stáváte vy“ oznámil redaktor časopisu Lev Grossman. Dále uvedl, že tvůrci i zákazníci internetových stránek vytvářených uživateli ukázali dosud nevídanou shodu a spolupráci. Vysvětluje, že dav převzal moc z rukou hrstky vyvolených a pomáhal si bez nároku na odměnu. Celosvětově rozšířený internet je nástrojem, který toto umožňuje. Nástroj, který shromažďuje příspěvky milionů lidí a dává jim význam. [ČTK, 2006]

- **audio RSS (podcast)** – zvukové nebo video záznamy, které autor umístí uje na internet, často ve formátu MP3
- **sociální sítě** – aplikace, které slouží pro výměnu informací mezi jejich uživateli (*MySpace, Facebook*)
- **publikování fotografií** (*Flickr*)
- **publikování video souborů** (*YouTube*<sup>14</sup>)
- **wiki** – označení portálů, které svým uživatelům umožňují přidávat obsah podobně jako v internetových diskusích, ale navíc také umožňují měnit stávající obsah (*Wikipedia*)
- **značkování (tagging)** – označování obsahu uživateli, neboli folksonomie (jako protiklad od taxonomie – což je klasické třídění odborníky)
- **doporučení nebo značkování (tagging)** užitečných/relevantních webových stran pro další internetové uživatele (*Digg, Reddit, de.lic.io.us*<sup>15</sup>.)<sup>16</sup>

Poslední zmíněný druh systémů aplikující myšlenku *UGC* se týká i zpravodajských zdrojů. Například systém *Digg.com*<sup>17</sup>, který se zdá být počtem uživatelů nejpobulárnějším, je jakýmsi zpravodajským rozcestníkem, do kterého uživatelé mohou přidávat nové odkazy, které sami třídí do jednotlivých kategorií a svým hlasováním určují pořadí zpráv na titulní stránce.

<sup>14</sup> Systém *YouTube* nedávno odkoupila společnost *Google* a věděla proč. Uživatelé *YouTube* na něm publikují 100 milionů video souborů, z toho denně je 65 000 nově uložených a 2,5 miliard videí bylo shlédnuto za jeden měsíc (červen, 2006).

<sup>15</sup> Ojedinělý systém *Del.icio.us* je prezentován jako úschovna odkazů. Po zaregistrování lze importovat své oblíbené odkazy a následně je rozřadit do náležitých tematických skupin (přiřadit tag a stručný popis). Odkazy, které si uloží nejvíc uživatelů, jsou pak na prvním místě popularity a v seznamu tagů, který je organizován jako tzv. *tag cloud* (mrak), jsou psány největším písmem.

<sup>16</sup> Pro výše uvedené aplikace existují programové nadstavby pro zpracování dat. Jednou z nich je *AJAX* (*Asynchronous JavaScript and XML*) – obecný pojem pro technologie vývoje interaktivních webových aplikací, které mění obsah svých stránek bez nutnosti jejich znovunačítání, a webová rozhraní *API* (*Application Programming Interface*) pro přístup nadstavbových aplikací třetích stran.

<sup>17</sup> V českých podmínkách existuje méně úspěšný *Linkuj.cz*.

UGC v sobě **však** nese i jisté problémy:

- právní **problémy**<sup>18</sup>
- autorskoprávní problémy
- spam a **lživá** reklama
- množství **nezajímavého** obsahu s nízkou kvalitou

Proto je u takových systémů potřeba jasných pravidel a autority (administrátor je nadřazený člen komunity a lze se na něj odvolat).

### 3.5 Typologie způsobů užití agregace zpráv

Uživatel má **několik** možností **jak** užívat nepřeborné množství služeb systémů, které poskytují agregaci zpráv.

#### 3.5.1 Sledování přehledu článků

Ve vybraném agregátoru **sleduje** na jednom webovém rozhraní aktuální zpravodajství různých médií. Má tak **před** sebou přehledy čerstvě publikovaných článků, v českém prostředí např. *Lidových novin*, *Mladé fronty*, *Aktuálně.cz*, *Práva* a jiné a vybírá si podle zájmů.

#### 3.5.2 Volba kategorie zájmů

Dle preference zájmů si **uživatel** zvolí jednu z několika hlavních kategorií nadefinovaných v systému, jako **například** hlavní zpravodajství, politika, byznys, sport, technologie, společnost či jiné. **Řada** systémů má tato témata nadefinovaná podobně, ale některé, jako např. český systém *NewsTin.com*, umožňují další výběr podkategorií. V případě sportu by **například** následovalo další rozdělení na míčové hry, cyklistiku, gymnastiku..., které **může** být následně rozdělené tezaurovým stylem

<sup>18</sup> Databáze krátkých snímků na serveru *YouTube* musela smazat 30 000 videí poté, co si organizace *Japanese Society for Rights of Authors (JASRAC)* stěžovala na porušování autorských práv. V říjnu 2006 rovněž společnost *Universal Studios* zažalovala dva méně známé servery pro sdílení videa *Group.com* a *Bolt.com* kvůli pirátským nahrávkám, které se na nich objevují.

na podřazené a podřazenější termíny dál a dál. Čím hlubší kategorizaci systém nabízí, tím relevantnější články uživatel na základě svého výběru témat může získat.

### **3.5.3 Definice vlastních kategorií (témat) neboli personalizace**

Pokud systém hlubší kategorizaci nenabízí, jak je tomu například v případě *Google News*, může si čtenář témata sám nadefinovat. Většinou to funguje stejnou formou jako při vyhledávání, kdy se zadá rešeršní dotaz, na základě kterého systém dodává agregované zprávy. *Google News* nabízí vytvoření vlastních témat, pod které následně spadají vyhledané články dle nadefinovaného dotazu.

### **3.5.4 Emailové zasílání zpráv neboli alertní služby**

Na stejném principu funguje zasílání článků elektronickou poštou, kdy na základě uživatelem nadefinovaných témat (stejným způsobem jako v předchozím případě), čtenář periodicky nachází ve svém emailu články obsahující žádaná klíčová slova.

### **3.5.5 RSS čtečky**

Webové rozhraní agregátorů zpráv nemusí být zapotřebí v případě použití RSS čteček. Stačí si stáhnout jednu z mnoha volně dostupných RSS čteček a nadefinovat si RSS kanály dle vlastního zájmu. Mnoho elektronických periodik dnes RSS kanály na svých webech poskytuje. Většinou jsou také vytvářené pro každé téma daného periodika zvlášť, a tak je možné získávat denní zpravodajství podle volby vlastní kategorie – jak bylo zmíněno v druhém případě. Takové zprávy, na rozdíl od nekomerčních agregátorů zpráv, můžeme ve RSS čtečce číst bez obtěžující reklamy kolem.



### 3.6 Metodologie tvorby a architektura webových agregátorů zpráv

Každý jednotlivý systém webové agregace zpráv využívá osobitého přístupu ke své metodologii tvorby, proto jejich postupy nelze přesně sjednotit. V této kapitole je uvedena obecná charakteristika způsobu agregace zpráv u dvou různých skupin systémů zabývajících se webovou agregací zpráv, proto se tyto procesy nemusí přesně shodovat s níže uvedenými příklady systémů.

#### 3.6.1 Základní metodologie webové agregace zpráv u jednodušších systémů

Systémy tohoto typu vynikají nízkými náklady a jsou méně náročné na lidskou intervenci, proto bývají většinou zdarma přístupné např.: *Krátce.cz*, *Články.cz*, *Právě dnes*.

Webové agregátory několikrát denně prověřují vybrané zdroje (tyto zdroje většinou nebývají od sebe odlišeny svojí kvalitou) a v případě, že narazí na nový článek, provedou několik důležitých kroků:

- **Zajištění odkazu na článek, případně vytvoření abstraktu** – Jedná se o krátký úvod velikosti jednoho odstavce. Pokud nelze zjistit je uvedeno prvních 20 – 40 slov textu.
- **Zajištění obsahu článku** – Většinou v rámci *HTML* nebo prostřednictvím kanálů *RSS*. Obsah článku není prezentován uživateli, je stahován pouze za účelem jeho indexace a pro účely vyhledávání.
- **Ověřování klíčových slov či jen jednotlivých rubrik zdrojů** – Podobné systémy většinou neposkytují detailní taxonomii, spíše nabízí několik málo hlavních kategorií – rubrik. Články jsou do jednotlivých kategorií jednoduše řazeny podle toho, ze kterých rubrik sledovaných zdrojů pocházejí články dané skupiny – např. všechny články excerptované z rubriky *sport* se agregují do stejné tematické kategorie.

### 3.6.2 Základní metodologie webové agregace zpráv u propracovanějších systémů

Lze sem zahrnout systémy typu: *Newstin.com*, *Moreover*, *Inform.com* aj. Takové systémy narozdíl od výše uvedených užívají **explicitní taxonomii** založenou na **inteligentní kategorizaci zdrojů** a poskytují **další služby s přidanou hodnotou** jako např. **tvorbu souvisejících článků**. Jedná se o kombinace lexikálních, sémantických a kontextuálních analýz a lidské intervence s automatizovaným systémem.

Takové systémy většinou uplatňují níže uvedené postupy:

- **Nalezení zdroje a jeho klasifikace** – Jedna z mála činností, která není plně automatizována. Zdroje mohou být sice vyhledány automaticky pomocí harvesteru<sup>19</sup>, jejich kontrola je však nezbytná stejně tak jako následná definice hodnoty jednotlivých zdrojů. Profesionálním webovým stránkám, například zpravodajským agenturám je přiřkládána vyšší hodnota než blogům – tato hodnota, mimo jiné, ovlivňuje výsledné řazení dokumentů dle relevance (ne všechny systémy toto uplatňují - některé jen vymezují blogy od klasických zpravodajských zdrojů speciální rubrikou).
- **Automatická identifikace jazyka** – Je prováděná pomocí lingvistických aplikací např. *Inxight Software*, český systém *Lingea* aj.. Je určena nejen pro identifikaci jazyka, ale také pro lematizaci a následné skloňování slov, které napomáhá kvalitní kategorizaci zpráv.
- **„Vyřezávání“ článků z HTML stránek dokumentu** – Všechny články jsou většinou ukotveny mezi reklamou a jiným obklopujícím textem nevztahující se k článkům. Pro přebírání článků jsou také využívány komunikační formáty *RSS*, *ATOM* nebo *RDF*.
- **Vytváření abstraktu** – viz výše
- **Ověřování podobnosti článků** – Jednotlivé články jsou porovnávány a pokud u nich dojde k podobnosti klíčových slov, shlukují se v rámci jedné kategorie (tak vznikají související články).

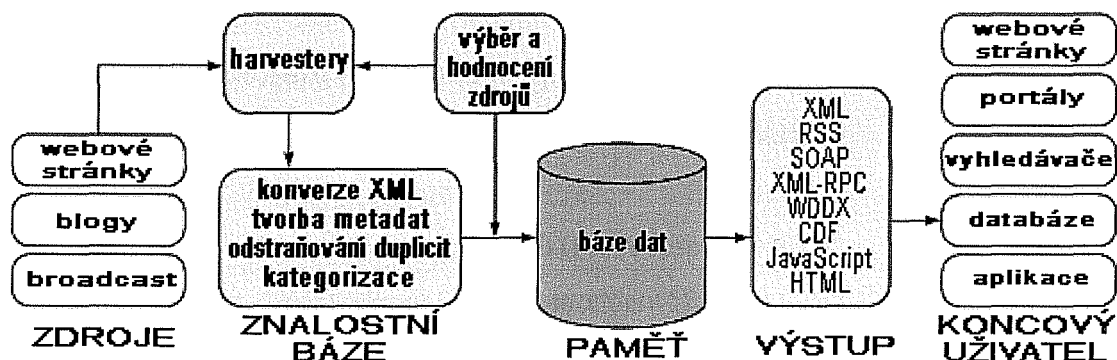
---

<sup>19</sup> Harvester je speciální softwarový nástroj využívaný ke sběru webových zdrojů.

- **Reduplikace** – Odstraňování duplicit – stejných článků v rámci jedné kategorie.
- **Tvorba metadat** – K článkům jsou na základě metadat vytvořeny entity jako např. názvy firem, produktů, geografických míst, osob aj.
- **Kategorizace** – Každá kategorie může být nadefinována speciálními vzorci pro identifikaci témat, kontextu a entit, které jsou manuálně vytvářené pro jednotlivé kategorie a tím napomáhají k co nejrelevantnějším výsledkům. Systém při zpracovávání článku srovnává jeho obsah s nadefinovanými kategoriemi, které následně třídí do příslušných tematických skupin – tak vzniká výsledné uživatelské rozhraní s organizovanými zpravodajskými informacemi.

Na obrázku 5 lze vidět konkrétní architekturu agregace zpráv u systému *Moreover*.

Obr.5: Ukázka XML architektury komerčního agregátora zpráv inspirovaného systémem *Moreover*



### 3.7 Současné světové trendy agregace zpráv

V lednu 2006 vyšel článek týkající se projektu Evropské unie zvaným *REVEAL THIS* [CHOWDHURY, 2006], který měl v rámci průzkumu zjistit jaké očekávání mají uživatelé od webových agregátorů zpráv a jakým způsobem hodnotí jejich jednotlivé služby.

Do systému bylo zahrnuto 5 webových agregátorů: *Google News*, *HeadlineSpot*<sup>20</sup>, *Newsburst*<sup>21</sup>, *Awasu*<sup>22</sup> a *TVEyes*<sup>23</sup> a zúčastnilo se ho 45 respondentů.

Vítězně z těchto systémů vyšel agregátor *TVEyes*, který získal nejlepší ohodnocení od uživatelů ve všech směrech. *TVEyes* není typickým agregátorem. Je to komerční systém poskytující zpravodajství z televizních a rozhlasových kanálů. Poskytuje přístup k aktuálním i starším zdrojům. Výsledek ukazuje, že uživatelé mají rádi multimediální zprávy, což si také uvědomila většina světových agregátorů, které multimediální zdroje čím dál více zahrnují do svých systémů.

Z analýz uživatelských očekávání vyplynulo, že dobrý webový agregátor zpráv by měl mít následující vlastnosti:

- pokročilé vyhledávací funkce (uvedlo 80 % respondentů)
- uživatelsky přívětivé rozhraní (78 % respondentů)
- vysoce kvalitní a renomované zdroje (76 % respondentů)
- přístupný archiv (71 % respondentů)
- možnost prohlížení (*browsing*) článků (69 % respondentů)
- chronologické třídění zdrojů (65 % respondentů)
- možnost personalizace (odpovědělo 62 % respondentů)

Mezi další časté odpovědi, které uvedlo kolem 50 % uživatelů patřila *sumarizace textu*, *geografická personalizace*, *multimediální zdroje*, *služby typu SDI* (průběžné informování – *alerts*).

---

<sup>20</sup> <http://www.headlinespot.com>

<sup>21</sup> <http://www.newsburst.com>

<sup>22</sup> <http://www.awasu.com>

<sup>23</sup> <http://www.tveyes.com>

Mnoho informačních profesionálů objevilo hodnotu využívání webové agregace zpráv. Tyto nástroje velice šetří čas každému kdo potřebuje sledovat více zpravodajských zdrojů za den. Jednotlivcům přináší přehled o jednotlivých událostech a firmám užitečný nástroj pro sledování konkurenčního zpravodajství.


Systemy agregací zpráv se staly integrální součástí mediální scény, přičemž jejich vliv poroste časem s rozšiřováním počtu uživatelů na internetu. Pro zpravodajské servery produkující vlastní články to znamená přijmout tuto skutečnost a uvědomit si co pro ně takové systémy přináší.

## 4 VÝBĚR JEDNOTLIVÝCH SYSTÉMŮ A JEJICH CHARAKTERISTIKA

Volně přístupné systémy využívají plně automatizovaný systém pro agregaci zpráv. Systémy typu *Google News* tak na jednu stranu vynikají svými nízkými náklady, na straně druhé jsou ale limitovány v poskytování relevantních informací. Na rozdíl od komerčních systémů, které poskytují lidskou intervenci ve svých procesech. Tato intervence vychází z podstaty zpravodajství jako takového, které již od svých počátků vyžadoval editorskou práci a lidský úsudek k vyhodnocení toho co je důležité.

### 4.1 Systémy zdarma přístupné ve světě

#### 4.1.1 *Google News*<sup>24</sup>

 *Google News* započal svoji kariéru v prosinci 2001, kdy byl vytvořen za víkend mladým inženýrem, který byl frustrován vyhledáváním zpráv po události 11. září 2001.

Agregace zpráv na *Google*, proslulém jako jeden z nejlepších internetových vyhledávačů na světě, byla oficiálně spuštěna v **září 2002**. Teprve ale od ledna 2006 se vymanila z beta verze. Jako ve všech svých poskytovaných službách se *Google* i v tomto směru snaží být výjimečný a netají se tím, že usiluje o celosvětové pokrytí. V současné době poskytuje již **36 regionálních edic v 12 jazycích**, na jejichž rozšíření se stále pracuje.

#### **Poskytované služby:**

##### **Personalizace zpráv**

Na základě vytvoření účtu, možnost vytváření vlastních sekcí a jejich kombinování s již existujícími z 36 regionálních edic. Lze si také nastavit tzv. **doporučování článků** (*recommended news*), které se zobrazuje ve speciální sekci pod hlavním zpravodajstvím (*top news*) na základě předchozího označování článků našeho zájmu. Pomocí algoritmů *Google* analyzuje tyto selekce a nabízí další relevantní články. Algoritmy porovnávají naši selekci se selekcí jiných uživatelů, dle

---

<sup>24</sup> <http://news.google.com/>

kteří doporučují články, jež četli jiní uživatelé, kteří se také věnovali podobným článkům, nacházejících se v zaznamenané historii. Údaje v historii lze též podle libosti promazávat. *Google News* uvádí, že čím déle uživatel bude používat tuto službu, tím lepší doporučení mu systém časem poskytne.

### **Průběžné informování (*alerts*)**

Lze zvolit okamžité – denní nebo týdenní emailové zasílání informací o nových člancích na vybrané téma. Témata se definují stejným způsobem jako rešeršní dotazy ve vyhledávači *Google*. Lze uplatnit veškeré pokročilé funkce webového vyhledávání. Není zde sice pokročilé rozhraní, ale dotaz lze nadefinovat přes pokročilé vyhledávací rozhraní a výsledek nakopírovat jako požadované téma pro emailové upozornění.

### **Mobilní zpravodajství**

Umožňuje prohlížet aktuální zpravodajství přes mobilní telefony poskytující webové služby.

### **Kanály (*feeds*)**

Podporuje *RSS* a *ATOM* kanály, které umožňují nastavení vybraných sekcí do čteček kanálů. Pro anglickou edici lze vytvořit i kanál uživatelem uzpůsobené stránky.

### **Archiv**

- Vyhledávání v archivu lze retrospektivně až 200 let zpátky.
- Při pokusu zadání dotazu:

*"news aggregator" OR "news aggregators" OR "news aggregation"*

se lze dozvědět, že nejstarší archivované články zmiňující agregaci zpráv, se nabízí již z června 2000 a dle časové osy je možné vidět, že je jich hned několik, zejména zmiňujících se o agregátoru zpráv *Moreover* (rovněž analyzovaného v kapitole 4.2.3).

Většina článků, které jsou zpřístupněny retrospektivně, je dostupná za poplatek jejich poskytovatelům.

#### 4.1.1.1 Newsmap<sup>25</sup>

Tento nástroj, spuštěný v březnu 2004, umožňuje na principu vizuálního zobrazování (*treemap visualization algorithm*) pozorovat neustálou proměnu značného rozsahu zpráv, které agreguje *Google News*.

*Newsmap* dovoluje vizualizovat aktuální zpravodajství podle geografického hlediska, celkem **11 zemí** (Austrálie, Španělsko, Francie aj.), a dle **7 tematických kategorií** (svět, business, zdraví, sport aj.) či času. Na jedné stránce obdržíme globální pohled na poslední publikované zprávy, seskupené dle předdefinovaných kategorií (svět, technologie, sport atd.)

Na webovém rozhraní se graficky zdůrazňují ty tituly, které jsou nejvíce populární, tzn., že mají nejvíce souvisejících článků. Ostatní články jsou pod nimi nabízeny menším písmem podle četnosti výskytu jejich tématu ve světových zdrojích.

Velikostně zde také vyčnívají kategorie, pod nimiž je publikováno nejvíce článků. Kromě toho se prostřednictvím různých barevných tónů zdůrazňují nejaktuálnější tituly.

Uživatelské rozhraní rovněž dovoluje zkoumat vztahy jednotlivých seskupení článků, tedy všechny tituly jedné kategorie jsou v jedné skupině.

Obr. 6: Systém *Newsmap*



<sup>25</sup> www.marumushi.com/apps/newsmap/newsmap.cfm



#### 4.1.2 Topix.net



Webový agregátor zpravodajství *Topix.net* sídlící v Palo Alto v Kalifornii byl založen v roce 2002<sup>26</sup>. Jeho funkčnost je postavena na speciálních algoritmech na základě kterých je soustavně monitorováno přes 10 000 zdrojů (noviny, magazíny, blogy, vládní zdroje, televizní a rozhlasové stanice). Jednotlivé excerpované dokumenty jsou následně roztrženy do 360 000 kategorií.

Mezi nimi lze mimo jiné nalézt kategorie:

- 30 000 amerických měst
- 5 500 firem
- 48 000 celebrit a hudebníků
- 1 500 sportovních týmů a osobností sportu

Je jediným agregátorem, který poskytuje kategorizaci článků též podle poštovních směrovacích čísel, které upřesňují všechna města ve Spojených státech amerických a v Kanadě.

Uživatel má možnost články prohlížet buď v rámci hlavní stránky, kde nalezne nejaktuálnější témata nebo skrze 360 000 kategorií. Je možné také vyhledávat v historii zpráv prostřednictvím jednoduchého a pokročilého rozhraní (využívající technologii *AJAX*).

V **jednoduchém rozhraní** je možné zadat název tématu, města nebo poštovního směrovacího čísla.

**Pokročilé rozhraní** nabízí:

- možnost volby zahrnutí nebo vyloučení blogů

---

<sup>26</sup> Tvůrci systému *Topix.net* před jeho vznikem vytvořili *Open Directory Project* (originálně nazývaný *NewHoo*), který byl prvním webovým katalogem a nyní patří mezi největší katalogové servery na internetu jež spoluvytváří lidé z celého světa. Nyní je majetkem společnosti *America Online* a je zdrojem pro jiné webové katalogy včetně *Google Directory*. Absence zpravodajství ve webových katalozích byla podnětem k vytvoření systému *Topix.net*.

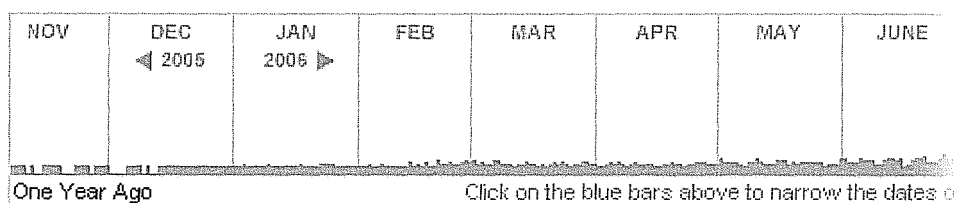
<sup>28</sup> <http://www.comscore.com>

- vymezení zdroje (např. vyhledá pouze články *New York Times*)
- vymezení URL adresy (např. *cnn.com*)
- vymezení lokace (poštovního směrovacího čísla nebo názvu města)
- vymezení země
- vymezení data publikování

Jako u většiny webových agregátorů lze články setřídit podle relevance (dána podle počtu souvisejících článků, využívá patentované technologie *NewsRank™*) nebo data.

Dalším interaktivním prvkem a také novinkou je časová osa, na které je možné vidět kolik pozornosti si jednotlivé téma ve zpravodajství získalo za období více než jednoho roku zpětně. Archiv je možné rovněž procházet.

Obr. 7. – Časová osa systému *Topix.net*



Pokud má uživatel zájem vědět např. co se psalo o *Google News* v listopadu, tak jednoduše zadá *Google News* a poté v časové ose klikne na listopad 2006.

V současné době je *Topix.net* ohodnocen na 72 milionů dolarů a během září 2006 tento systém navštívilo 3,9 milionů unikátních návštěvníků (*Comscore*<sup>28</sup>). Zatímco minulý rok jich bylo ve stejném období 3,6 milionů.

V listopadu 2006 *Topix.net* získal 15 milionů dolarů od tří hlavních zpravodajských společností: *Gannett Co.*, *The McClatchy Co.* a *Tribune Co.*. Peníze budou využity zejména pro rozšíření lokálního zpravodajství. *Topix.net* poskytuje zpravodajství pro více než 30 000 lokálních komunit.

**Hlavní služby:**

- vkládání vlastních článků – *Topix.net* umožňuje uživatelům přispívat vlastními články
- možnost vkládání komentářů k jednotlivým článkům
- emailová upozornění (*Alerts*)
- personalizace
- *Top-Blogs* - monitoring aktivit 100 000 blogů a hodnocení jejich příspěvků
- využití *RSS* kanálů v rámci všech kategorií

### 4.1.3 Inform.com<sup>29</sup>



Založen v roce 2004. Orientuje se na aktuální zpravodajství, které je zpřístupněno po dobu maximálně 4 dnů<sup>30</sup>. Zprávy monitoruje z více než 10 000 zdrojů včetně tradičního zpravodajství, video, audio souborů a blogů.

#### Hlavní charakteristika:

Všechny články zpřístupněné v systému *Inform* jsou opatřeny pěti odkazy (viz obr.7):

Obr. 7: Pět odkazů článku *Inform.com*



#### Blair pushes Mideast peace process

Reuters My News

British Prime Minister Tony Blair said on Saturday that Palestinians need to find a way to form a fully functioning government with which the rest of the world can deal. Blair began a Middle East trip by holding talks with Turkish Prime Minister Tayyip Erdogan.

related: stories 1 subjects 2 email 3 flag 4 5

#### 1.související články

- zprávy týkající se stejného tématu, určují relevanci zpráv

#### 2.související předměty

- zobrazí hlavní předměty (vznikají na základě automatické tvorby **metadat** k jednotlivým článkům) vztahující se k obsahu článku jako *témata, lokace, organizace, osoby, produkty*
- jednotlivá předmětová hesla odkazují na patřičné články z periodik či blogů, video a audio souborů obsahující daný předmět (viz obr. 8).
- vybraná předmětová hesla lze jednoduše vložit do personalizovaného rozhraní *Moje zprávy (My News)*

<sup>29</sup> <http://www.inform.com>

<sup>30</sup> Plánuje retrospektivní rozšíření.

### Obr.8: Související předměty

subjects @	email @	flag @	?   x
<b>Topics</b>	Government, Politics, Russian Politics, World Politics, Science and Technology, Technology, Venezuelan Politics		
<b>Places</b>	Russia, Venezuela, United States, Moscow Kremlin, Iran, South America, Moscow, Belarus		
<b>Organizations</b>	Kremlin, OPEC		
<b>People</b>	Hugo Chavez, Vladimir Putin, Sergei Chemezov, Bolivar, Fidel Castro		
<b>Products</b>	F-16 Fighting Falcon Aircraft		

**3.email** – možnost sdílení článků prostřednictvím elektronické pošty

**4.označení článku (Flag)** – zprávu lze označit pro pozdější použití, seznam označených článků se zobrazí v rámci rozhraní *Moje zprávy* (uživatelé si většinou vyberou předem několik článků, které si posléze najednou prohlídí ve svém personalizovaném rozhraní)

**5.Digg** – propojení na *Digg.com* kam lze vybraný článek vložit (články, které byly vloženy v systému *Digg* nejvíce uživateli se dostávají na hlavní webovou stránku)

V rámci personalizovaného rozhraní *Moje zprávy*

(obr. 9) má uživatel několik možností:

- prohlížení dříve označených článků (přes odkaz *Flag*)
- vymezení zdrojů
- vymezení předmětových hesel
- definování libovolného RSS kanálu

My News	
<b>Top Stories</b>	
<b>Flagged Articles</b>	
<b>Sources</b> <input type="button" value="READ"/>	
Gawker	x
New York Times	Edit   x
TechCrunch	x
<b>Subjects</b> <input type="button" value="READ"/>	
Competitive Intelligence	Edit   x
<b>Feeds</b> <input type="button" value="READ"/>	
Weather	Edit   x

Obr. 9: *Moje zprávy*

Inform také nabízí služby vydavatelům větších zpravodajských deníků. Svoji službou jim napomáhá zvyšovat návštěvnost jednotlivých stránek pomocí odkazů na další podobné články a na související předmětová hesla. Odkazy tvořené na základě systematické tvorby metadat přispívají k celkové dynamičnosti webových stránek.

The screenshot shows the Inform website interface with two main sections: 'RELATED CONTENT' and 'RELATED SEARCHES'. The 'RELATED CONTENT' section is divided into 'OKLAHOMAN ARTICLES' and 'OKLAHOMAN ARCHIVES'. The 'RELATED SEARCHES' section includes categories like TOPIC, PERSON, STATE, ORGANIZATION, COMPANY, and BODIES OF WATER, each with a list of search terms and radio buttons for selection.

RELATED CONTENT	RELATED SEARCHES
<p><b>OKLAHOMAN ARTICLES</b></p> <ul style="list-style-type: none"> <li>• 3 dead, 1 hurt in plane crash 12/16/2006</li> <li>• Plane crash kills 2 12/03/2006</li> </ul> <p><b>OKLAHOMAN ARCHIVES</b></p> <ul style="list-style-type: none"> <li>• 2,000 at City FAA Center Return to Jobs 11/17/1995</li> <li>• Tulsa dies in Owasso plane crash 07/03/2006</li> <li>• Plane crash kills pilot, injures two in Cushing 06/22/2003</li> <li>• Five hurt as passenger jet descends to avoid military planes near Tulsa 03/18/2003</li> </ul>	<p><b>TOPIC</b></p> <ul style="list-style-type: none"> <li><input type="radio"/> Air Disasters</li> <li><input type="radio"/> Transportation</li> </ul> <p><b>PERSON</b></p> <ul style="list-style-type: none"> <li><input type="radio"/> Roland Herwig</li> <li><input type="radio"/> Kera Phillippi</li> </ul> <p><b>STATE</b></p> <ul style="list-style-type: none"> <li><input type="radio"/> Oklahoma</li> </ul> <p><b>ORGANIZATION</b></p> <ul style="list-style-type: none"> <li><input type="radio"/> U.S. Federal Aviation Administration</li> <li><input type="radio"/> U.S. National Transportation Safety Board</li> <li><input type="radio"/> Oklahoma Highway Patrol</li> </ul> <p><b>COMPANY</b></p> <ul style="list-style-type: none"> <li><input type="radio"/> The Walt Disney Company</li> </ul> <p><b>BODIES OF WATER</b></p> <ul style="list-style-type: none"> <li><input type="radio"/> Grand Lake</li> </ul>

Obr.10: převzato z oklahomského zpravodajského zdroje NewsOk.com

Mezi nejznámější zpravodajské zdroje využívající této služby patří např. *Washington Post*, *New York Sun* a portál *Answers.com*.

#### 4.1.4 NewsNow<sup>31</sup>

System News Now, působící od roku 1997 byl prvním agregátorem zpráv ve Velké Británii a brzké době ho navštěvovalo 100 000 pravidelných uživatelů. Dodnes patří mezi hlavní služby monitoringu zpráv ve Velké Británii. V pěti minutových intervalech automaticky vyhledává ve 27 500 zdrojích (národní, regionální a mezinárodní zpravodajské zdroje, vládní a firemní stránky, blogy, dokumenty typu webzines, newsletters a newswires, stránky televizních stanic a další zdroje na vyžádání) ve 22 jazycích ze 139 zemí. Je využíván 1,1 miliony uživatelů.

Zdarma je na internetu zpřístupněná pouze část služeb, zbylé jsou poskytovány za pevný měsíční poplatek.

<sup>31</sup> <http://www.newsnow.co.uk>

News Now je vytvořen na dynamické platformě, kterou lze implementovat do firemních intranetů či webových stránek.

### Hlavní charakteristika:

- 30 denní archiv zpráv
- zprávy ve 22 jazycích ze 139 zemí
- možnost tvorby vlastních klíčových slov či frází a definování vztahů mezi nimi
- podpora zákazníkům
- možnost vymezení a vložení vlastních zdrojů
- emailová upozornění (*Alerts*)

Obr.11: Rozhraní webového agregátora zpráv News Now

The screenshot displays the News Now web interface. At the top, there is a navigation bar with links: HOME | ABOUT | PRESS | VACANCIES | ADVERTISE | PUBLISHER NETWORK | LIVE FEED | RELEASE WIRE | FAQ | BUSINESS SERVICES | CONTACT US | CLIENT LOGIN. Below this is a banner for 'NEWS NOW CURRENTLY' with the text 'Automatically searching 27,175 news sources every 5 minutes' and a 'pre-book at www.gatwickparking.com' advertisement.

The main interface is divided into several sections:

- Headline Search:** Includes a search box with 'Enter one word' and radio buttons for 'this newsfeed' (selected) and 'all articles'.
- Choose a newsfeed:** A list of categories with checkboxes and expand/collapse arrows:
  - Hot Topics (34)
  - Business & Finance (38)
  - Industry Sectors (42)
  - I.T. (38)
  - Internet (20)
  - Current Affairs (75)
  - Regions (107)
  - Sport (49)
  - Football (172)
  - US Sports (134)
  - Entertainment (34)
- Energy Politics:** A filter section with a date dropdown set to 'Sun 17 Dec' and an 'Earlier' button.
- Time Filters:** A vertical list of time-based filters:
  - 5 to 10 minutes old
  - 15 to 30 minutes old
  - 1 to 2 hours old
  - 2 to 4 hours old
- News Feed:** A list of news items with headlines and source/date information:
  - China's 2006 Crude Oil Import Estimated At 140 Million Tons *Nasdaq 09:22*
  - China 2006 crude imports seen at 140 mln tons, up 10.2 pct - Xinhua interactive investor 09:06
  - Profits to be split 70-30 between ETTC and Timor govt; Kuwait, East Timor establish... *Arab Times 08:03*
  - 'OPIC' nations agree to boost energy efficiency; Strengthening cooperation on strategic oil stocks *Arab Times 08:03*
  - Chile, Pakistan To Work Together In Mining Sector *Nasdaq 07:23*
  - Funding threatened by Congress's inaction *The Detroit News 06:56*
  - India, Japan join hands to script a new Asian era *Gulf Times 06:32*
  - Thanh Nien, BP continue aid to typhoon victims *Thanh Nien Daily 06:27*
  - Oil consumers join hands for security *Gulf Times 06:17*
  - Westinghouse bags China nuclear deal *Gulf Times 06:17*
  - Part of Greece-Turkey gas pipeline to operate in 2007: Greek minister *People's Daily 06:12*
  - Don't let the door hit you going out *Jacksonville Daily News, North*
- Top Stories (Last 24hrs):** A list of featured news items:
  - In One Industry, Great Danger From Falling Axes *The New York Times 23:06 16-Dec-06*
  - Iraq trade unions: No foreign oil control *The Asian Age 20:09 16-Dec-06*
  - White House Clamps Down On USGS Publishing *Slashdot 19:08 16-Dec-06*
  - The dollar melts as Iraq burns *The Star 14:32 16-Dec-06*
  - Iran May Begin Sharing Nuclear Technology *Voice of America 12:14 16-Dec-06*
  - US didn't keep promise: Scientists *rediff.com 11:59 16-Dec-06*
  - Part of Greece-Turkey gas pipeline to operate in 2007:... *People's Daily 06:12 17-Dec-06*
  - High crude prices worry Petroleum Minister *The Peninsula 04:24 17-Dec-06*
  - China's 2006 Crude Oil Imports Expected To Reach 140... *Nasdaq 03:39 17-Dec-06*
  - 'Longstocking' seduces Germany *The Times 01:05 17-Dec-06*
- Previous Top Stories:**
  - The Oil Connection in Iraq Study Group Report *ArabNews 00:06 16-Dec-06*
  - Britain: Iran Sanctions Deal 'Emerging' *Forbes.com 01:06 16-Dec-06*
  - Faint signs of hope

At the bottom left, there is a logo for 'BT Bringing it all together'.

## Popis zdarma přístupného rozhraní News Now

- zpřístupňuje pouze zprávy v anglickém jazyce
- zprávy jsou členěny do časových kategorií od nejaktuálnějších až po 30 dnů staré
- možné filtrování v rámci jednoho konkrétního data
- volba mezi 750 předmětovými kategoriemi
- fulltextové vyhledávání ve všech zprávách nebo v jednotlivé aktuálně zobrazené kategorii
- původ zpráv ze zahraničních serverů upřesňují vlajky daného státu – všechny jsou v anglickém jazyce
- ikona *SUB* u některých článků označuje zprávy vyžadující předplatné u odkazujícího zdroje



## 4.2 Komerční systémy ve světě

### 4.2.1 Factiva



*Factiva* byla od roku 1999 až donedávna produkována společnostmi Dow Jones a Reuters. V říjnu 2006 agentura Reuters prodala firmě Dow Jones svůj 50% podíl<sup>32</sup>, a tím se **Dow Jones stal plným vlastníkem společnosti Factiva.**

*Factiva* je unikátním zdrojem zpřístupňujícím klíčové informace především z oblasti obchodu a ekonomiky.

V srpnu 2006 *Factiva* dosáhla prvního místa na trhu, díky své službě aktuálního zpravodajství (*Current Awareness News*) a online výzkumu trhu (*Research Online Industry*). *Factiva* rovněž vede počtem svých předplatitelů, kterých je 1,86 milionů [Sullivan, 2006].

Systém *Factiva* užívá 84 % z 500 firem hodnocených časopisem *Fortune*, který každoročně hodnotí 500 největších amerických společností [CNNMoney.com, 2006].

V únoru 2005 *Factiva* obohatila své služby o obsah webového agregátoru *Moreover*, který byl integrován do služby *Factiva Insight*. *Moreover* poskytuje přístup k online zpravodajským článkům agregovaných z téměř 10.000 webových stránek zahrnujících prezentace světových zpravodajských zdrojů, regionálních, firemních publikací a firemních a vládních stránek. Obsah rovněž zahrnuje informace v 26 jazycích ze 125 zemí.

#### ***Factiva*, mj. obsahuje následující zdroje:**

- více jak 10 000 zdrojů ze 159 zemí, 350 regionů ve 22 jazycích
- přes 120 aktualizovaných zpravodajských zdrojů (*newswires*), vč. výhradní kombinace zdrojů od *Dow Jones*, *Reuters* a *Associated Press*. Celkem více než 500 zpravodajství, vč. *Asia Pulse*, *VWD*, *Agence France Presse*, *Agencia EFE* atd.

<sup>32</sup> Agentura Reuters získala za svůj podíl Factivy 160 milionů dolarů.

- aktuální denní obsah novin *Wall Street Journal*, *New York Times*, *Washington Post*, *Globe and Mail*, *Financial Times*, *Les Echos*, *South China Morning Post*, *Australian Financial Review*, *Sydney Morning Herald*, *Straits Times*, *Yomiuri Shimbun*, *Frankfurter Allgemeine Zeitung*, *Irish Times* a lokálních novin z různých částí světa
- přepisy klíčových televizních a rozhlasových zpráv od *BBC*, *ABC*, *CBS*, *NBC*, *Fox*, *CNN*, *NPR* apod.
- všeobecná a odborná periodika jako *Economist*, *Forbes*, *Fortune*, *Time*, *Newsweek*, *Finanz & Wirtschaft*, *Satellite News*, *BusinessWeek*, *Focus*, *L'Express*, *Bulletin*, *Far Eastern Economic Review* atd.
- fotografie a vyobrazení od *Reuters* a *Knight-Ridder*, firemní zprávy, historická data o trzích, odkazy na www zdroje
- blogy, elektronické nástěnky apod.
- *Search 2.0*

*Factiva* na začátku roku 2006 spustila nové vyhledávání *Search 2.0*, využívající vizualizace a interaktivní grafické navigace, které je stále ve fázi beta a je upravováno na základě získané zpětné vazby. *Search 2.0* je určené pro uživatele, kteří neznají vyhledávací strategie a jeho rozhraní je proto podobné vyhledávači *Google*. Výsledky vyhledávání jsou doplněny o tzv. *Discovery Pane*, který je setřídí a vizuálně zorganizuje do interaktivních tabulek a grafů. Umožňuje tak objevovat souvislosti ve vyhledaných datech, s nimiž je možné dále pracovat.

*Factiva* uvedla, že nové grafické filtrace a navigace změní firemní výzkum, protože uživatelé budou schopni snadněji nacházet vzájemné vztahy a souvislosti v rámci daného rešeršního výsledku [Hane, 2006].

*Search 2.0* vyhledává v člancích za **posledních 90 dnů**. Webové rozhraní je pouze v angličtině, ačkoliv některé zdroje jsou i v jiných jazycích.

Při zobrazování plných textů článků je možné také přejít na články s podobným obsahem, tzv. **související články**.

5

The screenshot displays the Factiva Search 2.0 interface. At the top, there's a navigation bar with 'Factiva Beta', 'factiva.com', 'Manage Alerts', 'Clipboard', 'Feedback', and 'Support'. The search term 'immigration' is entered in the search bar, with 'More Search Options' and 'Alert me' buttons. Below the search bar, there are filters for 'Publications' (Web News [12065], Pictures [468]), 'All' (Newspapers, Magazines, Newswires), and 'Date'. The main content area shows a list of search results, each with a title, source, and a brief snippet. On the right side, there are several analytics and filters: a bar chart showing 'Distribution: Weekly' from Jan 9 2005 to Apr 23 2005; 'News Clusters' listing terms like 'nursing homes work: fired workers'; 'Companies' listing DP World, AWB LTD, etc.; 'Sources' listing Associated Press, US Fed News, etc.; 'Subjects' listing Asylum/Immigrat..., Domestic Politics, etc.; 'Industries' listing Immigration/Natu..., Broadcasting, etc.; and 'Keywords' listing 'term care united states immigration reform illegal immigrants immigration officials immigration debate nursing homes health care care workers fired workers'.

**Popis jednotlivých funkcí Search 2.0, viz jednotlivé body na obr. 12:**

1. výsledky jsou prezentovány prostřednictvím tří odkazů:
  - publikace (rozdělené na články ze zpravodajských agentur, novin a časopisů)
  - webové zpravodajství (poskytuje webový agregátor *Moreover*)
  - obrázky
  
2. Výsledky jsou řazeny buď dle relevance nebo data.

3. Průběžné informování (*Alert me*) – nastavení emailového zasílání aktuálních informací týkajících se vybraných témat z taxonomie.
4. Další vyhledávací možnosti (*More Search Options*) – podobné webovým vyhledávačům. Dodatečné vyhledávací pole, které nabízí následující možnosti definování rešeršního dotazu:
  - se všemi slovy
  - s alespoň jedním ze slov
  - neobsahující slova
  - s přesnou frází
  - časové omezení (včera, minulý týden, za poslední měsíc)
5. Označení dokumentů (*Clipboard*) – až 25 dokumentů může být přidáno (v rámci jednoho přihlášení) pro zaslání emailem, tisk, či uložení.
6. Speciální filtry (*Discovery Pane*) – dále uvedené; je možné podle potřeby je schovat a vytvořit tak větší prostor pro samotné články.
7. Časová osa – seřazení článků dle vybrané časové osy.
8. Zpravodajské klastry (*News Clusters*) – klastry hlavních témat.
9. Firemní klastry – seřazeno podle nejvíce citovaných firem.
10. Klastry zdrojů (*Sources*) – seřazeno podle zdrojů, ze kterých byly nejvíce agregovány články.
11. Předmětové klastry (*Subjects*) – organizace článků dle předmětových hesel.
12. Klastry průmyslových odvětví (*Industries*).
13. Klíčová slova (*Keywords*) – seznam klíčových slov vyskytujících se v prvních 100 člancích výsledku.

## Taxonomie

Factiva má vlastní speciální patentovanou taxonomii<sup>33</sup> pro průmyslová odvětví, regiony, předmětová hesla a organizace:

- *Company Codes* (320,000+)
- *Industry Codes* (872+)
- *Regional Codes* (350+)
- *Subject Codes* (650+)

Je aplikována do všech obsahů, jazykových verzí a služeb společnosti. Factiva také ve svých službách využívá speciální jazykovou technologii od společnosti *Inxight Software, Inc.*.

## Pokročilé vyhledávání

Pokročilé vyhledávání je určeno pro profesionální rešeršéry.

Zde je několik jeho základních funkcí:

- vyhledávání ve volném textu (*free text searching*) – ruční zadání dotazu pomocí dotazovacího jazyka
- možnost rozšiřování dotazů přidáváním termínů ze speciální taxonomie
- využití Booleovských a dalších pokročilých operátorů
- specifikace typu zdrojů
- RSS kanály – v rámci alertních služeb je možné zahrnout jakýkoliv kanál zprostředkovaný agregátorem *Newsgator*, s kterým Factiva spolupracuje

---

<sup>33</sup> Dow Jones Reuters Business Interactive (US). Polyarchical data indexing and automatically generated hierarchical data indexing paths. Jonathan Guy Grenside Cook aj. (inventors). MPT6: G06F 017/30. United States Patent, 6938046. 2001-03-02.

#### 4.2.2 Thomson Business Intelligence

### **THOMSON BUSINESS INTELLIGENCE**

*Thomson Business Intelligence* (dále jen *TBI*) je

portál **obchodně-ekonomických, analytických, zpravodajských a marketingových informací**. Kromě dalších zdrojů obsahuje známý *News Research* (dříve *NewsRoom* – e-knihovnu cca 11 000 novin a časopisů) a *Market Research* (dříve *Profound* – e-knihovnu cca 220 000 zahraničních marketingových studií v plném znění). U *News Research* může nyní uživatel kromě mnoha jiných možností specifikovat hledání podle zdroje, včetně omezení na určitý typ periodika. Dále byla u *TBI* posílena funkčnost a praktičnost alertních služeb, adresného monitoringu na základě uložených filtrů, možnost odkazování na konkrétní zprávy a dokumenty pevnou *URL* adresou i z prostředí uživatelských intranetů. Také se zvýšily možnosti personalizace, tedy upravení vzhledu a funkčnosti rozhraní podle potřeb uživatele.

*TBI* je název pro společnou platformu informačních služeb *Dialogu* pro obchodní informace, zpravodajství a průzkum trhu:

***TBI* představuje unikátní informační zdroj např. pro:**

- tvorbu plánů strategického rozvoje
- sledování nových produktů a technologií
- sledování konkurence
- přehled obchodních a odvětvových trendů a analýz aj.

**Od června 2005 se změnilly názvy služeb poskytovaných *TBI*:**

- *Today's News* (dříve *NewsEdge*)
- *Breaking News* (dříve *Live News*)
- *News Research* (dříve *NewsRoom*)
- *Market Research* (dříve *Profound*)
- *Broker Research* (dříve *Intelliscope*)

## Today's News

Každý den poskytuje analytiku hodnocené zpravodajské obsahy (z více než 90 000 článků) na předem definovaná témata a průmyslová odvětví.

## Breaking News

Na uživatelem předem nadefinovaná témata, *Breaking News* nabízejí okamžité zobrazení titulků klíčových zpráv (*streaming headlines*) z více než 2 300 zdrojů z celého světa. Pomocí této služby lze například sledovat:

- předpovědi hospodářských výsledků
- ekonomické indikátory
- geopolitický vývoj
- události v konkrétních odvětvích
- vývoj společností a značek
- uvedení produktů na trh
- oznámení o nových zákonech a předpisech

*Breaking News* využívají např. top manažeři, marketéři, manažeři prodeje, specialisté na *competitive intelligence*, tzv. *account managers* a *strategičtí plánovači* na celém světě v celé řadě odvětví včetně médií konzultačních společností, farmaceutické, technologické a telekomunikační firmy a výrobců.

## Market Research (dříve Profound)

Nabízí interaktivní přístup do několika skupin informačních zdrojů z oblasti obchodu a ekonomiky, firemních a marketingových informací, novinového a agenturního zpravodajství či analýzy obchodních, společenských i politických klimát jednotlivých zemí světa. Jde o tzv. *one-stop-shop* službu určenou střednímu a vrcholovému managementu, analytickým a marketingovým firmám, státním institucím i politickým orgánům.

## News Research

*News Research* je rovněž *one-stop-shop*, kde se na webovém portálu nachází přes 11 000 novin, časopisů, dokumentů typu *newsletter* a podobných zdrojů obchodně-ekonomického, geopolitického a zpravodajského charakteru. Jde o jeden z nejvýznamnějších zpravodajských portálů dneška, s možností předplatného (a neomezeného využití) anebo s dílčí úhradou za každou zůžitkovanou informaci.

## Broker Research

Broker Research je jednotný a jeden z nejkompelnějších zdrojů obtížně dostupných brokerských zpráv. Tyto zprávy kombinují hluboké znalosti konkrétních odvětví, expertní názory a analýzu posledních novinek v oblasti obchodu a průmyslu. Zprávy pocházejí od více než 800 analytiků.

*Broker Research* obsahuje přes 2 miliony analytických zpráv, které pokrývají přes 38 000 společností z celého světa.

### 4.2.3 Moreover



Systém *Moreover*, vzniklý v roce 1998, aktuálně sbírá obsah z více než 25 000 vybraných webových stránek včetně zdrojů na bázi předplatného (*Wall Street Journal* a *Financial Times*<sup>34</sup>) z 200 zemí v 75 jazycích. Zpracovává 250 000 článků denně. Jako většina komerčních systémů i *Moreover* klade důraz na kvalitní agregaci relevantních dokumentů. Ke každému článku jsou přiřazena **metadata** skládající se z více než 45 polí (název článků, publikace, datum a místo vzniku aj.) Narozdíl od agregátorů typu *Google News*, *Moreover* vybírá relevantní články užitím 380 témat týkajících se zejména zpravodajství z byznysu.

### Typy zdrojů obsažených v systému *Moreover*:

<sup>34</sup> Pomocí filtru si uživatelé mohou navolit, zda mají zájem o články ze zdrojů s předplatným, nebo je chtějí eliminovat. Některé zdroje, jako například *Wall Street Journal*, lze prohlížet v rámci předplatného systému *Moreover*.



- hlavní zpravodajské pokrytí (zahrnuje: *New York Times*, *WSJ Online*, *BBC*, *Guardian*, *CNN*, *Wired*, *Bloomberg*, *AP*, *Reuters* aj.)
- portály jednotlivých průmyslových odvětví a tzv. vortály
- regionální zdroje
- mezinárodní zdroje
- zpravodajské agentury
- webové stránky výzkumných, vládních a institucionálních organizací
- firemní stránky nejúspěšnějších firem dle statistik S&P 500, FTSE 100 a FTSE 250
- diskuzní skupiny a elektronické nástěnky (*Usenet groups*, *SlashDot* aj.)
- blogy (7 milionů)
- audio a video kanály (35)
- jiné zdroje – na základě přání uživatele

Moreover zprostředkovává obsah, který je uplatnitelný v uživatelských aplikacích, webových stránkách, vyhledávacích, firemních intranetech a portálech.

## 4.3 Systémy zdarma přístupné v České republice

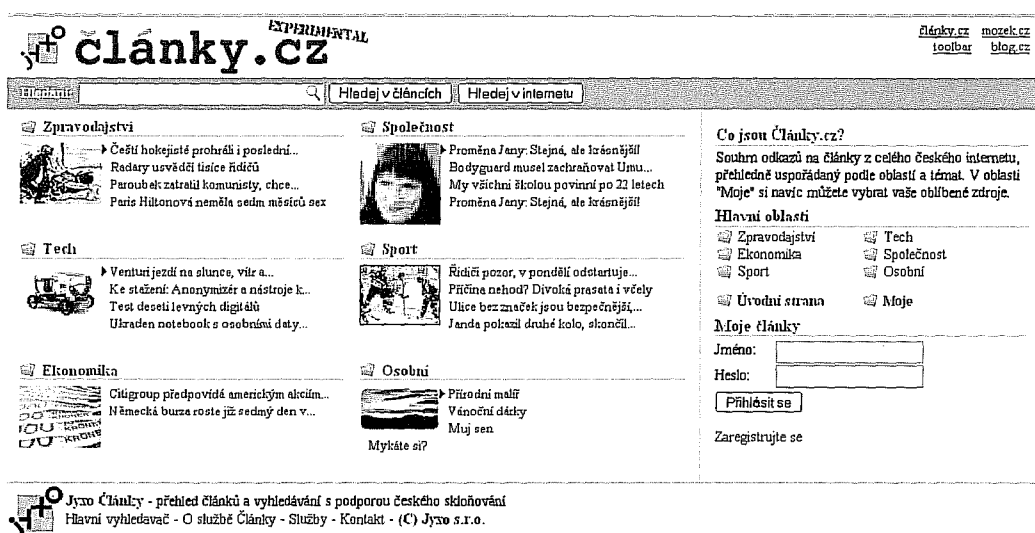
### 4.3.1 Články.cz

**články.cz** EXPERIMENTAL

Články.cz provozuje český vyhledávací systém *Jyxo*.

Články jsou indexovány od roku **2003**. Jedná se o české zprávy ze **7 000 zdrojů** (tradiční zpravodajské zdroje a blogy), indexují se také anglické články, o jejich uveřejnění se v současné době jedná. V průměru za den *Jyxo* zpracovává na **9 600 článků**, které jsou automaticky tříděné do **šesti hlavních kategorií**. V rámci kategorií jsou zprávy dále řazeny dle jednotlivých zdrojů. Související články systém nevytváří. Uživatel si může navolit zdroje svého zájmu pomocí personalizace *Moje zprávy*, kde si může nadefinovat i jakýkoliv RSS kanál.

Obr.13: Články.cz, stále ve fázi beta



#### 4.3.2 Nový den<sup>35</sup>

Český systém *Nový den*, si za poměrně krátkou dobu – **nový den** od svého spuštění, v roce 2004, vydobyl označení české *Google News*. V listopadu 2004 přešel do vlastnictví portálu *Atlas.cz*. Americkému systému se svojí koncepcí *Nový den* skutečně podobá, stahuje aktuální zprávy v krátkých intervalech a sdružuje je podle témat a zdrojů, které řadí jako většina podobných systémů podle důležitosti (v českých podmínkách je však takových systémů zatím poskrovnu), která je vytvářena podle počtu souvisejících článků<sup>36</sup>. Zpravodajství svého domácího serveru však přikládá větší váhu a tak jeho články často spadají do hlavních zpráv.

Tento systém je zajímavý také tím, že je projektem pouze jednoho muže, Jiřího Lahvičky. Agreguje články ze **150 zdrojů/rubrik** (nevyužívá *RSS*, pouze *HTML*), tzn. asi ze 30 jednotlivých serverů (pouze hlavní zpravodajské servery poskytující obecné zpravodajství). **Týdně** je zpracováno v průměru **25 000 článků** (tzn. denně zhruba 3 500 zpráv).

<sup>35</sup> <http://novyden.atlas.cz>

<sup>36</sup> Pro každou dvojici článků se vypočítává číslo od 0 do 1000, které vyjadřuje vzájemnou podobnost dvou článků, na základě kterých se pak vytváří stromová hierarchie článků, z nichž následně vznikají finální klastry – související články.

Agregované texty článků systém používá jenom pro zpracování, uloženy jsou pouze statistické údaje, např. o výskytu jednotlivých slov, titulek a úvod článku pro použití jako anotace. I tato data jsou zpravidla mazána po 60 dnech, takže žádný archiv zde vytvářen není. Jiří Lahvička to zdůvodňuje tím, že většina médií má ve smlouvě s ČTK, že články musí po 30 dnech mazat, což znamená, že čím starší článek, tím spíše už není dostupný ani v původním médiu. Nový den nemá explicitní taxonomii, každou skupinu článků interně řadí do jedné až dvou ze šesti základních rubrik (na základě toho, ve kterých rubrikách sledovaných zdrojů jsou články z dané skupiny).

Obr. 14: Agregace zpráv Nový den na Atlasu

The screenshot shows the Atlas.cz website interface. At the top, there is a search bar with the text "Hledej" and a "přihlásit" button. Below the search bar, there is a navigation menu with categories: Domov, Svět, Ekonomika, Sport, Kultura a společnost, Věda a technika, and O Novém dni. The main content area is titled "Hlavní zprávy" and features a prominent news item: "Topolánek: O důvěře vládě by se mělo hlasovat do půlky ledna". This article includes a sub-headline "Atlas Zprávy, před 35 minutami" and a brief summary: "Premiér a předseda ODS Mirek Topolánek soudí, že sněmovna by mohla o důvěře kabinetu hlasovat do poloviny ledna. Prezidentovi by chtěl návrh nové vlády předložit nejspíše ve čtvrtek. Uvedl to v pořadu České televize (ČT) Otázky Václava Moravce. Vedení...". Below the summary are three links: "Vláda půjde pro důvěru v lednu, slíbil Topolánek (iDNES)", "Poslanci rozhodnou o důvěře do půlky ledna (Lidovky)", and "Topolánek: pro důvěru si půjdeme do poloviny ledna (Novinky)". To the right of the article is a small image of a man in a suit, with the caption "České noviny".

Below the main article, there are two columns of news items. The left column is titled "Domov" and features the article "Opakované volby v Havířově vyhráli komunisté" with a sub-headline "iDNES, před 7 hodinami" and a brief summary: "Opakované komunální volby v Havířově vyhrála KSČM. Dostala 30,79 procenta hlasů. Druhá se ziskem 28,41 procenta skončila ODS, která v říjnovém hlasování zvítězila. K urnám podle volebního serveru přišlo necelých 22". To the right of this article is a small image of a hand holding a ballot, with the caption "České noviny".

The right column is titled "Svět" and features the article "Abbás se sešel s volební komisí: volby nejdříve za tři měsíce" with a sub-headline "Atlas Zprávy, před 2 hodinami" and a brief summary: "Předseda palestinské samosprávy Mahmúd Abbás se dnes sešel s ústřední volební komisí kvůli detailům přípravy předčasných prezidentských a parlamentních voleb, které vyhlásil v sobotu. Přes velké protesty vládního hnutí Hamas,". To the right of this article is a small image of a group of men, with the caption "České noviny".

### 4.3.3 *Krátce.cz*<sup>37</sup>



Krátce.cz bylo vytvořeno v září 2003. Zpřístupňuje články z 320 jednotlivých serverů. Systém umožňuje vyhledávání i v starších zprávách, ale jen v úvodech (perexech) zpráv, jelikož celé články systém neuchovává – pouze jejich odkazy.

*Krátce.cz* mezi zdroje zahrnuje také **blogy**, které jsou v samostatné rubrice *Weblogy*, v kterých je 65 zdrojů, jedná se ale většinou o osobní, tematicky nevyhraněné blogy.

Ostatní tematické blogy jsou zařazené v příslušné rubrice, kterých je v poměru k ostatním zdrojům zhruba jedna třetina. V průměru agregátor zpracuje **650 článků za jeden den**.

Tvůrci *Krátce.cz* plánují kompletní předělání způsobu třídění zdrojů do kategorií. V současné době existují dvě úrovně kategorií. Nicméně některé zdroje nejsou úplně zřejmě zařaditelné a proto se v určité kategorii nemusí objevit články na dané téma, jelikož se jimi daný zdroj zabývá pouze okrajově. Nový systém kategorizace bude pouze jednoúrovňový a půjde jen o hrubé rozdělení zdrojů. Druhou úroveň budou tvořit témata, kterých bude podstatně více než je v současných kategoriích, ty ovšem nebudou rozdělovat zdroje, ale samotné články.

Vše bude probíhat automaticky na základě analýzy slov obsažených v článku.

Například v tématu *Linux* se zobrazí i články o *Linuxu* vydané obecnými zdroji ne jenom technickými.

---

<sup>37</sup> <http://www.kratce.cz/>

## 4.4 Komerční systémy v České republice

### 4.4.1 NewsTin.com



*NewsTin.com* je nově vznikající projekt společnosti *I2S a.s.* se sídlem v Praze.

Společnost *I2S* byla založena v roce 1998 a její hlavní specializace se týká dvou okruhů činností:

- *Text Mining* (vytěžování informací<sup>38</sup>)
- *Defence & Security* (obrana a bezpečnost)

Od poloviny roku 2005 společnost *I2S*, v rámci své specializace na metody typu „*text mining*“, započala s vývojem systému *NewsTin* zabývajícího se agregací zpráv. Systém *NewsTin* nemá v evropských podmínkách zatím žádného obdobného konkurenta a na světovém trhu se snaží konkurovat i těm největším.

V současné době podporuje **6 jazyků v 7 edicích** a plánuje další jazykové rozšíření.

Agregace zpráv v systému *NewsTin* je založena na hluboké taxonomii, která se skládá z velkého množství kategorií (témat), jejichž rozšíření bude čítat až **100 000 jednotlivých kategorií**. Každá kategorie je zde nadefinována speciálními vzorci pro identifikaci témat, kontextu a entit, vytvořenými společností *I2S*, které jsou manuálně definovány pro všechna témata ve všech jazycích separátně.

Díky této důkladné práci je uskutečnitelné následné automatické třídění relevantních článků do náležitých kategorií. Definice těchto kategorií vytváří tým tzv. taxonomistů, kteří jsou specialisty na jazyk, v němž jednotlivá pravidla tvoří.

Pro rozpoznání jazyka, lematizaci a následné skloňování slov systém *NewsTin* využívá softwarovou aplikaci společnosti *Inxight Software Inc.* (společnost *I2S* je jejím *OEM* partnerem), přední světové firmy v oblasti vytěžování informací a znalostního managementu.

---

<sup>38</sup> Český ekvivalent není ustálený.

Články systém agreguje z běžných *HTML* stránek a prostřednictvím komunikačních formátů *RSS*, *ATOM* a *RDF* z předem nadefinovaných zdrojů, které pravidelně prohlíží.

Kapacita systému umožňuje zpracovat až **1 000 000 článků denně**.

Systém *NewsTin* má dynamickou platformu s možností aplikace i na jiné systémy.

### **Metodika základních postupů agregace zpráv systému *Newstin***

- **nalezení zdroje a jeho klasifikace** – nalezení a identifikace zdroje (periodik, blogů a dalších zdrojů), jejich klasifikace dle geografické lokace, typu dokumentu a kvality
- **automatická identifikace jazyka** stažených článků
- **vyhledávání** – tvorba plnotextového indexu pro vyhledávání, který také podporuje pokročilé vyhledávací funkce
- **podobnost článků** – analýza podobnosti článků umožňuje tvorbu klastrů souvisejících článků
- **reduplikace** - ověřování a odstraňování duplicit
- **tvorba metadat** - identifikace různých důležitých informačních elementů, na základě kterých lze rozpoznat, jestli se jedná o jména osob, názvy firem, geografických míst či produktů
- **indexace**
- **automatická kategorizace** – kategorizace dokumentů do tisíců tematických, logických a hierarchických kategorií
- **konečná podoba webové aplikace** – intuitivní a uživatelsky přívětivé rozhraní pro koncové uživatele

## **Popis webového rozhraní *Newstin.com* (viz obr. 15):**

**1. Registrace** – v rámci registrace uživatel získá možnost personalizace zájmových témat (*My Topics* – viz níže)

**2. Nastavení (*Settings*)** – na základě registrace uživatel dále může definovat:

- jazykovou edici (momentálně je k dispozici americká, britská, německá, francouzská, španělská, italská a česká)
- typ edice (volně přístupná či profesionální; volně přístupná edice nepřístupňuje velkou část tématu – byznys)
- způsob načítání článků (v rámci stejného či nového okna)
- vlastní témata (*My Topics*)

**3. Hlavní stránka (*Front Page*) nabízí:**

- šest hlavních témat (*World, Domestic News, Politics, Business, Technology a Health*)
- hlavní zprávy (*Top Stories*) – zde se zobrazují články s největším počtem souvisejících článků, tj. o kterých se nejvíce píše

**4. Moje témata (*My Topics*)** – zde si uživatel nadefinuje svá oblíbená témata z nepřeberného množství kategorií, které si může vybrat při prohlížení témat (viz bod 6)

**5. Ukazatel nadřazených kategorií** – díky této informaci uživatel snadno zjistí, kde se zrovna nachází, či se případně může inspirovat nadřazenými kategoriemi, podřazená témata se mu zobrazí v pravém sloupci jako články týkající se tématu (viz bod 9)

**6. Procházení témat (*Browse Topics*)** – umožňuje prohlížení všech témat

**7. Vyhledávání** – vyhledávat lze dvěma způsoby:

- v tématech (*Search Topics*)
- ve všem (*Search All*)

**8. Změna jazyka** – daná vlajka uživatele odkáže na dokumenty momentálně prohlíženého tématu v příslušném jazyce

9. Související témata (*Related Topics*) – nabízí podřazená témata týkající se aktuálně prohlížené kategorie

10. Zdroje (*Sources*) – počet zdrojů podobných danému článku

Obr.15: Webové rozhraní systému Newstin

The screenshot shows the Newstin web interface. At the top right, there are links for 'LOG OUT', 'HELP', 'PRODUCTS', and 'SETTINGS' (1). Below the navigation bar, there are links for 'Front Page', 'Top Stories', 'World', 'U.S.', 'Politics', 'Business', 'Technology', and 'Health' (3). A search bar is located on the right side (7). The main content area features a section titled 'Business Intelligence' (5) with several articles. The first article is 'Visibility's Business Intelligence Profiled in Manufacturing Computer ...' (10), which includes the text: 'For manufacturers, BI (business intelligence) should be about quantifying business assumptions that were previously too costly and time-consuming to challenge. ...'. Other articles include 'UL Announces New Qualified Contractor Program' and 'DIAGNOS and its Partner Cairnaly Deliver Advanced emailing Analytics ...'. On the right side, there is a 'RELATED TOPICS' section (9) with sub-sections for 'Competitive Intelligence' and 'Industrial Espionage'. The 'Competitive Intelligence' section includes an article titled 'HITIS - Life Science & Biotech Intelligence Expands Web Presence with ...'. The 'Industrial Espionage' section includes an article titled 'GFI white paper warns of data theft dangers through 'pod slurping''. A search bar is also present in the top right corner (7).



#### 4.4.2 Právě dnes<sup>39</sup>



*Právě dnes* monitoruje celkem **1 600 zdrojů** (z České republiky, Slovenska a USA), celý seznam jednotlivých zdrojů je možné najít na hlavní stránce.

Články není možné řadit podle relevance, jsou řazeny pouze dle data agregace. Vedle titulků článků se zobrazují barevné čtverce, které indikují stáří jednotlivých zpráv:

- nejnovější zpráva, ■ dnešní, ■ max. 3 dny stará, □ starší

Zprávy jsou tříděny do celé řady oblastí:

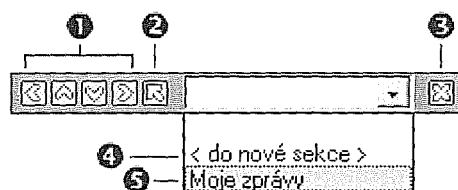
*z domova, ze světa, ze sportu, z ekonomiky, techniky, politiky, kultury, dopravy, z regionů, z bulváru, pro studenty, o podnikání, o marketingu, o počítačích aj.*

Systém neagreguje pouze články, ale také např. informace *o cenách* (v rámci rubriky obchody), *kurzy měn a akcie* (ekonomie).

Uživatel má možnost, stejně jako ve většině systémů agregace zpráv, nastavení svého personalizovaného rozhraní *Můj Právě dnes*.

Po povinné registraci se uživatel může pustit do úprav své osobní stránky.

Stiskem tlačítka pro úpravu polohy informačního boxu se nad ním zobrazí editační lišta (viz obr. 16), pomocí které lze přesouvat box v rámci jedné stránky, vytvářet nové sloupce a nové sekce.



Obr. 16 – Editací lišta

Významy jednotlivých tlačítek na editační liště jsou následující:

1. tlačítka pro přenos boxu v rámci jedné stránky
2. přenos boxu z podsekce do nadřazené sekce
3. odstranění boxu z osobní stránky
4. vytvoření nové podsekce a následné přesunutí boxu do ní
5. vybráním názvu podsekce se box do ní přesune

<sup>39</sup> <http://www.pravednes.cz/>

Vedle zdarma přístupného rozhraní na serveru *Právě dnes*, jsou poskytovány dvě hlavní komerční služby:

- MonitorMAIL
- MonitorBOX

### **MonitorMAIL**

Služba nabízí průběžné zasílání informací na předem nadefinované téma prostřednictvím elektronické pošty. Nalezené odkazy se hromadí na serveru ve zvláštní schránce. Schránka je v pravidelných intervalech kontrolována a pokud nějaké odkazy obsahuje, je z nich vytvořen souhrnný email, který je následně odeslán.

### **MonitorBOX**

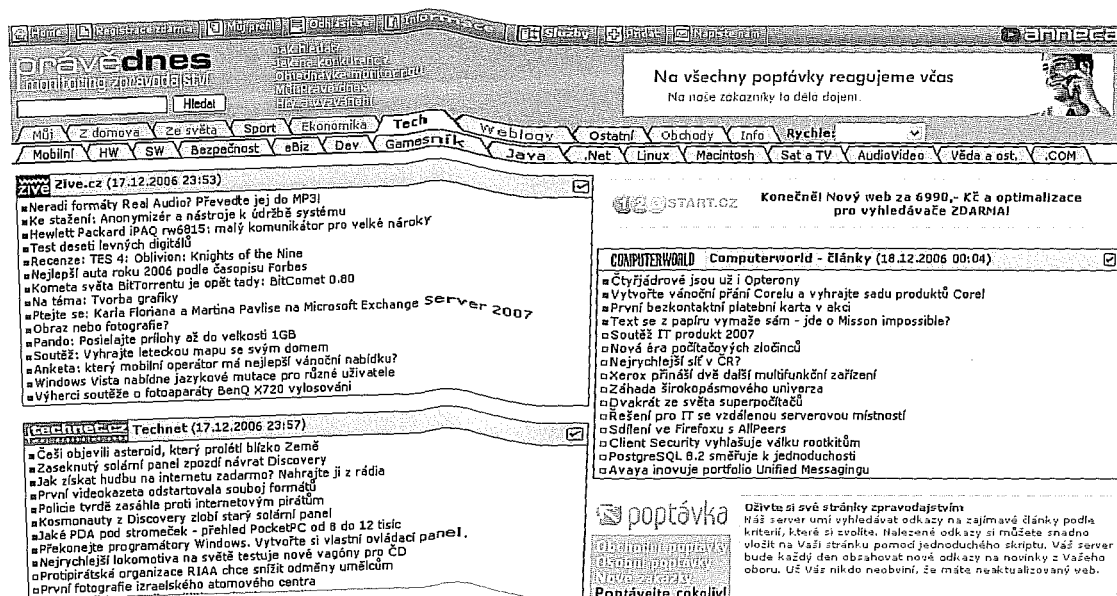
Služba je vytvořena pro implementaci na jiné informační systémy. Podobně jako u služby *MonitorMAIL* se i zde nalezené odkazy článků týkající požadovaného tématu shromažďují na serveru, kde připravují dávku, kterou si vzdálený systém může v pravidelných intervalech stahovat.

Tyto dvě služby jsou poskytovány i zdarma, ale vyhledávání v člancích je omezeno pouze na hledání v titulcích zpráv. Filtrování jednotlivých zpráv je možné pouze podle jazyka.

Narozdíl od placených služeb, kde lze nastavit filtrace nejen podle jazyka, ale i podle kategorie a zdroje, navíc je zde poskytována asistence nastavení. Systém vyhledává i v anotacích a plných člancích.

Podle portálu *Toplist*, který vytváří statistiky návštěvnosti webových stránek, v listopadu 2006 systém *Právě dnes* navštívilo 367 900 unikátních návštěvníků. Nejvíce uživatelů začíná čtení článků na hlavní stránce, další zajímá zejména technická rubrika poté bulvár.

Obr. 17: Webová stránka technické rubriky portálu Právě dnes



#### 4.4.3 Newton I.T.



Firma *Newton I.T.*, založena v roce 1994, je největší společností v České republice poskytující elektronický monitoring médií a mediální analýzy. Zabývá se shromažďováním, zpracováním, ukládáním, vyhledáváním a distribucí informací z médií v elektronické i tištěné podobě. Svoji webovou platformu pro práci s monitoringem médií spustila jako jedna z prvních na světě v roce 2000.

Zprostředkovává informace z těchto typů zdrojů:

- tištěné noviny a časopisy z České a Slovenské republiky a Polska (celkem 850 titulů)
- celoplošné deníky včetně regionálních mutací
- regionální deníky
- časopisy – ekonomické, odborné, počítačové, automobilové, společenské
- tištěná zahraniční média (10 000 světových, zejména západoevropských a amerických titulů)
- televizní a rozhlasové stanice

- 100 hodin doslovně přepsaného televizního vysílání denně
- 100 hodin doslovně přepsaného rozhlasového vysílání denně
- agenturní zpravodajství
- kompletní agenturní zpravodajství ČTK
- agenturní zpravodajství TASR
- webové zpravodajství (na základě dohody s vydavateli zajištěny plné texty z 300 zpravodajských internetových serverů)

Archiv veškerých těchto zdrojů sahá do roku 1993.

*Newton I.T.* nabízí tyto produkty a služby:

- **Media monitoring dokument**<sup>40</sup> – vybrané zprávy jsou zasílané emailem
- **Media monitoring software** – nástroj pro další analýzu vybraných zpráv
- **Internetový media monitoring** – výběr zpráv dostupný na internetu
- **Media monitoring Lotus Notes** – výběr zpráv v prostředí Lotus Notes
- **Media monitoring PLUS** – anotace zpráv, denní přehledy témat, překladatelské služby
- **MediaSearch** – mediální databáze online
- **Mediální analýzy** – služby s přidanou hodnotou
- **Databáze hospodářských informací** – zprávy z tisku pouze s ekonomickou tematikou
- **Český a Slovenský výběr** – denní komentovaný přehled hlavních zpráv z celostátních deníků
- **Mediální výběry** – denní nebo týdenní specializované výběry nejdůležitějších zpráv z vybraných oborů

---

<sup>40</sup> Názvy služeb jsou vlastními názvy společnosti Newstin I.T. Domnívám se, že se jedná o nevhodnou kombinaci češtiny a angličtiny.

- **Tender monitoring** – přehled aktuálních výběrových řízení
- **Služby pro vydavatele** – tvorba elektronických archivů

### Media monitoring dokument – MMD

Představuje základní formu monitoringu médií. Zahrnuje přehled nalezených zpráv dle nadefinovaných témat, doplněný o plné zdrojové texty se zvýrazněním klíčových slov, která zadané téma vystihují. Tyto dokumenty jsou zasílány elektronicky nebo tištěně dle předdefinovaných formátů (DOC, RTF, TXT, či HTML).

Obr.18: Ukázka služby Media monitoring dokument

**NEWTON MEDIA MONITORING**  
INFORMATION TECHNOLOGY

**Přehled zpráv 29.8.2006 - 8.9.2006**

Section	Date	Source	Page	Count
<b>Tiskový mluvčí Euro</b>	4.9.2006	<u>Obchodník pro veřejný sektor</u>	str. 29 Lidé a podniky (nad)	1
				Tiskový mluvčí
<b>Hospodářské noviny</b>	4.9.2006	<u>Cikni bude dělat mluvčího Julinkovi</u>	str. 4 Z domova (veš)	2
				Tiskový mluvčí
<b>PERSONÁLIE</b>	8.9.2006	<u>PERSONÁLIE</u>	Hospodářské noviny str. 17 Z domova (žtk)	3
				Tiskový mluvčí
<b>Lobby</b>	29.8.2006	<u>Personálie</u>	Lobby str. 4 Stalo se	4
				Tiskový mluvčí
<b>Marketing magazine</b>	5.9.2006	<u>Marketing magazine</u>	str. 54 Management & Know - How - Hlá v pohybu	5
				Tiskový mluvčí

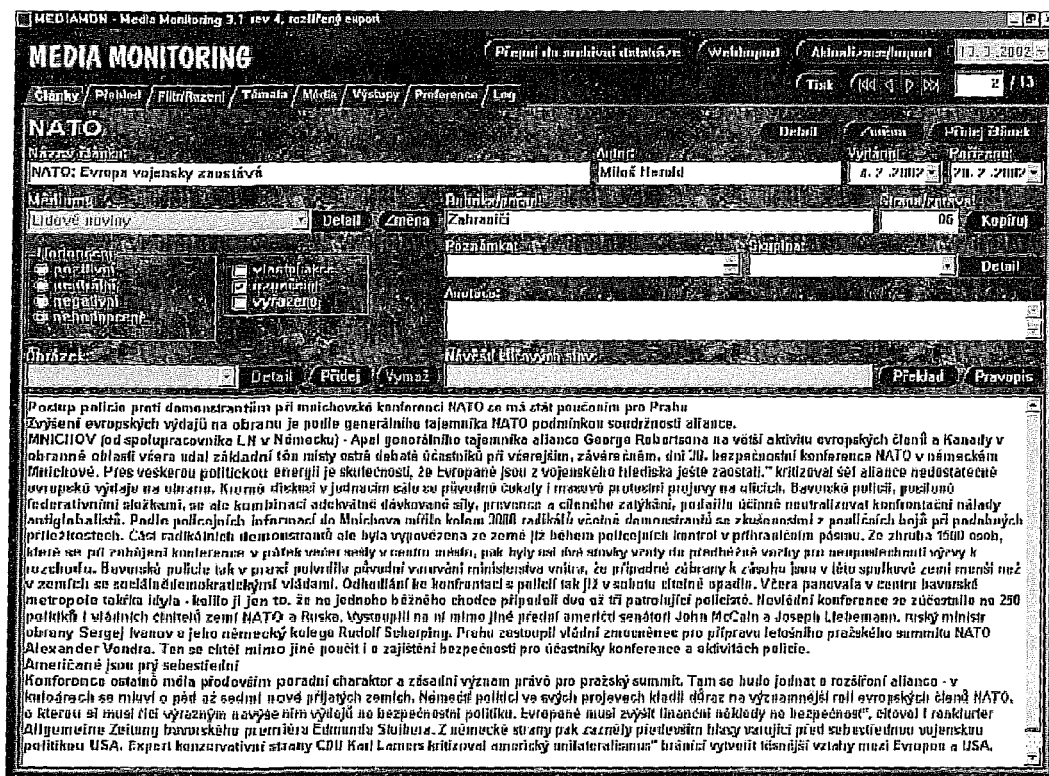
Hotovo    Adblock    Nyní: Slunečno, 12° C    Pon: 17° C    Úte: 15° C

### Media monitoring software

Newton I.T. poskytuje vlastní vytvořenou softwarovou aplikaci, která umožňuje:

- hodnocení jednotlivých článků
- doplnění vlastních poznámek
- filtrování a třídění dle libovolných kritérií
- doplnění vlastní či objednané anotace a překladu
- zvýraznění klíčových slov
- výstupy v RTF nebo HTML formátech
- archivování zpráv
- tisk tematických výběrů

Obr.19: Aplikace služby Media monitoring software



## Internetový media monitoring

Newton I.T. poskytuje databázový systém přístupný přes webové rozhraní, který umožňuje:

- pravidelnou aktualizaci
- možnost integrace do firemních intranetů
- rozdělení zpráv do témat určených uživateli
- přizpůsobení designu
- zaslání zpráv emailem dle zadaných kritérií
- možnost zajištění anglické anotace a překladů zpráv
- zvýraznění klíčových slov
- vícejazyčné uživatelské rozhraní (česky, anglicky, slovensky a polsky)

Obr.20: Internetový media monitoring

**APRA**  
Asociace Public Relations Agentur

**MEDIÁLNÍ SLUŽBY NEWTON I.T.**

**MONITORING MÉDIÍ**  
Poslední aktualizace: 16.10.2006

Jednoduché vyhledávání

rozšířené hledání

Data jsou aktualizována pravidelně každé pondělí.

**Témata:**

	nové	vše
■ APRA	22	4847
■ PR	21	4590
■ LSPR	0	162
■ PR Klub	0	294
■ Akademie PR	0	23
■ AKA	0	425
	43	10331

**Služby:**

**APRA**

**Komunální a senátní volby**  
Mladá fronta DNES | 14.10.2006 | 19:15 | pořad: Události

...Senátní volby jsou podle odborníka na public relations Vladimíra Bystrova letos nevýrazné. Případní senátoři sází pouze na svou tvář. Prezenci komunálních kandidátů pochválil. Vladimír BYSTROV, člen výkonné rady Asociace PR agentur APRA - Komunální volby mají kampaň letos v mých o...

**V Liberci soutěží evropské krásky**  
Mladá fronta DNES | 14.10.2006 | rubrika: Kraj Liberecký | strana: 4 | autor: (sol)

...Studuje obor public relations na soukromé škole mezinárodních a veřejných vztahů a doufá, že ve finále se jí podaří získat cenné umístění. „Do soutěže mě přihlásila moje agentura. Kolektiv byl ze začátku velmi dobrý, teď před finále se už ale začíná objevovat učitá rivalita mezi děvčaty,“ prohlásila...

**Stavba obchodního centra může začít**  
Mladá fronta DNES | 12.10.2006 | rubrika: Kraj Pardubický | strana: 4 | autor: MILAN ZLINSKÝ

...Jsme rádi, že jsme se s hotelem nakonec domluvili,“ řekl Tomáš Staněk z firmy Ogilvy Public Relations, která investorovi zajišťuje propagaci. Šérové hotelu Labe se začátkem září na poslední chvíli odvolali proti stavebnímu povolení. ...

**Blažek z Letiště skákal padákem**  
Mladá fronta DNES | 11.10.2006 | rubrika: Lidé | strana: 10 | autor: (ježi)

...Ale od prvních sekund po seskoku z letadla jsem si let vzduchem začal náramně užívat,“ tvrdí představitel Pavla Kučery, který je v seriálu Letiště šéfem Aeroklubu a letecké školy. K seskoku v tandemu přemluvil Blažek také svou přítelkyni Jolanu, s níž má dvouletou dceru Johanku. ...

## **Media monitoring Lotus Notes**

Jedná se o aplikaci s vybranými funkcemi služby *Media monitoring software*.

Toto řešení nabízí:

- práci v síti včetně nastavení přístupových práv a využití všech obecných funkcí systému Lotus Notes
- rozdělení zpráv podle uživatelem určených témat
- možnost hodnocení
- tvorbu poznámek
- filtrování a třídění dle libovolných kritérií
- doplnění vlastních či objednaných anotací a překladů
- zvýraznění klíčových slov

## **Media monitoring PLUS**

Tato služba nabízí tvorbu:

- anotací – možnost implementace do internetového media monitoringu
- denních přehledů – obsáhlejší anotace z více zpráv najednou
- překladatelské služby (anotace, denní přehledy, plná znění zpráv, mediální analýzy)

## **MediaReport**

Vytvořeny pro vyhodnocování a měření PR aktivit. Nabízí každý měsíc shrnutí nejdůležitějších informací tabulkovou formou.

Služba *MediaReport* poskytuje:

- přehled o výsledcích PR aktivit klienta
- informaci o počtu článků a jejich hodnotě v jednotlivých skupinách médií i v konkrétních titulech



## **MediaSearch – mediální databáze**

Služba *MediaSearch* nabízí:

- online vyhledávání v archivu novin, časopisů a přepisů televizních a rozhlasových pořadů
- fulltextové vyhledávání v databázi více než 600 tištěných zdrojů, ve zpravodajství stovek internetových serverů a přepisech téměř 30 televizních a rozhlasových stanic z České republiky, Slovenska a Polska
- retrospektivní vyhledávání do roku 1996

## **Mediální analýzy**

Mediální analýza je strukturovaným rozbořením toho, jaký obraz vytvářejí média o sledovaném subjektu či problematice ve stanoveném časovém úseku. Je nadstavbou klasického monitoringu médií a rozebírá z různých hledisek chování médií vůči předmětu analýzy. Mapuje obsahovou, tematickou a časovou strukturu mediální prezentace určitého subjektu během stanoveného období na zvoleném vzorku médií. Subjektem analýzy přitom může být například společnost jako celek, nějaká událost, osoba či místo. Mediální analýzy jsou vytvářeny dle individuálních požadavků klientů, kteří mají možnost výběru z desítek posuzovaných kritérií, jež odpovídají standardní mezinárodní metodice. Analýzu lze podobně jako monitoring vypracovávat pravidelně či jednorázově z historických dat.

- Internet média analyzér
- Internet média analyzér je webová aplikace umožňující:
- interaktivní práci s daty
- ohodnocení či označení důležitých informací a třídění do vlastních složek
- přidání poznámek soukromých či veřejných
- zobrazování výstupů do tabulek a grafů
- export do formátu *HTML* nebo *RTF*

## **Databáze hospodářských informací – DHI**

Poskytuje tematický soubor hospodářských i makroekonomických zpráv zaměřených na konkrétní ekonomické subjekty působící v České republice. Slouží ke sledování aktuálního stavu ekonomické situace jednotlivých společností – i jeho vývoje v čase. Produkt je určen především pro podnikatelskou sféru, finanční instituce, obchodníky na finančním trhu, poradenské firmy apod. Obsahuje výběr plných znění:

- aktuálních i historických zpráv z celostátních a regionálních periodik. Do *DHI* se nezařazují duplicitní články publikované v jednotlivých médiích a některé méně významné zprávy.

### **Český a slovenský výběr**

*Český a slovenský výběr* nabízí manuální výběr zpráv a událostí z hlavních českých a slovenských deníků společně s komentovaným souhrnem pro rychlou orientaci v aktuálních politických a ekonomických tématech. Český výběr vychází v české, anglické a německé mutaci.

### **Český oborový výběr**

Poskytuje přehled o dění ve vybraných oborech. Přináší:

- denní přehled analytiky vybraných zpráv (*nejdůležitějších*), o kterých média v daných odvětvích informují
- týdenní souhrny
- výběr z několika jazykových variant (česky, anglicky, německy)

Služba je poskytována pouze v těchto vybraných oborech:

- automobilismus
- bankovníctví
- energetika

- farmacie
- nápoje
- telekomunikace

### **Tender monitoring**

Poskytuje informace o nových obchodních příležitostech. Přináší přehled aktuálních výběrových řízení, poptávek a dražeb.

**Další projekty Newton I.T., zdarma přístupné na internetu:**

**Mediainfo.cz** – *analytický web o českých médiích*

- aktuální informace o medializaci více než 800 osobností z řad politiků a umělců
- možnost vlastního výběru a porovnávání jednotlivých osobností
- pravidelné obsahové analýzy internetových serverů a televizního zpravodajství
- volně publikovatelné analýzy řady vybraných témat

**Headlines.cz** – *internetový magazín pro širokou veřejnost*, členěný do několika desítek témat. Jedná se o elektronickou podobu zpráv z tištěných médií, celoplošných deníků, vybraných ekonomických, politických či společenských týdeníků a dále plné přepisy vybraných televizních pořadů. Z těchto médií jsou na serveru Headlines.cz publikovány pouze vybrané (*nejvýznamnější*) zprávy. Jednotlivá témata je možné také získávat elektronickou poštou. Tato možnost je nabízena i pro tvůrce webových stránek, kteří mohou vybraná témata bezplatně umístit na svůj server.

[Newton Information Technology, 2006]

## 5 ZÁVĚR

S rozvojem internetu se budou jistě dál rozvíjet i agregační služby. Monitoring elektronických zdrojů je minimálně stejně důležitý jako monitoring tištěných publikací.

A významu bude nabývat, jelikož v současné době na internetu vznikají nové a nové organizace, jako i elektronické publikace (neexistující v tištěné podobě).

V digitální ekonomice potřebují nové organizace efektivní prostředky, aby se staly úspěšnými a byly tak schopné obstát v konkurenci. Proto je důležitý management organizací, kterému J.B. Harreld [HARRELD, 1998, s. 60-76] přikládá důležité role v nové ekonomice.

Jedněmi z nich jsou:

- řízení informačních toků
- řízení inovací a tvorby intelektuálního vlastnictví

K tomu bude využita celá řada technologických nástrojů a metod, zejména:

- vytěžování médií (*media mining*)
- zdroje znalostí (*knowledge repositories*)
- nástroje spolupráce a součinnosti (*collaborative tools*)

[PAPÍK, 2001].

Webová agregace zpráv se pro mnoho organizací stává nepostradatelným nástrojem pro řízení firemní reputace (*reputation management*) a konkurenčního zpravodajství (*competitive intelligence*). Stejně tak má svůj význam pro jednotlivce, kteří nemusí navštěvovat velké množství zpravodajských serverů, protože jim systém všechny články zpřístupní v rámci jednoho rozhraní.

Webové agregátory ocení i uživatelé, které zajímá co o daném problému píšou různé informační zdroje, rovněž tak jim pomohou vyhodnocovat důležitost zprávy, která je evidentní právě podle počtu jejich uveřejnění na různých serverech.

Zpravodajské agregátory zaznamenávají nárůst návštěvníků a pro většinu uživatelů se stávají „výchozím bodem“. Tedy domácí stránkou, z níž teprve

přecházejí na konkrétní webové stránky svého zájmu. Dá se říci, že v tomto směru změnila technologie agregace zpráv způsob užívání internetu.

Tzv. sociální sítě a folksonomie zmíněné v kapitole 4.2 by mohly ve vývoji webové agregace zpráv hrát také určitou roli. Stále častěji se mluví o nové koncepci *Web 2.0*, která do centra zájmů klade uživatele a využívá výhody jeho možné interakce. Proto je pravděpodobné, že bude kladen čím dál větší důraz na systémy typu *Digg.com*, jak ukazuje jejich popularita již nyní, kdy svojí návštěvností předčí i nejužívanější agregátory zpráv jako např. *Google News*.

Takové systémy by ovšem nemohly fungovat bez důvěry k uživatelům, kteří se na jejich tvorbě spolupodílejí (např. vkládáním článků). Vždy se totiž najdou jedinci, kteří mohou této důvěře zneužít ve svůj prospěch či jen narušit funkčnost systému, ať už vkládáním reklamy, nebo jen nesprávným zařazením jednotlivých článků.

Proto budou mít na trhu stále význam systémy, které jsou méně otevřené vůči uživatelské interakci a staví na týmu profesionálů spolupracujících na kvalitní organizaci informací.

## SEZNAM POUŽITÉ LITERATURY

Albertina icome. *AiP : Albertina icome Praha* [online]. Praha : c2003-2006 [cit. 2006-12-01]. Dostupný z WWW: <<http://www.aip.cz>>.

...

BARTOŇ, Petr, LAHVIČKA, Jiří. *SkrzNaSkrz* : [www.SFERA.cz](http://www.SFERA.cz). *Lupa : server o českém internetu* [online]. 2000, č. 9 [cit. 2006-11-11]. Dostupný z WWW: <<http://www.lupa.cz/clanky/skrznaskrz-www-sfera-cz/>>. ISSN 1213-0702.

...

BEDNÁŘ, Vojtěch. *RSS kamarád do deště . PC World.cz : magazín digitálního věku* [online]. 2004, roč. 12 [cit. 2006-11-11]. Dostupný z WWW: <<http://www.pcworld.cz/pcw.nsf/c5a232a5bdf1591fc1256efd00563e9b/94c5d80626f12cec1256fa1004a80e1?OpenDocument>>.

...

BEDNÁŘ, Vojtěch. *RSS v moderním světě: syndikace má budoucnost. Lupa : server o českém internetu* [online]. 2006, č. 2 [cit. 2006-12-07]. Dostupný z WWW: <<http://www.lupa.cz/clanky/rss-v-modernim-svete-syndikace-ma-budoucnost/>>. ISSN 1213-0702.

...

BUREŠ, Jiří. *RSS 2.0. Interval.cz : webdesign a e-komerce denně* [online]. 2004 [cit. 2006-11-11]. Dostupný z WWW: <<http://interval.cz/clanky/rss-2-0/>>. ISSN 1212-8651.

...

CNNMoney.com. *Fortune500 : Our annual ranking of America's largest corporations* [online]. Florida : CNNMoney.com, 2006 [cit. 2006-12-01]. Dostupný z WWW: <<http://money.cnn.com/magazines/fortune/fortune500/>>. ISSN 0015-8259.

...

DAYTON, Denise. *All the News That's Fit to Print (And Then Some!) : Google News or Yahoo! News. Searcher*. 2005, vol. 13, no. 2, s. 55-57.

...

DEUZE, Mark. *Online Journalism : Modelling the First Generation of News Media on the World Wide Web. First Monday : Peer-reviewed Journal on the Internet* [online]. 2001, vol. 6, no. 10 [cit. 2006-11-11]. Dostupný z WWW: <[http://firstmonday.org/issues/issue6\\_10/deuze/index.html](http://firstmonday.org/issues/issue6_10/deuze/index.html)>. ISSN 1396-0466.

...

o

DEUZE, Mark. *The web and its journalism : considering the consequences of different types of newsmedia online. New Media and Society*. 2005, vol. 5, no. 2, s. 203-230.

...

Dow Jones Reuters Business Interactive (US). Polyarchical data indexing and automatically generated hierarchical data indexing paths. Jonathan Guy Grenside Cook aj. (inventors). MPT6: G06F 017/30. United States Patent, 6938046. 2001-03-02.

...

FOSS, C.L. Tools for reading and browsing hypertext. *Information Processing and Management : an International Journal*. 1989, vol. 25, is. 4, s. 407-418.

...

Google. *Google news* [online]. Mountain View (California). c2006 [cit. 2006-12-01]. Dostupný z WWW: <<http://news.google.com>>.

...

GRÜNDEL, Hans, et al. Clipping and Analyzing News Using Machine Learning Techniques. In *Discovery Science : 4th International Conference, DS 2001, Washington, DC, USA, November 25-28, 2001. Proceedings*. Heidelberg : Springer Berlin, 2001. 13 s. Dostupný z WWW: <<http://www.springerlink.com/content/d6jt0jq30uh6bc86/>>. ISSN 0302-9743.

...

HANE, Paula J. Factiva to Launch Search 2.0 Beta. *Information Today* [online]. 2006, vol. 1 [cit. 2006-12-01]. Dostupný z WWW: <<http://www.infotoday.com/newsbreaks/nb060116-2.shtml>>. ISSN 8755-6286.

...

HARRELD, J. B. Building smarter, faster organizations. In *Blueprint to the digital economy. Creating wealth in the era of e-business*. New York : McGraw Hill, 1998, s. 60-76.

...

HENDERSON, Pat. Clippings to Computers : The Great Newspaper Migration. *Library Mosaics*. 2001, vol. 12, no. 3, s. 17-18. ISSN 1054-9676.

...

HLAVENKA, Jiří. Trafika přichází o Mladou Frontu : co bude dál?. *Živě.cz* [online]. 2000, roč. 9 [cit. 2006-11-11]. Dostupný z WWW: <<http://www.zive.cz/h/Uzivatel/AR.asp?ARI=17308&CAI=2114>>. ISSN 1214-1887.

...

HORRIGAN, John. *Online News : For many home broadband users, the internet is a primary news source*. Washington, 2006. Výzkumná zpráva. Pew Internet & American Life Project. Dostupný z WWW: <[http://www.pewinternet.org/pdfs/PIP\\_News.and.Broadband.pdf](http://www.pewinternet.org/pdfs/PIP_News.and.Broadband.pdf)>.

...

CHOWDHURY, Sudatta, LANDONI, Monica. *News aggregator services : user expectations and experience*. Glasgow, 2006. Výzkumná zpráva.

...

KOCH, Tom. *Journalism for the 21st century : online information, electronic databases, and the news*. New York : Greenwood Press, 1991. 374 s. Adamantine studies on the 21st century . ISBN 0744900328.

...

KOPTA, Martin. Kdo nesyndikuje, ať není. *Lupa : server o českém internetu* [online]. 2003, č. 12 [cit. 2006-12-07]. Dostupný z WWW: <<http://www.lupa.cz/clanky/kdo-nesyndikuje-at-neni/>>. ISSN 1213-0702.

...

MOTZ, Albert. The Battle Over Blogs. *CMA : Canadian Marketing Association* [online]. 2006 [cit. 2006-12-01]. Dostupný z WWW: <<http://www.the-cma.org/council/blogsbattle.cfm>>.

...

Newton Information Technology. *Newton I.T. : Information Technology* [online]. Praha : c2006 [cit. 2006-12-02]. Dostupný z WWW: <<http://www.newtonit.cz/>>.

...

O'REILLY, Tim. *What Is Web 2.0 : Design Patterns and Business Models for the Next Generation of Software* [online]. c2006 [cit. 2006-12-11]. Dostupný z WWW: <<http://www.oreilly.com/contact.html>>.

...

PAPÍK, Richard. Vyhledávání informací II. : Uživatelské rozhraní a vlivy oboru "human-computer interaction". *Národní knihovna : knihovnická revue*. 2001, č. 2, s. 81-90. Dostupný také z WWW: <<http://knihovna.nkp.cz/NKKR0102/0102081.html>>.

...

PAPÍK, Richard.. Společnost orientovaná na znalosti a digitální ekonomiku. *Ikaros* [online]. 2001, roč. 5, č. 1 [cit. 2006-12-16]. Dostupný na World Wide Web: <<http://www.ikaros.cz/node/691>>. URN-NBN:cz-ik691. ISSN 1212-5075.

...

PUMERA, Martin. Portály do kyberprostoru. *PC World.cz : magazín digitálního věku* [online]. 1999, roč. 3 [cit. 2006-11-11]. Dostupný z WWW: <<http://www.pcworld.cz/pcw.nsf/472b9f1e744c0ca0c12568c8004f4248/0a1096ca43431710c12568bb006286e8?OpenDocument>>.

...

SEFTON-WILLIAMS, Peter. *Re-Inventing Aggregation : The Rise of Aggregated Online Current Awareness in the Enterprise: Analysis and Recommendations Whitepaper*. London, 2005. Výzkumná zpráva. Electronic Publishing Services Ltd. Dostupný z WWW: <<http://www.verisign.com.au/guide/realtime-publish/aggregation.pdf>>.

...

Služba vystřihuje za nás. *Pražské noviny*. 18.2.1997, č. 2, s. 1.



...

SULLIVAN, Shanon. Factiva achieves #1 market position according to analyst firm. *PRWeb : Press release Newswire* [online]. 2006 [cit. 2006-12-01]. Dostupný z WWW: <<http://www.prwebdirect.com/releases/2006/8/prweb427960.htm>>.

...

ŠMÍD, Milan. Žurnalistika v informační společnosti : internetizace a digitalizace. *Pražské sociální vědní studie : Mediální řada MED-003* [online]. 2006 [cit. 2006-11-11]. Dostupný z WWW: <[http://www.fsv.cuni.cz/FSV-533-version1-MED003\\_Smid.pdf](http://www.fsv.cuni.cz/FSV-533-version1-MED003_Smid.pdf)>.

...

*The State of the News Media 2006 : an Annual Report on American Journalism* [online]. Washington : The Project for Excellence in Journalism, c2006 [cit. 2006-11-11]. Dostupný z WWW: <<http://www.stateofthenewsmedia.org/2006/>>.

...

Thomson Business Intelligence. *Thomson* [online]. c2006 [cit. 2006-12-01]. Dostupný z WWW: <<http://www.thomsonbusinessintelligence.com>>.

...

WARD, Mike. *Journalism online*. Oxford : Focal Press, 2002. 214 s. ISBN 0240516109.