

POROVNÁNÍ TUZEMSKÝCH BÁZÍ DAT SLEDUJÍCÍCH ČLÁNKY Z ČESKÝCH PERIODIK

Aleš Brožek

Poskytování údajů z novin a časopisů patří k běžné službě nabízené ve veřejných knihovnách. Vždyť archivace novin a časopisů, které vedle knih stále patří mezi nejvíce zastoupené druhy informačních materiálů ve veřejných knihovnách, je jednou z hlavních knihovních činností.

Práce s novinami a časopisy má v našich knihovnách dlouholetou tradici. Pro rychlejší a snazší vyhledávání potřebných článků vydávala Státní knihovna ČSR a ČSSR desítky let bibliografický katalog Články v českých časopisech. Navíc si mnozí aktivní knihovníci budovali po dlouhá léta výstřížkové archivy. V případě bibliografického katalogu si však oprávněně stěžovali na zpoždění ve vydávání Článků v českých časopisech, které se postupem doby zvětšovalo, i na jeho poměrně omezený rejstříkový aparát. V případě výstřížkových archivů se museli potýkat vedle jeho značné pracnosti při budování a udržování i s problémy klasifikace a třídění výstřížků, aby dokázali nalézt ty požadované.

Zavádění výpočetní techniky do knihoven nabízí knihovníkům novou cestu, jak výše uvedené problémy řešit. Je to budování a využíváníází dat, které sledují články z českých periodik. Ukazuje se, že se tato cesta zalíbila mnoha českým knihovníkům a informačním pracovníkům. Poslední vydání Přehleduází dat využitelných v knihovnách, které vyšlo na konci roku 1995 na disketě s hypertextovým vyhledáváním a je přístupné dokonce i na Internetu, zachytilo více než padesát takovýchází. Znamená to, že zhruba 30 procentází budovaných v knihovnách nebo pro knihovníky a informační pracovníky je takového typu. Jejich úroveň je však značně rozdílná. Jejich budování je náročné jak na čas, tak na schopnosti zpracovatelů a jejich počet. V zahraničních knihovnách se proto ustupuje od tvorby vlastníchází a místo toho se kupují profesionálně vytvářené báze dat.

Předpokládám, že i v našich knihovnách budeme brzy svědky tohoto trendu, a to i přesto, že mnozí zastánci budování vlastníchází obhajují své báze dost pádnými argumenty. Zdůrazňují místní specifika a lepší informovanost pracovníka v knihovně, pokud časopis excerpuje, než když „jen“ pracuje sází dat. Musíme si však uvědomit, že i ten nejvýkonnější pracovník stačí sledovat omezený počet titulů, protože jeho pracovní náplň musí zahrnovat

i styk s uživatelem, a že je v knihovnách poměrně značná fluktuace takových pracovníků. Je o ně totiž zájem i mimo knihovnickou oblast.

Úkolem této přednášky je seznámit přítomné se třemi profesionálně budovanými bázemi dat, jejichž základní charakteristiku podává tabulka 1. Cílem přitom není stanovit jejich pořadí podle kvality pomocí jejich porovnání. Není to ani možné, protože každá z nich je jiného typu. Jedná se o bibliografickou, faktografickou i plnotextovou bázi dat, přičemž každá je z určitého hlediska výhodnější než ostatní.

Tab. 1

Základní údaje o tuzemských bázích dat sledujících česká periodika

Název báze	Česká nár. bibliografie - články	Monitor tisku	Abacus Monitoring
Producent	Národní knihovna ČR + Albertina Icome	Albertina data	Abacus 2000
Typ	bibliografická	faktografická	plnotextová
Počet položek pro vyhledávání	dvacet	jedenáct	sedm (u rejstříkové konv.)
Retrospektiva	1991-červen 1996	1.1.1993 - červen 96	1.6.1995 - 30.6.1996
Počet sledovaných novin a časopisů	více než 500	130	120
Počet záznamů	335 632	147 308	106 664
Měsíční přírůstek	5 000 - 6 000	6 000	6 000

První z nich, „Českou národní bibliografii – články“, knihovníci znají nejlépe. Je to dáno jednak tím, že je to do určité míry elektronická forma dobře známých Článků v českých časopisech a že jejími producenty jsou knihovníci (s výjimkou odborníků z Albertiny icome, kteří vymysleli vyhledávací software). Na rozdíl od svého tištěného předchůdce však nevzniká jen prací pilných knihovnic z Národní knihovny ČR (i když všechny záznamy z roku 1991 a většina novějších záznamů má „křestní list“ napsaný v Národní knihovně ČR), ale od roku 1992 do báze přispívají knihovníci státních vědeckých knihoven a Moravské zemské knihovny. V novější době se k nim přidali také pracovníci STK, ÚZPI, ÚIV a dvou ústavů Akademie věd.

Vzhledem k nejdelší retrospektivě „České národní bibliografie – články“ nepřekvapí největší počet záznamů v porovnání se zbylými dvěma bázemi. Obdivuhodná je spíše velikost excerpční základny, která je více než čtyřnásobkem excerpční základny ostatních bází. Je potěšitelné, že excerpční základna reaguje na nově vzniklé tituly (v bázi najdeme záznamy i z nového časopisu Xantypa), že se báze neomezuje na humanitní a společenskovedné časopisy

a že začíná sledovat i technické časopisy (v poslední době např. AutoTip a Computerworld), takže nabývá charakteru univerzálnosti. Uvědomíme-li si však, že měsíční přírůstek „České národní bibliografie – články“ dosahuje zhruba stejného čísla jako u dvou dalšíchází, znamená to, že zařazování údajů o člancích do této báze je mnohem výběrovější. To ostatně dokládá i tabulka 2, která sledovala, kolik článků z týdeníku Českomoravský Profit splnilo kritéria pro zařazení do jednotlivýchází.

Tab. 2

Sledovanost Českomoravského Profitu č. 18 až 26 z roku 1996

	ČNB-články	Monitor tisku	Abacus	EPIS
aktualizace báze	červen 1996	6/1996	6/1996	červenec 96
došlo do knihovny	29.7.1996	26.7.1996	18.7.1996	17.7.1996
počet zachycených článků v:				
č. 18 z 30.4.96	7	42	25	67
č. 19 ze 7.5.96	4	29	26	49
č. 20 z 14.5.96	8	37	22	64
č. 21 z 21.5.96	4	34	33	59
č. 22 z 28.5.96	7	0	25	59
č. 23 ze 4.6.96	4	39	26	61
č. 24 z 11.6.96	7	0	23	68
č. 25 z 18.6.96	5	42	27	64
č. 26 z 25.6.96	nezprac.	40	32	61
celkem	46	263	239	552

Pozn.:

EPIS je zkratka pro Ekonomicko Právní Informační Servis produkovaný firmou K&Psoft. Jedna z jehoází (Profit - Anotace) je zaměřena na sledování novin Českomoravský profit

Pohledem do tabulky 1 zjistíme, že „Česká národní bibliografie – články“ má prioritu i v počtu položek pro vyhledávání. Například v porovnání s plnotextovým „Abacus Monitoring“ lze vyhledávat téměř v trojnásobku položek. Při jejich pečlivějším studiu však musíme připustit, že méně někdy znamená více. Vyhledávání podle druhu článku, svazku nebo čísla sešitu, myslím, nepatří mezi to nezbytné. Další položky jako autor, název článku, název zdrojového dokumentu a datum vydání jsou vyhledatelná ve všech

třech bázích. Charakteristické pro „Českou národní bibliografii – články“ je možnost hledat pomocí Mezinárodního desetinného třídění (kdo z neknižovníků ho však zná a využívá?!) a pomocí regionálního třídění. Posledně jmenované položce mohou jen zazlívát, že je to písmenný kód, který by měl být pro potřeby uživatelů báze rozkódován. Možnost vyhledávat pomocí klíčových slov je nutnost, ostatní dvě báze totiž umožňují hledat jakékoliv slovo v anotaci, respektive abstraktu, což je v případě „Abacus Monitoring“ plný text článku. Nepříjemné je, když zpracovatel šetřil v „České národní bibliografii – články“ klíčovými slovy, což naštěstí je výjimka. V průměru lze v „České národní bibliografii – články“ najít minimálně pět klíčových slov (obr. 1 a). Naproti tomu nedostatkem je absence tezauru pro klíčová slova či alespoň seznamu klíčových slov. Při vyhledávání tak musíme počítat i se synonymy a pravopisnými dubletami.

Obr. 1 a) Záznam z České národní bibliografie

Autor

HÁBÍK Jan

Název

Ostře sledované vlaky

Druh článku

rp

Název zdrojového dokumentu

ČESKOMORAVSKÝ PROFIT

Svazek

7

Číslo sešitu

18

Datum vydání

30.04.1996

MDT

656.1/.5

656.2

338.2

Klíčová slova

pozemní doprava

železniční doprava

železniční doprava

hospodářská politika

restrukturalizace drah

modernizace

Česká republika

r. 1996

doprava

ekonomika

ČR

Osoba a instituce, o níž článek pojednává
České dráhy
Regionální třídění
cz
Identifikační číslo záznamu
AN96ABO16 00055

Báze „Monitor tisku“ je faktografická, což nutí zpracovatele uvést v abstraktu konkrétní údaje, číselné hodnoty a další náležitosti, aby vystihl to nejdůležitější z článku a uspokojil uživatele do té míry, že nebude muset číst celý článek. Jak se to podařilo v případě článku o Českých drahách, můžete posoudit porovnáním záznamu z „Monitoru tisku“ (obr. 1 b) a z „Abacus Monitoring“ (obr. 1 c), kde je text nezkrácený a v originále zabírá 180 řádků ve sloupcích širokých 45 mm.

Obr. 1 b) Záznam z Monitoru tisku

Albertina - Monitor tisku
Albertina data, s.r.o.
Josef Fric SOFTWARE

Záznam : 1/1
Číslo záznamu: 135193
Název článku : Ostře sledované vlaky
Anotace : Konsolidace a restrukturalizace české železniční dopravy
Autor : Jan Hábík
Datum zařazení : 10.05.1996
Firma(y) :
České dráhy, a.s.
Zdroj(e) :
Českomoravský Profit, 30.4.1996, č.:18, str.:3
Obor(y) :
Pozemní doprava; potrubní přeprava
Tem. okruh(y) :
ZPRÁVY Z PODNIKŮ
Reorganizace
restrukturalizace
změna názvu
Privatizace
investice
FNM

Abstrakt :

Česká republika má 9500 kilometrů železničních tratí a má tak na svoji rozlohu a počet obyvatel nejhustší železniční síť na světě. Udržování 6200 kilometrů celostátních a 3300 kilometrů regionálních tratí je nákladné a mezi regionálními tratěmi je 13,7 % (1300 km) ztrátových. Projekt restrukturalizace českých drah počítá s redukcí nejztrátovějších tratí a celá redukce regionálních tratí může dosáhnout až 15 % celé sítě. Mezi jednotlivé kroky transformace, které společnost České dráhy plánuje, patří: zjištění nákladů a výnosů na všech tratích (do května 1996), pokus o snížení nákladů racionalizací provozu a údržby, zavedením úspornější technologie (do prosince 1996), volba tratí nabízených k privatizaci a převedení osobní dopravy na autobusy a mikrobuses (v květnu 1997), zrušení provozu na trati v případě, že se nenajde zájemce o privatizaci (do konce roku 1997). Celý proces proběhne do konce roku 1999 nejméně u 15 % tratí. Majetek bude možno získat za účetní hodnotu, tratě a kolejová vozidla se budou prodávat i pod vyvolávací cenu. Očekávaný přínos má Českým drahám vynést nejméně 1 mld. Kč.

V květnovém jízdním řádu počítají České dráhy s redukcí počtu osobních vozů asi o 5 % a počet vlaků v regionální dopravě se sníží o 1,2 %. Přesto bude stále o 400 % převažovat nabídka nad poptávkou. Dalším snížením (v září 1996) by se měla dostat železnice na úroveň roku 1993. Teprve v roce 1998 by mělo být dosaženo poměru mezi nabídkou a poptávkou pod 300 %. České dráhy začnou prodávat svůj zbytný majetek již letos, a to asi v hodnotě 1 mld. Kč (pozemky, budovy, byty, stroje a zařízení). Podnikatelům se tak nabízejí možnosti, které se už nebudou opakovat.

Obr. 1 c) Záznam z Abacus Monitoring (zkrácený, jedná se o plnotextovou databázi):

Ostře sledované vlaky

Datum : 02.05.1996

Anotace : Část majetku Českých drah bude již letos na prodej

Autor : Jan Hábík

Zdroj : ČESKOMORAVSKÝ PROFIT z 30.4.1996 č. 18 s. 3

Firma : ČESKÉ DRÁHY (ČD)

Na našich peronech necítíte vůni dále a kouzlo cestování, ale chlad a nejistotu. I v poledne má na nádraží člověk dojem, jako by mu odjel poslední flamendrák. Ale zdání - chválabohu - klame. Vlaky a koleje se opět dostávají do záře reflektorů. Nerozsvítil je filmový režisér, ale vláda, které není lhostejné, že

.....
.....
Přemlouvat vás nemohu, chci vás jen upozornit, že dráhy začnou prodávat svůj zbytný majetek. Letos asi v hodnotě jedné miliardy korun za pozemky, budovy, byty, stroje a zařízení, informoval nás Vladimír Budinský, ministr dopravy ČR. Je pochopitelné, že čtvereční metr nádraží v Praze - Bubnech se bude prodávat za jinou cenu než v Poličce. Přesto se tu každému nabízí možnosti, které se již nebudou opakovat (viz i graf Prodej zbytného majetku).

Zůstanete tedy sedět doma, nebo vás ostře sledované vlaky vylákají z poklidu?

Graf: Průměrná délka tratě

Graf: Prodej zbytného majetku

Graf: Nabídka a poptávka v osobní železniční dopravě

Pro vyhledávání lze v případě „Monitoru tisku“ využít i položek číslo záznamu a datum zařazení, bez nichž bych se dokázal obejít. Je sice zajímavé sledovat datum zveřejnění versus datum zařazení, které se pohybuje v rozmezí jen několika dnů, ale nedovedu si představit, jak často budu hledat podle data zařazení. (To má snad význam jen pro ty, kteří pečlivě sledují v bázi „Monitor tisku“, kdy co nového přibylo.) Tabulka 2 prozradila na „Monitor tisku“ určitou nespolehlivost, když některá čísla Českomoravského Profitu nebyla (omylem?) zpracována. Pro vyhledávání mají větší význam položky firma (tato položka prakticky odpovídá položce „osoba a instituce, o níž článek pojednává“ v „České národní bibliografii – články“), obor a tematický okruh. Zatímco pro položku obor existuje seznam, pro tematický okruh nikoliv, takže má kritická slova zaměřená na absenci tezauru pro klíčová slova v bázi „Česká národní bibliografie – články“ platí i pro tematický okruh v bázi „Monitor tisku“.

Báze „Abacus Monitoring“ sleduje nejmenší počet novin a časopisů ze všech tříází a má i nejmenší počet záznamů, protože je také nejmladší. Její záznamy jsou však plnými texty článků, což znamená, že uživatel prakticky nemusí jít do originálního zdroje. (V případě článku o Českých drahách však možná uživatel požádá knihovníka o donesení 18. čísla Českomoravského Profitu, protože si bude chtít prohlédnout některý z grafů, na jejichž existenci báze „Abacus Monitoring“ pouze upozorňuje. Tato báze má nejmenší počet položek sloužících pro vyhledávání. Při použití rejstříkové konvence je to název článku, autor, zdroj, firma, datum vydání, anotace a abstrakt. Je sice pravda, že díky plnému textu uživatel najde článek o Českých drahách hledáním sousloví „vůni dalek“ nebo „kouzlo cestování“ či slova „flamendrák“, pokud ho vůbec napadne, že autor článku právě tato „poetická“ slova a právě v takovém pádu použil. (Pro zajímavost lze uvést, že výše uvedená sousloví byla jednoznačná a našla se opravdu jen v jediném článku ze sta tisíce v bázi!) Bázi „Abacus Monitoring“ by však v tomto případě prospělo zavedení další položky, která by obsahovala i výrazy nevyskytující se v plném textu a přitom vhodné pro vyhledávání. Vzorem pro tuto položku by se mohla stát položka obor z báze „Monitor tisku“ nebo Mezinárodní desetinné třídění či klíčová slova z báze „Česká národní bibliografie – články“. Inspirujeme-li se totiž heslem „pozemní doprava“ z báze „Monitor tisku“, článek o Českých drahách v bázi „Abacus Monitoring“ nenajdeme, protože se v textu článku nevyskytuje.

Pro úplnost dodávám, že počet článků z Českomoravského Profitu, které byly „zařaditelné“ doází (každý časopis vždy obsahuje i texty, které se snad do časopisu hodí, ale je zbytečné je zařazovat doází) byl stanoven v tabulce

2 s využitím báze EPIS, která se specializuje na sledování Českomoravského Profitu. Poslední ukázka (obr. 1 d) je právě ze zmíněné báze a demonstruje, jak je zde článek o Českých drahách zpracován. Takové báze zaměřené na jeden titul novin či časopisu jsou vedle tištěných obsahů jednotlivých čísel další pomůckou, hledáme-li článek pro uživatele a přitom víme, v kterém časopise byl otištěn. Selže-li i taková pomůcka, nezbyvá než si vzít několik posledních čísel, půlročník i celý ročník časopisu do ruky a listovat a listovat.

Obr. 1 d) Záznam z EPISu

Profit Anotace 1996	
Název:	Ostře sledované vlaky
Autor:	Jan Hábík
Anotace:	Drama odehrávající se na českých železnicích v souvislosti s privatizací majetku Českých drah. Názor ministra dopravy ČR Vladimíra Budinského.
Klíč. slova:	Nabídka a poptávka v osobní železniční dopravě
Délka článku:	střední
Lokalizace:	Profit ročník: 1996, číslo: 18, strana: 3