

S nejbohatší databází

Soubor překladových slovníků Millennium ve verzi 4 se na trhu ještě ani pořádně neohřál, a už je tu následník a má se čile k světu. Recenzi “čtyřky” jsme přinesli v Chipu 4/00, takže se nyní soustředíme především na odlišnosti novinky oproti této verzi. Leccos si budete moci také sami ověřit – Millennium jste v upravené podobě dostali jako přílohu tohoto Chipu.

Instalace

Aplikace byla nainstalována pod Windows 95, NT 4.0 a 2000. V prvním případě proběhla instalace zcela bez problémů, pouze po zadání odblokovacího klíče program hlásí, že odblokování vezme na vědomí po restartu aplikace, otevření úplné databáze však nastane až po restartu celého systému. Pod NT zas okno zadávající uživatelskou instalaci hlásí nedostatečnou velikost volného prostoru na HD, i když ve skutečnosti bylo (podle Exploreru) volných více než 3,5 GB; hláška ale nijak nebrání instalaci. (Pravděpodobná příčina: HD má kapacitu 25 GB, partition je 5,85 GB.) Odblokování databáze zde nastane bez problému okamžitě po restartu aplikace. Žádná ze závad se už neprojevila pod Windows 2000.

Diskový prostor potřebný pro jednotlivé databáze není zanedbatelný: angličtina 304,7 MB, němčina 213,3 MB a francouzština 104,4 MB. Obslužný program je relativně skromný, pouhých 1,4 MB. Na systémovém disku se dále instaluje Microsoft DAO 3.5, jehož soubory mají 2,4 MB (je-li v systému MS Office 97, je již nainstalován). Instalace zvukové podpory Microsoft Agent zabere v plné verzi 22,8 MB.

Oproti minulé verzi se objem databází rozrostl natolik, že kvůli komprimaci odpadla možnost otevírat databáze z CD a na pevném disku mít pouze uživatelské rozhraní. CD je univerzální a obsahuje všechny databáze, které se nainstalují výběrově pouze podle jazyků. Po zaplacení odblokovacích klíčů pro vybrané moduly je ovšem možné obsazený prostor disku zmenšit odstraněním nevyužívaných modulů.

Uživatelské rozhraní

Uživatelské rozhraní bylo rozšířeno u funkce zkracující zadaný řetězec. Dříve se odebírala písmena od konce řetězce, nyní lze směr odebírání volit: buď od konce (implicitní nastavení), nebo od začátku (viz obr. 3) Tento režim je vhodný zejména pro němčinu, kde se tak v případě složeného slova spíše dostaneme ke kmeni.

Nové jsou také tři funkce pro ukládání do speciálních TXT souborů. Jednak se automaticky ukládá Historie, která dříve podchycovala jen posledních 20 hledání. Dále lze do pojmenovaného TXT souboru ukládat vybrané termíny (obr. 4). Tímto způsobem lze např. vytvořit glosář a po transformacích jej importovat přímo do nástrojů systémů CAT; kde se pak všem členům překladatelského týmu kontextově nabízí při překladu (u systému Trados po převodu na tabulku Wordu, v případě DéjVue na tabulku Excelu). Tím je zajištěna dokonalá terminologická homogenita rozsáhlého překladového projektu.

Dalším pojmenovaným souborem typu TXT je soubor zachycující zjištěné chyby v databázi. Při jejím rozsahu je občasný výskyt překlepu přirozený – každý uživatel, který na něj narazí, má nyní jednoduchou možnost, jak se zasloužit o jeho opravu. Když po určité době tento soubor odešle autorské firmě, může se těšit, že od příští verze již bude chyba odstraněna.

Databáze

Při posuzování databází, které pod souhrnným záhlavím obsahují mnoho dílčích modulů, bylo nutné každou z nich porovnávat s řadou slovníků. Z každého dále uvedeného slovníku bylo vzato vždy 50 náhodně vybraných hesel a bylo zjišťováno, kolik z nich se v analyzované databázi nachází. Při velikosti srovnávaných databází jsou takovéto údaje pochopitelně jen informativní; teprve při velikosti srovnávacího vzorku alespoň o dva řády vyšší by byla věrohodnost výsledku uspokojivá. I při tomto omezení lze však prohlásit, že výsledky přesahující 80 % znamenají velmi dobrou oborovou shodu.

Výsledky porovnání jsou dále uváděny v procentech, objemy jednotlivých oborů v počtu dvojic “jazyk výchozí – cílový”, nikoli v heslech; totéž platí o rozšířeních oproti verzi 4. Všechny ceny jsou

bez DPH.

Angličtina

Obor obchodní/právní

Elman: A-Č ekonomický slovník, 1995 – shoda nad 95 %; Chromá: A-Č právní slovník, 1995 – shoda nad 95 %; Mueller-Lutz: N-A-F-Č pojišťovací slovník, 1991 (níže jen "Lutz") – shoda nad 95 %; Hynek/Vítovský: A-Č a Č-A slovník výpočetní techniky a IT, FRAUS 2000 (porovnávání při současně zapnutém technickém oboru) – nalezeno cca 55 %; Říha: Internet A-Č výkladový slovník – nalezeno 45 %; Dyson: Novellovský slovník sítí, 1995 – nalezeno cca 50 %; Hoenigová et al.: A-Č terminologie bezpečnosti IT, 1996 – nalezeno cca 40 %; A-Č hospodářský slovník, FRAUS 2000 – shoda nad 80 %; Le Docte: A-Č-F-N právní slovník (níže jen "Le Docte"), 1997 – nalezeno nad 65 %. Rozšířen o 22 000, celkem 152 000.

Obor technický

A-Č technický SNTL – shoda nad 95 %; A-Č elektrotechnický SNTL – shoda nad 95 %; Procházka: Slovník pro výstavbu A-Č-N-F, 1992 (níže jen "Procházka") – shoda nad 90 %; Dvořák et al.: Č-A-N slovník automobilové techniky, 1995, (níže jen "Dvořák") – shoda cca 85 %. Rozšířen o 22 000, celkem 689 000.

Obor lékařský

Paroubková et al.: A-Č lékařský slovník, 1996 – shoda nad 95 %. Celkem 152 000.

Obor chemický

Jouklová et al.: A-Č a Č-A chemicko-technologický slovník, SNTL – nalezeno 35 %. Celkem 152 000.

Obor speciální

Král et al.: A-N-Č-R slovník veterinárního lékařství, 1990 (níže jen "Král") – při porovnávání pouze s tímto oborem shoda malá, překvapivě více veterinárních termínů je v oboru technickém, při zapnutých všech oborech nenalezeny ani některé běžné veterinární výrazy (jalovost, mršina); Horák: A-Č a Č-A lesnický a myslivecký slovník, 1991 – shoda nad 95 %; Horák: A-Č a Č-A slovník lesnický a dřevařský, 1999 – porovnávána hlavně dřevařská terminologie – shoda nad 95 %; Peprník: Slovník amerikanizmů – shoda nad 95 %. Rozšířen o 113 000, celkem 185 000.

Obor vojenský

Rozšířen o 10 000, celkem 57 000.

Anglická obchodní korespondence

Uživatel zaplativší obor obchodní/právní dostává tuto databázi o 5500 položkách zdarma. Samostatně stojí 1000 Kč.

Němčina

Obor obchodní/právní

Lutz – shoda nad 95 %; N-Č/Č-N hospodářský slovník, FRAUS 1997 – shoda cca 90 %; Le Docte – nalezeno nad 60 %. Rozšířen o 121 000, celkem 231 000.

Obor technický

N-Č technický slovník SNTL – shoda nad 95 %; Michalíček: N-Č a Č-N hornický slovník – shoda nad 65 %; Procházka – shoda cca 90 %; Dvořák – shoda nad 95 %. Rozšířen o 102 000, celkem 448 000.

Obor lékařský

Celkem 60 000.

Obor chemický

Gottwald et al.: N-Č a Č-N chemicko-technologický slovník, SNTL – nalezeno nad 90 %. Celkem 160 000.

Obor speciální

Král – výsledek porovnání velmi obdobný jako v angličtině, některé výrazy spíše v technickém nebo lékařském oboru. Rozšířen o 21 000, celkem 121 000.

Obor vojenský

Celkem 57 000.

Ceny oborů pro angličtinu/němčinu

Všeobecný	zdarma
Obchodní/právní	2 600 Kč
Technický	5300 Kč
Lékařský	2000 Kč
Chemický	2300 Kč
Speciální	3500 Kč

Vojenský 3000 Kč
Multioborový 10 800 Kč (všechny výše uvedené obory s výjimkou vojenského)
K ceně za jednotlivé obory se ještě přičítá 300 Kč za distribuční CD (stejný pro všechny slovníky).

Francouzština

Francouzština nabízí v současnosti pouze obory obchodní/právní a technický společně, celkem 258 000 dvojic, bez možnosti oddělené platby. Připojen je všeobecný slovník o rozsahu 152 000 dvojic.

Porovnání bez rozdělení na obory: Neumann/Hořejší et al: Velký F-Č slovník, ČSAV 1992 – nalezeno cca 60 %; Lutz – nalezeno cca 35 %; Le Docte – nalezeno nad 55 %; Dubský et al: F-Č a Č-F obchodní slovník, 1998 – nalezeno cca 45 %;

Slovník nepodchycuje řadu běžných zkratk, např. ONU, ONUDI, OIT, OECE.

Cena 10 800 Kč.

Poznámky a připomínky

(Není-li řečeno jinak, následující poznámky se týkají anglických databází.)

Spotřeba vnitřní paměti (NT/2000)

Bez připojené databáze pouhé 4 MB, po připojení databáze a následném hledání postupně vystoupá nad 20 MB a tam už setrvá po celou dobu používání. Množství zabrané paměti nad základních cca 12 MB je úměrné počtu nalezených výskytů. I když později jsou hledána paměťově skromnější slova, program paměť neuvolní. (Vzhledem k tomu, že Word si nárokuje přes 16 MB RAM, nechceme-li se dočkat zdržování swapováním na HD, je dobré mít nad operačním systémem alespoň 40 MB volných.)

Měření časů

Časy byly měřeny jako rozdíly mezi dvěma následujícími údaji spotřebovaného času CPU uváděného Správcem úloh (Celeron 500 MHz, 128 MB RAM) – jde o extrémní případy s tisíci výskytů. U běžných dotazů (pro desítky výskytů) je čas hledání neměřitelný. Vyhledávací algoritmus rozděljuje odpovědi na dotaz do dvou kategorií: řetězce začínající řetězcem zadaným (údaj před lomítkem) a řetězce obsahující zadaný řetězec kdekoliv – do této kategorie se při běžném dotazu zařazuje jen první slovo ze zadaného řetězce (údaj za lomítkem).

Dotaz	Poč. výskytů	Čas [s]
stroj	2932/4899	15
papír	458/1486	5
účet	401/524	2

Pokud chceme hledat víceslovní spojení vyskytující se kdekoliv v řetězci, je nutné přestavit režim vyhledávání. Program zcela oprávněně varuje, že hledání může trvat dlouho – prohledává se totiž plnotextově celá databáze. Dotaz "stroj na" se vyhledával 26 sekund a nabídl 3372 výskytů. U počítačů s pomalejší CPU může tento režim hledání zabrat i několik minut.

Nekonzistence v databázi

Databáze obchodní/právní (O/P) obsahuje heslo "najatý zabiják (vrah)", zatímco databáze speciální má "najatý zabiják n. vrah". Tento úkaz se objevuje vícekrát. Interní program pro vyřazování duplicit po přidávání nových segmentů zřejmě porovnává pouze na absolutní shodu; vhodnější by ovšem byl "fuzzy" přístup.

Chyba zobrazení českého písma

Vyskytuje se pod Windows 95 (obr. 5); pod NT či 2000 není.

Jazyk komunikace s uživatelem

I když byla připojena francouzská databáze, možným komunikačním jazykem zůstala pouze čeština, němčina nebo angličtina.

Počet překladů

Databáze je v současnosti natolik rozsáhlá, že už by zasloužila podrobnější dělení oborů podle specializací. Například při zapnutí pouhých dvou oborů (obchodní a technický) na dotaz offset dostaneme 104 možných překladů. S takto velkou paletou možností má i profesionální překladatel co dělat. Technický obor by si nejspíš zasloužil i hierarchické členění, např. elektrotechnika dělená na silnoproud, slaboproud a elektroniku.

Uživatelé francouzštiny

jsou na tom v tuto chvíli nejhůře, neboť nemají možnost odděleného nákupu jednotlivých oborů a platí za multioborový slovník o rozsahu 410 000 dvojic stejně jako uživatelé anglické databáze,

která má 1 218 000 dvojic.

Informační technologie

Terminologie z této oblasti si zaslouží rozšíření. Některé výrazy, s nimiž se běžně setkáváme v odborných publikacích již několik let, ve slovníku nejsou: outsourcing, firewall, disk mirroring, hop count, handler.

Závěr

V současnosti nejrozsáhlejší databáze od zavedení nakladatelství se dále rozrostla. Tematicky velmi dobře pokrývá většinu klasických technických aplikací, v průběhu posledních dvou verzí úspěšně expandovala i do oblasti přírodních věd. Stejně jako u ostatních hromadně prodávaných databází, překotný vývoj a s ním i terminologie v oboru výpočetní techniky není dostatečně rychle promítán do nabízené slovní zásoby – elektronický obchod, internet a bezdrátové telekomunikace jsou ve slovníku poněkud pozadu. Zůstávají také některé otazníky z minulosti a slovníku by už prospělo podrobnější rozřídění databáze na podobory a specializace. Vedoucí postavení Millenniuma co do rozsahu databáze dále trvá.

Miroslav Herold

Millennium5

Překladové slovníky (A, N a F) pro Windows 95 a vyšší.

Hardwarové nároky: výkonnější počítač pro Windows s dostatkem operační paměti a diskového prostoru (viz text).

Výrobce/poskytl: Commercial Service, Brno (www.pc-slovníky.cz)

Cena: multioborový 10 800 Kč, ceny jednotlivých oborů viz text.