

Obsah

1. úvod	2
1.1. Co lze očekávat od hledání v plných textech <i>Ottovy encyklopedie</i>	3
1.2. Co je na CD-ROM disku <i>Ottovy encyklopedie II.</i>	4
1.3. Co lze v <i>Ottově encyklopedii verze II.</i>	5
2. Instalace		
2.1. Několik rad před instalací	6
2.2. První instalace	8
2.3. Další (opětovná instalace)	9
2.4. Odinstalace <i>Ottovy encyklopedie II.</i>	9
3. Vyhledávání v plných textech		
3.1. Spuštění aplikace	10
3.2. Vyhledání hesla	11
3.3. Zobrazení výsledku hledání	12
3.4. Struktura hesel	13
3.5. Provázanost textů mezi sebou, provázanost s verzí I.	14
3.6. V případě problémů...	15
4. Doplnky		
4.1. Použití aplikace <i>Tovek Agent</i>	16
4.2. Pomůcka k tvorbě dotazů	20
5. Závěrem	24

Vážený zákazníku,

děkujeme Vám, že jste se rozhodl(a) zakoupit *Ottovu encyklopedii* obecných vědomostí na CD-ROM ve verzi II., respektive jejím prostřednictvím rozšířit možnosti verze I., kterou již vlastníte. Ve verzi II. můžete nyní hledat a nacházet informace a souvislosti v plných textech *Ottova slovníku* a výsledky Vašeho hledání pak dále libovolným způsobem použít a zpracovat.

Ottova encyklopedie obecných vědomostí na CD-ROM je projektem převodu *Ottova slovníku* naučného – největší české encyklopedie a pilíře českého vědění do moderního elektronického hávu - nyní na médium CD-ROM, v budoucnu pak na DVD a Internetu.

Projekt je rozdělen do tří po sobě jdoucích etap, přičemž každé etapě přísluší konkrétní (římsky číslovaná) verze *Ottovy encyklopedie* obecných vědomostí.

Verze I. je „digitální replikou“ velkého *Ottova slovníku* naučného ve 28 svazcích, poprvé knižně vydaného v letech 1888 – 1909. Tuto verzi, čtouce tyto řádky, znáte a používáte, nebo alespoň máte k dispozici.

Verze II. je rozšířením užitečných vlastností a možností verze I. *Ottovy encyklopedie* o vyhledávání v plných textech. Této verzi je věnován další text této příručky.

Verze III. doplní obě předchozí verze o digitalizovanou podobu *Ottova slovníku* naučného nové doby, který vyšel v letech 1930 – 1943 ve formě 12 svazků dodatků k “velkému” *Ottovu slovníku*.

Volitelným doplňkem jsou pak mapové přílohy, které jsou k dispozici na zvláštním CD-ROM disku.

Přečtěte si, prosím, tuto příručku co nejpečlivěji, jelikož informací, které musíte vstřebat proto, aby Vám *Ottova encyklopedie* ve verzi II. úspěšně sloužila je poněkud více než v případě verze I.

Přání mnoha pěkných chvil při čtení *Ottovy encyklopedie* z úvodu uživatelské příručky k verzi I. doplňujeme nyní o přání co největšího užitku z verze II....

1.1. Co lze očekávat od hledání v plných textech Ottovy encyklopedie

Ottova encyklopedie je novým "digitálním" kabátem Ottova slovníku naučného, největšího česky psaného encyklopedického díla v historii. Přes svůj věk (přibližně sto let) je Ottův slovník neustálým zdrojem poznání minulosti a s jistým (filosofickým) nadhledem i současnosti.

Ottův slovník naučný je dílem více než 1000 autorů, na přibližně 29 tisících stranách obsahuje přes 180000 hesel. Z pohledu dalšího zpracování tvoří obrovskou, unikátní databázi znalostí, tedy informací. Vzhledem k době svého vzniku a množství autorů - je to databáze jen velmi málo strukturovaná...

V koncepci verze I. jsme si dali za úkol, aby přístup k datům Ottovy encyklopedie byl velmi podobný práci s knihou. Kterékoliv heslo nalistujete čtyřmi kliknutími myši. Kromě textu hesla samotného pak máte k dispozici také odkazy na hesla související, pokud autoři souvislosti vyznačili. Například pokud hledáte informace o Karlu IV., najdete v hesláři heslo Karel IV., císař římský a král český.

Záměrem verze II. je pak rozšíření hledacích možností o hledání v plných textech. Kompletní texty Ottova slovníku jsou nyní rozebrány až na jednotlivá slova v obrovské indexované databázi, do které lze velmi rychle přistupovat. Napíšete dotaz »Karel IV.« a v několika vteřinách (dáno rychlostí počítače) pak zjistíte, že o Karlovi IV. je řeč v dalších 296 heslech(!). Navíc můžete do těchto hesel rychle nahlédnout a porovnat je mezi sebou v oddělených textových oknech. Přitom na jediné kliknutí myši máte stále k dispozici digitální repliku Ottova slovníku naučného, kde si můžeme heslo přečíst v plné kráse včetně všech obrázků, tabulek, vzorců matematických i chemických.

Pokud však chcete dobrou odpověď, musíte položit dobrou otázku. Pokud Vás zajímá například, jak Karel IV. ovlivnil Prahu, kdy v ní pobýval, zprávníte dotaz na »Karel IV, Praha« a zjistíte, že kombinace těchto slov se vyskytuje již "jen" ve 119 heslech. A tak dále...

1.2. Co je na CD-ROM disku Ottovy encyklopedie verze II.

Na CD-ROM disku *Ottovy encyklopedie verze II.* naleznete

- texty 28 svazků *Ottova slovníku naučného* z let 1888 – 1908 doplněné indexovými soubory, které umožňují hledání libovolných slov, nebo jejich skupin, či vzájemných souvislostí
 - se zachováním vzájemné provázanosti hesel
 - se zachováním hierarchie hesel
 - se zachováním fontového bohatství
 - s možností kdykoliv odskočit pouhým kliknutím myši na příslušné heslo do *Ottova slovníku naučného* (tedy do verze I. *Ottovy encyklopedie*) a mít tak k dispozici veškerou krásu tohoto velkolepého díla včetně všech obrázků, tabulek, matematických či chemických vzorců
- softwarový balík *Tovek Tools 2.05*, který Vám pro práci s plnými texty *Ottovy encyklopedie* zprostředkovává možnosti patrně nejvýkonějšího fulltextového systému *Topic* americké firmy *Verity* (po instalaci máte možnost nejen pracovat s plnými texty *Ottovy encyklopedie* a také si tento mocný systém vyzkoušet na vlastních datech ¹⁾)

¹ *Tovek Tools* jsou produktem české softwarové firmy *Tovek*. Pro účely vyhledávání v plných textech *Ottovy encyklopedie* byly realizovány některé úpravy tak, aby bylo možno použít volně šířitelnou demontsační verzi tohoto produktu..

1.3. V Ottově encyklopedii na CD-ROM verze II. lze

- prohledávat plné texty encyklopedie v prostředí **Windows 95/NT**² pomocí robustního vyhledávacího aparátu Topic[®] firmy Verity[™] přičemž
- hesla, která obsahují Vámi hledané slovo, či kombinace slov můžete zobrazit v samostatných oknech, v kterých budou hledaná slova zvýrazněna podsvícením
- tyto hesla si rovněž můžete jediným kliknutím myši zobrazit v Ottově encyklopedii verze I., čím se Vám otevře plný rozměr Ottova slovníku včetně všech obrázků, vzorečků a tabulek a různých méně používaných fontů
- texty Ottovy encyklopedie jsou Vám k dispozici pro další použití ve formátu RTF (Rich Text Format) pro další využití v jakémkoliv textovém editoru v prostředí Windows, případně prostřednictvím funkce "Kopíruj/Vlož" (Copy/Paste)
- verze II. Ottovy encyklopedie je publikována na samostatném CD-ROM disku, na kterém je též samostatně funkční; doporučujeme je však zkopírovat na pevný disk počítače a v CD-ROM mechanice mít vloženu verzi I. (podrobnosti v sekci Instalace)

² provoz pod Windows 3.1x je možný jen omezeně, vzhledem k tomu, že v aplikaci jsou z důvodu maximální funkčnosti využity některé vlastnosti 32-bitových operačních systémů, které již není možné do Windows 3.1x implementovat..

2. Instalace

2.1. Několik rad před instalací Ottovy encyklopedie

Jak již bylo řečeno *Ottova encyklopedie* ve verzi II. je tak koncipována takovým způsobem, aby bylo možné

- rychle nalézt heslo či hesla s žádanou informací či informacemi včetně zvýraznění všech souvislostí
- na jediný klik myši si prohlédnout vyhledané heslo v digitální replice *Ottova slovníku naučného* (tedy ve verzi I. *Ottovy encyklopedie*) v plné kráse, včetně všech obrázků, tabulek, či vzorečků.

Vzhledem k velkému objemu dat je *Ottova encyklopedie* distribuována na samostatném CD-ROM disku. Jako taková umožňuje provoz v zásadě na pět způsobů podle níže uvedené tabulky. Každý způsob je určitým kompromisem mezi rychlostí hledání, prostorem na disku, který věnujete *Ottově encyklopedii* a dostupností digitální repliky *Ottova slovníku*.

- **VYBERTE** si ještě před začátkem instalace *Ottovy encyklopedie* způsob, který nejlépe vyhoví Vám a Vašemu počítači. Z pohledu maximální funkčnosti Vám doporučujeme variantu 4, která je ovšem nejnáročnější na diskový prostor. Jakoukoliv změnu v konfiguraci *můžete* kdykoliv uskutečnit při novém spuštění instalačního programu.
- **POZOR!** Jakoukoliv změnu ve způsobu provozu *provádějte zásadně* prostřednictvím instalačního programu, jelikož každému ze způsobů provozu přísluší specifické záznamy v konfiguračních souborech.

Varianty instalace, konfigurace a provozu Ottovy encyklopedie obecných vědomostí verze II.

	způsob a popis provozu	výhody	nevýhody
1.	provoz pouze z CD-ROM disku - v CD mechanice je CD-ROM disk verze II.	- zabere minimální prostor na pevném disku (cca 9MB)	- při požadavku náhledu do digitální repliky nutno měnit CD-ROM v mechanice (to může způsobovat systémové problémy na ne zcela vyladěných systémech) - pomalé vyhledávání z důvodu vysoké přístupové doby CD-ROM mechanik
2.	kombinovaný provoz (z pevného disku i CD-ROM) - indexové soubory na pevném disku - v CD mechanice je CD-ROM disk verze II.	- rychlé vyhledávání i při velmi složitých dotazech	- zabere velký prostor na pevném disku (cca 190MB) - při požadavku náhledu do digitální repliky nutno měnit CD-ROM v mechanice (to může způsobovat systémové problémy na ne zcela vyladěných systémech)
3.	kombinovaný provoz (z pevného disku i CD-ROM) - indexové soubory na pevném disku - v CD mechanice je CD-ROM disk verze I.	- rychlé vyhledávání i při velmi složitých dotazech - okamžitý přístup do digitální repliky, tedy verze I. Ottovy encyklopedie	- zabere velký prostor na pevném disku (cca 190MB) - nepodsvěčuje vyhledaná slova ani souvislosti - texty Ottovy encyklopedie nejsou k dispozici pro další použití
4.	kombinovaný provoz (z pevného disku i CD-ROM) - indexové soubory na pevném disku - v CD mechanice je CD-ROM disk verze I.	- rychlé vyhledávání i při velmi složitých dotazech - okamžitý přístup do digitální repliky, tedy verze I. Ottovy encyklopedie	- zabere velký prostor na pevném disku (cca 400MB)
5.	provoz na 2 CD-ROM mechanikách (z pevného disku i CD-ROM) - v první mechanice je CD-ROM Ottova encyklopedie verze I. - v druhé mechanice je CD-ROM Ottovy encyklopedie verze II.	- okamžitý přístup do digitální repliky, tedy verze I. Ottovy encyklopedie - zabere minimální prostor na pevném disku (cca 9MB) - možnost používat Ottovu encyklopedii v CD-ROM serverech	- pomalé vyhledávání z důvodu vysoké přístupové doby CD-ROM mechanik - nutnost pořízení druhé CD-ROM mechaniky - BIOS některých výrobců nepodporuje 2 CD-ROM mechaniky, případně mohou nastat jiné hardwarové či softwarové problémy

2.2. První instalace

Windows 95 si automaticky detekují první vložení CD-ROM disku *Ottovy encyklopedie verze II.* a nabídnou možnost instalace, kterou můžete potvrdit a spustit či nechat na později.

Ve **Windows NT** spusíte, prosím, k instalaci program **INSTALL.EXE**, v adresáři **INSTALL.II** CD-ROMu *Ottovy encyklopedie II.*

Podmínkou pro instalaci *Ottovy encyklopedie obecných vědomostí verze II.* je předchozí úspěšná instalace verze I. Pokud nemáte verzi I. instalovánu, instalační program Vám tento fakt oznámí a ukončí činnost. Problém napravte řádnou instalací verze I. a poté vložte CD-ROM s verzí II.: resp. spusíte znovu **INSTALL.EXE**.

Instalační program provede instalaci následujících komponent

- i. Adobe Acrobat Reader 3.01
- ii. Tovek Tools 2.05
- iii. dat pro vyhledávání v plných textech, včetně jejich konfigurace a napojení
- iv. písma pro korektní zobrazení řečtiny a některých speciálních znaků

V případě první instalace provede program automaticky kroky i. - iv.

V kroku iii. Vám dá na výběr příslušnou variantu provozu pouze z CD-ROM disku či kombinovaně z pevného disku i CD-ROM. Rozhodněte se podle svých požadavků na rychlost vyhledávání a podle svých systémových možností (detaily v předchozí sekci 2.1.)

2.3. Další (opětovná) instalace

V případě opakované instalace, či reinstalace za účelem změny umístění, či konfigurace plnotextového vyhledávání spouštějte ve všech systémech (**Windows 95 i Windows NT**) program **INSTALL.EXE**.

Poté budete mít možnost instalovat pouze komponentu, s kterou jste měli problémy (Adobe Acrobat, Tovek Tools, písma), případně změnit variantu provozu *Ottovy encyklopedie*, či umístění datových souborů (podrobněji v části 2.1.).

Pokud nevíte, s kterou komponentou jste měli problémy, reinstalujte všechny.

24. Odinstalace *Ottovy encyklopedie*

Ottova encyklopedie není vybavena žádným speciálním programem na odinstalaci všech komponent. Samostatně je tedy nutno odinstalovat Adobe Acrobat (položka *Uninstall Acrobat Reader 3.01* ve skupině *Ottova encyklopedie*), samostatně Tovek Tools dle návodu v Nápovědě tohoto programového balíku. Datové soubory pro vyhledávání *Ottovy encyklopedie* pak vymažte pouhým vymazáním z disku

3. Vyhledávání v plných textech

3.1. Spuštění aplikace

Verze II. Ottovy encyklopedie (tedy vyhledávání v plných textech) je úzce navázána na Ottovu encyklopedii obecných vědomostí na CD-ROM a k úplné funkčnosti předpokládá její přítomnost, byť i výměnou CD-ROM disku.

Varianta 1., 2. (dle tabulky v sekci 2.1)

Ottovu encyklopedie spouštějte prostřednictvím aplikace Tovek Agent, kterou naleznete v programové skupině Tovek Tools. Na heslo Ottovy encyklopedie do digitální repliky máte možnost nahlédnout prostřednictvím volby "Otevřít dokument". Bude vyzváni k vložení CD-ROM disku Ottovy encyklopedie verze I. **Před návratem k aplikaci Tovek Agent uzavřete Ottovu encyklopedii verze I.** Aplikace (Tovek Agent) si totiž pamatuje, že v mechanice je CD-ROM verze II. a druhá aplikace (Adobe Acrobat) předpokládá vloženou verzi I., mohou nastat nečekané systémové problémy (zejména ve Windows 95).

Varianta 3., 4., 5. (dle tabulky v sekci 2.1)

Spuštění aplikace vyhledávání v plných textech se realizuje z prostředí Ottovy encyklopedie verze I. tedy z prostředí aplikace Acrobat Reader kliknutím tlačítka **F** v jeho horní liště dle následujícího obrázku, která spustí aplikaci **Tovek Agent** z instalovaného programového balíku **Tovek Tools**.



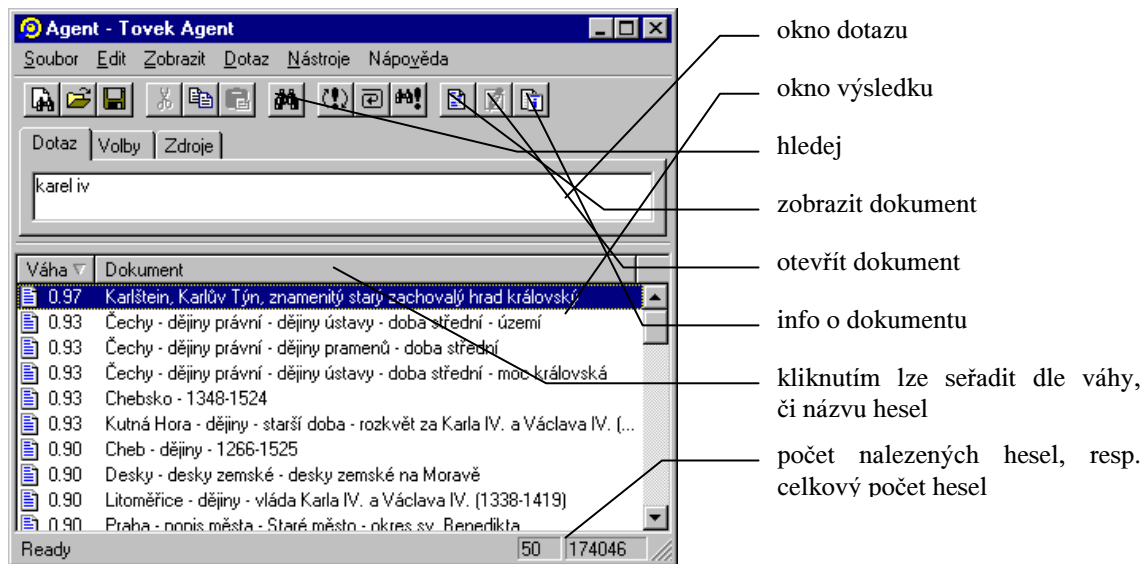
Tlačítko **O** v horní liště slouží k "zaostření" na cíl. Používejte jej v případě, že se chcete vrátit na začátek hesla, na které jste se dostali z Tovek Agentu při volbě "Otevřít dokument" a dále pokud Vás volba "Otevřít dokument" zavedla do již otevřeného souboru.

3.2. Vyhledání hesla

Základní prostředí programu Tovek Agent je znázorněno a stručně popsáno na následujícím obrázku.

Hledání probíhá jednoduše: do okna dotazu napíšete hledané slovo, či slova (bez interpukčních znamének), a stisknete Enter. V okně výsledků se Vám pak objeví seznam hesel, které hledané slovo obsahují. Hesla jsou seřazena podle váhy, která na základě relativního a absolutního počtu výskytu hledaných slov určuje, do jaké míry to, které heslo odpovídá Vašemu hledání.

Každé heslo si lze prohlédnout, prostřednictvím volby "Zobrazit dokument", odskočit na něj do digitální repliky volbou "Otevřít dokument" a zjistit výskyt hledaných slov pomocí volby "Info o dokumentu".

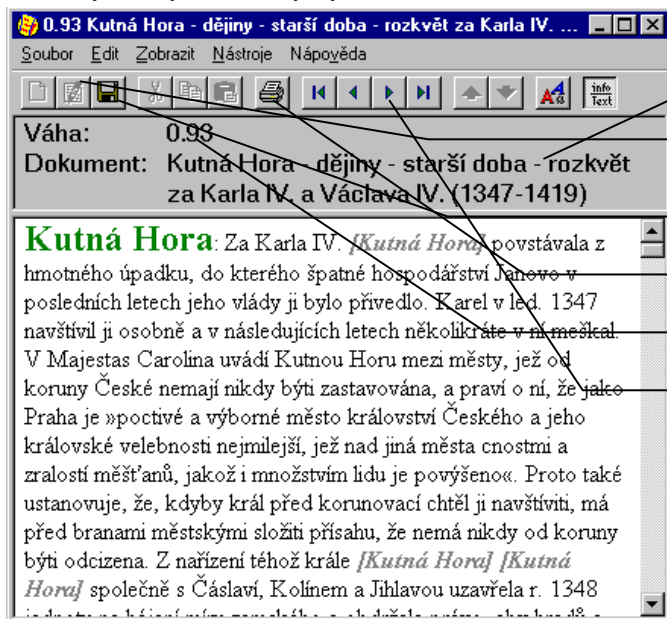


3.3. Zobrazení výsledku hledání

Zobrazit kterékoliv heslo, obsahující hledané slovo v samostatném textovém okně lze dvěma způsoby.

1. kliknutím na tlačítko "Zobrazit dokument" jak je popsáno v minulé sekci
2. kliknutím pravého tlačítka myši na konkrétním hesle v okně výsledků Tovek Agent

Heslo je pak zobrazeno v separátním textovém okně s plovoucím zlomem (okno lze tedy libovolně zvětšovat či zmenšovat). Kliknutím tlačítka "Otevřít dokument" lze rychle odskočit do digitální repliky pro jeho věrný obraz v knize. Dále lze heslo tisknout, uložit do souboru (formát RTF), použít funkci Kopíruj & Vlož a rychle vyhledávat výskyt hledaného slova, či kombinace slov dle následujícího znázornění.



specifikace nalezeného hesla

otevřít dokument (verze I. Ottovy encyklopedie)

uložit heslo (ve formátu RTF)

tisk hesla

Kopíruj & Vlož

další výskyty hledaného slova

detaillní popis aplikace Tovek Agent naleznete v sekci 4.1, podrobný návod ke tvorbě dotazů pak v sekci 4.2

3.4. Struktura textů

Pro přehlednost zpracování *Ottovy encyklopedie* jsme v textech verze II. zavedli následující úpravy

- definovali jsme **hesla 1. a 2. úrovně**, přičemž
heslo 1. úrovně začíná začátkem odstavce a je tištěno tučným písmem, například **Alžběta**
heslo 2. úrovně pak je dále rozvinuté a specifikované heslo 1. úrovně;
začíná odstavcem, či přímo v textu pouze tučně tištěnou iniciálou vyššího hesla 1. úrovně, přičemž další slovo (či slova) upřesňující význam je tištěno s proklady mezi písmeny - například **A. St u - a r t o v n a**, nebo **A. H a b s b u r s k á** (míněna Alžběta Stuartovna a Alžběta Habsburská);
- Pro účely indexace plných textů jsme tuto iniciálu rozvinuli na plný text hesla. Podle pouhé iniciály není možné vyhledávat, není tedy možné iniciále **A.** přiřadit slovo Alžběta³. Rozvinuté heslo je pak v kurzívě, v hranatých závorkách v barvě světle šedé a vždy v prvním pádu např. *[Alžběta]*
- hesel 1. úrovně je v *Ottově* slovníku naučném přibližně 140 000 a hesel druhé úrovně asi 40 000
- rozsáhlá, zejména geografická hesla jsme "strukturovali" jejich rozdělením na menší úseky, například část o chrámu sv. Víta v hesle Praha je strukturována jako
Praha - popis města - hrad královský - velechrám sv. Víta
či vláda Karla IV. pak jako
Praha - dějiny politické - vláda Karla IV.
- heslo 1. úrovně je v **červené (Alžběta)** barvě, heslo 2. úrovně v barvě **zelené (Alžběta)**

³ je zajisté možné zajistit dynamický rozvoj iniciál, přímo za běhu aplikace, při současných výkonech počítačů a jejich periferií pouze za cenu neúnosného zpomalení

3.5. Provázanost textů mezi sebou, provázanost s verzí I.

Aby možnosti vyhledávání informací ve verzi II. *Ottovy encyklopedie* byly co možná nejšířší, vybavili jsme i tuto aplikaci rozsáhlým odskokovým aparátem. Pomocí něj lze sledovat jednak provázanost hesel mezi i sebou a jednak kdykoliv nahlédnout příslušné heslo v *Ottově encyklopedie* verze I.

- ve verzi II. jsou k dispozici všechny odskoky typu "viz", známé z verze I.; zvýrazněny jsou **modře podtržením**, tedy způsobem obvyklým ve světě Internetu
- každé heslo 2. úrovně obsahuje na svém konci **modře podtržený** hypertextový odskok na nejbližší vyšší heslo 1. úrovně
- každé heslo 1. úrovně pak obsahuje na svém konci seznam všech přináležejících hesel 2. úrovně ve formě **modře podtržených** hypertextových odskoků
- obrázky jsou v textech zvýrazněny kurzívou v barvě modrozelené v hranatých závorkách, včetně popisek, podle kterých lze rovněž vyhledávat například *[viz obrázek č. 3219. Planetové soukolí.]* Celá závorka s číslem obrázku i s jeho popisou je interaktivní, kliknutí na ni vyvolá odskok do digitální repliky - verze I. *Ottovy encyklopedie*.
- tabulky jsou zvýrazněny stejným způsobem, například *[viz obrázek]* a jsou rovněž interaktivní, tedy umožňují odskok do digitální repliky při kliknutí na ně
- texty jsou vybaveny základní výbavou písem a jejich řezů, tedy kurzívou, tučným písmem, ruštinou a řečtinou. V *Ottově slovníku* je přes 1000 speciálních symbolů, které jsou v plných textech nahrazeny slovem ZNAK. Pro jejich plné zobrazení, důležité zejména v případě hesel z oborů exaktních věd je třeba otevřít digitální repliku kliknutím na tlačítko "**Otevřít dokument**"
- stejná situace jako se speciálními symboly, je se složitějšími vzorečky - matematickými i chemickými, které jsou v plných textech nahrazeny slovem vzorec, pro jejich plný náhled je opět nutno otevřít digitální repliku tlačítkem "**Otevřít dokument**"

3.6. V případě problému...

Ottova encyklopedie obecných vědomostí verze II. je vzhledem k velkému množství dat poměrně složitá aplikace (ve vteřinách prohledáváte téměř 30000 stran textu, což by donedávna bylo pravděpodobně na osobních počítačích téměř nemožné...).

Před uvolněním jsme celou aplikaci důkladně prověřovali mnohdy i za velmi bizarních situací. Všechny problémy, které se nám podařilo nasimulovat, vždy nějak souvisí s neúplnou konfigurací, či s jejím porušením. Níže uvádíme seznam těch nečastějších včetně chybových hlášení, kterými se případně projevují

- **CHYBOVÉ HLÁŠENÍ TOVEK AGENT:** ***Žádný zdroj není aktivní - použijte aplikaci Tovek Admin a nějaký zdroj vytvořte, či zpřístupněte***

řešení:

klikněte na záložku Zdroje v aplikaci Tovek Agent,

- v případě, že je přítomen zdroj "Ottova encyklopedie", zapněte jej (kliknutím do bílého čtverečku vedel jeho názvu), pak klikněte na tlačítko "Použít pro další"
- v případě, že nevidíte zdroj "Ottova encyklopedie", proveďte výše zmíněný úkon v aplikaci Tovek Administrator, pokud ani v Tovek Administrátoru nevidíte zdroj "Ottova encyklopedie", reinstalujete celý produkt znovu

- **CHYBOVÉ HLÁŠENÍ TOVEK AGENT:** ***Neznámá chyba VDK (Generic Fail): TT2: -100002, TDK: -2 nejsou texty, či jsou neúplné***

řešení:

spusťte instalační program a reinstalujte položku datové soubory pro vyhledávání

4. Doplnky

4.1. Použití aplikace Tovek Agent

Tovek Agent - vyhledávání

Pomocí Agentu se vyhledávají ve zdrojích informace uložené v zaindexovaných dokumentech. Inicializaci a zpřístupnění zdroje realizuje instalační program. Zdrojem je v našem případě *Ottova encyklopedie*.

Okno dotazu

Okno dotazu (horní okno) slouží pro zadávání dotazů. Způsob kladení dotazů je popsán v kapitole 4.2, Pomůcka při tvorbě dotazů. Jednotlivé karty slouží pro různé typy dotazů.

Text dotazu je možné vložit napsáním, přes schránku Windows (copy, paste) nebo vybráním a přetažením textu z okna jiné aplikace. Touto aplikací může být i prohlížeč dokumentů s podsvícenými klíčovými slovy. Takto je možné velmi rychle zpřesňovat dotazy.

Okno výsledků

Ve spodním okně se zobrazí výsledky hledání seřazené podle relevance, což je jejich váha vzhledem k dotazu.

U každého dokumentu je uváděna váha (relevance), zdroj (kolekce) a jeho identifikace. Implicitně je seznam nalezených dokumentů seřazen podle váhy. Lze ho však řadit podle všech kritérií, a to kliknutím na záhlaví sloupce.

Ve spodní části okna (status bar, stavový řádek) se zobrazuje počet dokumentů v seznamu a celkový počet prohledaných dokumentů v aktivních zdrojích.

Zobrazení nalezeného dokumentu

Dokument lze zobrazit třemi způsoby. Dle nastavení v menu *Nástroje | Nastavení* se dvojklikem na jméno dokumentu získá jedna z možností:

Zobrazit dokument - zobrazí text dokumentu s podsvícenými klíčovými slovy

Otevřít dokument - dokument se otevře v aplikaci, ve které byl vytvořen

Info o dokumentu - stručná informace o dokumentu (počet nalezených slov, délka, datum)

Ostatní volby lze získat kliknutím pravým tlačítkem myši. Pod Windows 95 je implicitní volba (dvojklik) tučná.

Menu Soubor

Obsahuje základní funkce pro práci se soubory.

Nový - otevře nový dotaz.

Otevřít - zobrazí dialog pro otevření dříve uloženého dotazu.

Uložit - uloží aktivní dotaz do souboru typu Tovek Query - TQR.

Uložit jako - uloží aktivní dotaz pod jiným jménem.

Konec - ukončí aplikaci Tovek Agent.

Menu Edit

Zpřístupňuje standardní editační funkce *Zpět (Undo)*, *Kopiruj (Copy)*, *Vystříhni (Cut)*, *Přilep (Paste)*, *Vyber vše*.

Menu Zobrazit

Toolbar - zobrazí/vypne lištu s ikonami.

Status Bar - zobrazí/vypne spodní řádek s nápovědou.

Rozdělení - umožní změnu poměru mezi polem dotazu a polem výsledků.

Menu Dotaz

Hledej - spustí vyhledávání .

Zdroje - obsahuje nastavení jednotlivých zdrojů a počtů dokumentů, které mají být zobrazeny.

AutoRun - je-li tato položka zatržena, probíhá vyhledávání okamžitě při zadávání dotazu. Tato funkce je implicitně zapnuta.

Iterace - po jednou provedeném vyhledávání omezí tato funkce další hledání pouze na předchozí nalezené dokumenty.

Stálý dotaz - při zatržení této položky běží dotaz neustále. Je-li do kolekcí zaindexován nový dokument (např. pomocí běžící aplikace Tovek AutoIndex), je na něj uživatel upozorněn, odpovídá-li dotazu.

Jednoduchý dotaz - typ dotazu, kdy se zadávají slova oddělená čárkou nebo booleovskými operátory.

Volný text - typ dotazu, při kterém lze zadat několik slov najednou bez jakékoliv syntaxe (např. celý odstavec). Tovek Agent najde dokumenty, které nejvíce odpovídají zadanému textu.

Topic - typ dotazu akceptující syntaxi systému TOPIC, která umožňuje např. uplatnění vážení slov, dalších operátorů apod.

Dotazy jsou popsány v další kapitole.

Menu Nástroje

Vyjmout dokumenty - tato funkce je aktivní tehdy, je-li nějaký dokument v seznamu vybrán. Umožňuje dokument odstranit z kolekce a znemožnit tím jeho další vydání. Používá se např. pro odstranění již smazaných dokumentů z kolekcí.

Kontrola integrity - zkontroluje, zda všechny nalezené dokumenty v seznamu skutečně existují. Označí dokumenty, které byly smazány nebo nebyly nalezeny. Tyto dokumenty lze poté z kolekcí odstranit funkcí *Vyjmout dokumenty*. Pozor, tuto volbu používejte pouze pro soubory!

Nastavení - zde se nastavuje funkce pro dvojklik myši na nalezeném dokumentu, rychlost funkce *Auto-Run* a zapíná se možnost vidět kompletní názvy dokumentů. Poslední volba je pro vyhledávání ve speciálních zdrojích, které byly připraveny podle postupů v Tovek Tools Developer's Kitu. Jméno dokumentu může obsahovat informace, které jsou pro většinu uživatelů nezajímavé a proto jsou implicitně skryty; touto volbou se však dají zobrazit.

4.2. Pomůcka při tvorbě dotazů

Co je to dotaz

Dotaz v Tovek Tools je seznam slov, která se hledají v dokumentech.

Na základě dotazu, resp. zadaných slov, jsou vybrány dokumenty, které tato slova nebo jejich odvozeniny obsahují. Seznam dokumentů je pak seřazen podle váhy, tj. podle míry relevance, do jaké obsah dokumentu odpovídá dotazu. Tato váha se počítá podle relativní četnosti slov v textu. Na prvních místech se tedy objeví dokumenty, ve kterých se slova objevují relativně nejčastěji (vzhledem k délce textu).

Jak tvořit dotazy?

Jednoduchý dotaz

Jednoduchý dotaz je základním typem dotazu. Sestává ze seznamu slov oddělených čárkou nebo booleovskými operátory (AND, OR, NOT).

Příklad dotazu typu *Jednoduchý dotaz*

Karel IV, Praha, Čechy

Slova se spojují pomocí operátorů a ty pak rozhodují o seznamu výsledků a vahách dokumentů.

Podrobněji o tvoření *Jednoduchého dotazu*

Lze použít klasické operátory ‘and’, ‘or’, ‘not’ známé z binární logiky, dále pro fulltextové vyhledávání speciální operátor ‘<near>’, ale hlavně nejdůležitější a zároveň na použití nejjednodušší operátor Accrue, jehož zápis je čárka ‘,’. Dále lze použít speciální znaky ‘*’ a ‘?’ a s nimi související modifikátor ‘<many>’.

Operátor Accrue

Máme-li zdroje s desítkami dokumenty, potřebujeme mít na předních místech seznamu výsledků pro nás nejdůležitější dokumenty. Dá se tedy říci, že potřebujeme zjistit, jak málo nebo jak hodně nás dokument zajímá. Tím se dostáváme k operátoru **Accrue**, jehož charakteristika při vyhledávání by se dala vyjádřit ‘čím více, tím lépe’. To je přesně ta vlastnost, kterou pro určení nejdůležitějších dokumentů potřebujeme. Jedná se vlastně o spojení požadavku na úplnost a přesnost (spojení operátorů AND a OR) do jednoho.

Zadáme-li tedy jednoduchý dotaz ‘lev, tygr’, vyhledají se nám dokumenty, kde se něco píše o lvech nebo o tygrech. Výsledný seznam bude seřazen tak, že texty s největší vahou na prvních místech budou obsahovat obě slova v nejvíce výskytech, následovány dokumenty, ve kterých jsou obě slova, ale ne tak často. Na konci budou texty obsahující pouze jedno z těchto slov. Seznam tedy bude obsahovat všechny dokumenty, kde se píše o lvech nebo o tygrech, ale ty nejdůležitější budou na začátku.

Operátory AND, OR, NOT

Operátory ‘and’ a ‘or’, logické ‘a’ a ‘nebo’, používáme pro požadovaný výskyt obou slov najednou nebo alespoň jednoho z nich v dokumentu. Dotaz ‘lev and tygr’ znamená, že hledáme pouze ty texty, kde se píše jak o lvech, tak o tygrech. Dotaz ‘lev or tygr’ vyhledá dokumenty obsahující alespoň jedno z těchto slov. Operátor ‘or’ je podobný operátoru Accrue - rozdíl je pouze ve vahách výsledků. Pomocí operátoru Accrue se získá mnohem přesnější obrázek o relevanci daných dokumentů.

Pokud se má vyloučit dokument obsahující některé slovo, použije se operátor ‘not’, logické ‘ne’. Dotaz ‘lev and not tygr’ tedy vyhledá ty dokumenty, kde se mluví o lvech a zároveň ne o tygrech. Operátory se také dají mezi sebou kombinovat. Například dotaz ‘lev and tygr, zebra and žirafa’ vrátí pouze ty dokumenty, kde je něco o lvech a o tygrech nebo o zebrách a žirafách s tím, že pokud se někde píše o obojím, budou tyto dokumenty na předních místech.

Další možnosti

V dotazech také můžeme použít speciální znaky pro hledání, tzv. **wildcards** (nebo též hvězdičková konvence). Znak '*' umí nahradit jedno a více písmen, znak '?' vždy písmeno jedno. Oba znaky mohou být použity kdekoliv v slově dotazu. Například '*ník' vyhledá texty se slovy 'úředník', 'zedník', 'patník' nebo 'pětník' a dotaz 'l?v, tygr' najde vše o lvech, lovu a tygrech. Kombinujeme-li však výrazy se speciálními znaky nebo s operátory 'and', 'or', je vyjadřování relevance nepřesné, resp. se o něj přichází a všechny vyhledané dokumenty mají váhu 1. Dá se tomu předejít modifikátorem '<many>' (např. '<many> l?v'), který umožní počítat i relativní četnost těchto neúplných slov.

Další velmi užitečný operátor je '<near>'. Ten by se dal charakterizovat jako 'čím blíže, tím lépe'. To znamená, že vyhledáváme-li v dokumentech podle dotazu 'lev <near> tygr', bude relevance dokumentů tím větší, čím tyto dvě slova budou k sobě v textu blíže. Chceme-li omezit vzdálenost mezi slovy na konkrétní počet slov, doplníme za 'near' lomítko a číslo udávající požadovaný počet včetně obou našich slov. Dotaz 'lev <near/3> tygr' vyhledá ty dokumenty, kde mezi hledanými slovy bude pouze jedno nebo žádné.

Odvozené tvary

Jistě vás již napadlo, co bude například s dokumenty se slovem 'bankou', pokud máme zadán dotaz 'banka'. I toto slovo se vyhledá, protože při vyhledávání se používá tzv. stemming, což je vyhledávání podle základu slova. Při dotazu 'banka' se naleznou nejen ty dokumenty, kde je přímo slovo 'banka', ale též kde se nachází 'banky', 'bankou' a 'bankami', nikoli však už 'bankomat'. Pokud se má vyhledat jen a pouze jeden tvar, uvádí se v dotazu v uvozovkách - např. „banka“.

Daná technologie má v češtině jedno omezení - neumí změnu kmenové souhlásky, tj. např. nezmění 'kozel' na 'kozla'. Protože existují teoreticky až tři, ale prakticky jen dva kořeny u těchto slov, je dobré uvádět oba tvary spojené operátorem Accrue (čárkou), například 'kozel, kozla'. V případě potřeby je možné použít i jinou technologii (není součástí standardní instalace), ovšem za cenu výrazného zpomalení vyhledávání a indexace.

Volný text

Dotaz typu Volný text hledá dokumenty podle jakéhokoli textu.

Tento typ dotazu není tvořen žádným výrazem, neboť může obsahovat jakákoliv slova, například celý odstavec. Váha dokumentů ve výsledném seznamu je pak dána jak počtem nalezených slov, tak podobností konstrukce zadaného textu s dokumentem. Text do editovacího pole můžeme zadat vepsáním hledaných slov nebo přímým přetažením vybraného textu (drag&drop) např. z MS Wordu.

Příklad dotazu typu *Volný text*

Karel IV. založil v roce 1348 universitu

5. Závěrem

Mé poděkování patří

- Pánu Ježíši Kristu za to, že nám dal sílu, schopnosti a možnosti v projektu *Ottovy encyklopedie* pokračovat a dovolil předložit Vám její další část k ještě většímu užitku či potěšení
- Ondřeji Šandovi za skvělé uchopení celé problematiky přípravy aplikace vyhledávání v plných textech i dat *Ottovy encyklopedie* pro ni
- Michalu Hruškovi z firmy Tovek za profesionální spolupráci
- mé ženě Jitce a dětem Ondrovi a Adélce za schopnost a ochotu překousnout chvíle, kdy měl "Otík" přednost
- Vám všem, zákazníkům, kteří jste si *Ottovu encyklopedii* obecných vědomostí doposud zakoupili a tím nám poskytli prostředky pro pokračování projektu