

FINDING KONTRA SEARCHING PŘI HLEDÁNÍ INFORMACÍ V KNIHOVNÍCH SYSTÉMECH

Jan Kastl, Vysoká škola ekonomická v Praze

Přístup k nejrůznějším informacím prostřednictvím sítě Internet je pro řadu lidí v současnosti již běžnou záležitostí. Logicky potom vystupuje do popředí otázka, jak nejspíše takovéto informace anebo informace podobné nalézt v rozsáhlém světě sítě Internet.

Nejběžnější postupy jsou asi dva. Buď uživatel využije služby některého Web-search(e) a pak se většinou probírá řadou internetových WWW-stránek, zda některá obsahuje hledané informace. Zpravidla se takto ale nenaleznou informace o existujících knihách k dané problematice. Potom se uživatel obvykle propojí do (svého) oblíbeného knihovního systému některé „bohaté“ knihovny a zde se (většinou pomocí dotazovacího jazyka) snaží vyhledat publikace žádaného zaměření. (A ty pak různým způsobem shání.)

V obou případech se z hlediska postupů i principů pro uživatele jedná o dosti podobnou situaci. V obou případech činnost také v češtině obdobně popisujeme, jako vyhledávání či hledání. Přesně takto byly činnosti zmiňovány v příručce „Informační služby v počítačových sítích“ [1] , jejíž části se mnohdy objevovaly i v jiných našich publikacích. Zmiňovaná příručka, přesněji řečeno příručky, byly překlady „Guide to Network Resource Tools“ [2] , které každým rokem aktualizovalo sdružení EARN a volně poskytovalo pro nekomerční využití. V originálech příručky se však objevují dva různé termíny: finding – pro nalézání internetových zdrojů a searching – pro hledání v databázi přístupné na některém internetovém počítači.

Je třeba poznamenat, že zmiňované tři verze příručky jsou z let 1993-94 (a asi s půlročním odstupem se pak objevovaly české překlady), tedy z doby, kdy ještě WWW-servery nebyly tak rozšířeny. Přesto toto pojmové rozlišení vyjadřuje koncepčně hlubší rozdíly mezi získanými výsledky. Finding znamená nalezení (počítačových) souborů nějakým způsobem přístupných v Internetu (přes FTP , Gopher, WWW). Searching naopak hledá záznamy (počítačové rekordy) v nějakém konkrétním (informačním) systému. A v tom je možná vidět i dosti velký rozdíl mezi typickými internetovými informacemi, t.j. soubory oproti záznamům v nějakém klasickém informačním systému, například rešeršním systému nebo v katalogu knihovního systému.

A můžeme pokračovat dále v tomto srovnávání. Proto je (či býval) nejčastější přístup k informačnímu systému pomocí TELNETu a přístup k internetovým souborům přes již zmiňované FTP a nyní WWW. V posledních

letech ovšem jednoznačně v Internetu začalo dominovat WWW svým principem orientované souborově. To zároveň vysvětluje nutnost implementace různých programových nadstaveb pro WWW-přístup do klasických systémů poskytujících informace. K nalezení souborů (volně) přístupných na internetových počítačích byly vytvořeny různé prostředky. Pro soubory přístupné přes anonymní FTP je asi nejznámější Archie, pro soubory přístupné přes Gopher je to Veronica a pro soubory přístupné přes WWW, tedy hlavně pro WWW-stránky, jsou to t.zv. Web-searche, jako například AltaVista, Yahoo, Lycos i další. Jest zřejmé, že je nerealizovatelné, aby šlo i přes nejdokonalější Web-search nalézt každou informaci obsahující dané slovo, kterou zpřístupňuje některý WWW-server ve světě. Proto Web-searche využívají různé ohraničující algoritmy tak, aby zprostředkovaly pokud možno nejvýstižnější odkazy. Zpravidla se proto vyhýbají odkazům, které se váží na programové nadstavby do jiných informačních systémů. Jejich typickou činností je tedy finding a jejich pojmenování jako Web-search či Web search engine tak bohužel nekoresponduje s terminologií v příručkách sdružení EARN. Nicméně je snad zřejmé, proč přes Web-searche (až na výjimky) nenalezneme k danému termínu odkazy například až z WebALEPHu. (Mezi výjimky patří, pokud někdo uvedl na WWW-stránce odkaz až na konkrétní informaci ve WebALEPHu.)

Lze říci, že snaha po nějakém způsobu integrace finding(u) a searching(u) je v současnosti problémem ve vývoji dokonalejších prostředků pro hledání informací v Internetu. Jistou nadějí v tomto směru snad slibuje využití tezaurově orientovaných slovníků, přes které by vyhledávací stroje mohly uživatele navigovat i na obecnější informační zdroje, kde by uživatel sám svůj dotaz zopakoval anebo zpřesnil. Odtud pak plyne jednoduchá úvaha, že to, co zatím nedokáže Web-searche, by mohlo jít „manuálně“ dodělat na straně informačního systému (alespoň pro nejobecnější úroveň). Představme si třeba katalog určité knihovny přístupný (nějak) ze sítě Internet. Asi každá takováto knihovna provozuje též WWW-server s informacemi o provozu knihovny, o poskytovaných službách i o dalších aktivitách. A na některé (z předních) stránek je obvykle odkaz pro přístup do knihovnických katalogů, a to buď TELNETem anebo opět přes WWW využívající obvykle další programovou nadstavbu. (Anebo samozřejmě obojí.)

Nebylo by asi příliš obtížné umístit na WWW-server také informační stránku nebo stránky, kde by byly uvedeny nejvýznamnější tematické okruhy, které jsou pro knihovní fond (= katalog) nějakým způsobem charakteristické. V analogii PSH (Polytematický strukturovaný heslář) Státní technické knihovny v Praze řekněme „tematický heslář“ (s termíny nejvyšších vrstev). Pokud by pak uživatel Internetu zadal nějaký takový termín pro nalezení odkazů přes Web-search, měl by (po čase) dostat i odkaz na WWW-stránku tohoto tematického hesláře. – A z ní by samozřejmě mělo být zřejmé, jakým způsobem se lze k takovýmto informacím v knihovním katalogu dostat.

Zároveň je třeba respektovat způsob a schopnosti současného programo-

vého vybavení Web-searchů, aby znaly odkazy na stránky s oněmi termíny. Stránky tematického hesláře tedy umístít co nejlíže úvodní WWW-stránce, vyhnout se pokud možno programovým odkazům, aby URL-specifikace k stránkám obsahujícím termíny byly co nejkratší, nejjednodušší, bez upřesňujících parametrů a podobně.

Možná jeden ilustrující příklad. Řekněme, že nějakého člověka v zahraničí zajímají naše knihy nebo časopisy o letectví. Zadá-li tento termín programu Web-search, možná dostane odkaz na WWW-server redakce časopisu o letectví. Logicky nejspíše dospěje k tomu, aby zkusil zadat i termín doprava a třeba snad i technika, technický a podobně. A takto by mohl dospěti až k odkazům na některou WWW-stránku Státní TECHNICKÉ knihovny. Pro uživatele neznalého češtiny, by ale takovýto postup mohlo provádět i programové vybavení Web-searche, ve kterém by byly uplatněny jednoduché prvky umělé inteligence s tezaurovou podporou. Bohužel ale musím říci, že třeba i současný vývoj u nás v podpoře umělé inteligence při vyhledávání na Internetu se zabývá mnohem složitějšími ale také mnohem méně využitelnějšími prvky z umělé inteligence.

Nejen proto by mohly manuálně vytvořené WWW-stránky tematického hesláře u nás přispět k snadnějšímu nalezení informací, zvláště pokud by existovala i varianta zmíněných stránek bez našich diakritických znamének. Jak bylo vlastně naznačeno již v úvodu, uživatel znalý našich poměrů ale nejspíše okamžitě vyzkouší pro publikace o letectví informační zdroje STK (aniž by prováděl předchozí úvahu). Vychází přitom ze zkušeností a znalosti našich poměrů. Mohli bychom obrazně říci, že k tomu využívá principů přirozené inteligence, jež je dosud daleko obtížnější v umělé inteligenci nějak algoritmičsky simulovat. V oblasti takovéto „přirozené inteligence“ se ale při vyhledávání přes Web-searche asi nedá zatím očekávat výrazný posun. A v tomto směru by možná tematický heslář mohl zprostředkovat i souvislosti jinak těžko odvoditelné.

Řekněme, že by se informace o publikacích o letectví, letecké dopravě a podobně daly časem nalézt třeba i v děčínské knihovně. K této souvislosti se ovšem dá asi dost těžko dospět v podstatě lexikografickými vazbami vůči jménu knihovny. Důvodem přitom může být to, že v Děčíně nyní existuje bakalářské studium Dopravní fakulty ČVUT a větší zájem čtenářů o literaturu s dopravní tematikou se logicky bude snažit knihovna v Děčíně uspokojovat. A právě takovýto posun tematického zaměření je pak možno v termínech hesláře uživatelům poskytnout dříve, než by se tato nová „všeobecná znalost“ rozšířila postupně i jen u nás do jiných oblastí.

- [1] Informační služby v počítačových sítích. Praha, VC ČVUT 1993. 117 s.
- [2] Guide to Network Resource Tools (Version 1.0). EARN Association 1993. (ISBN 2-910286-01-0)