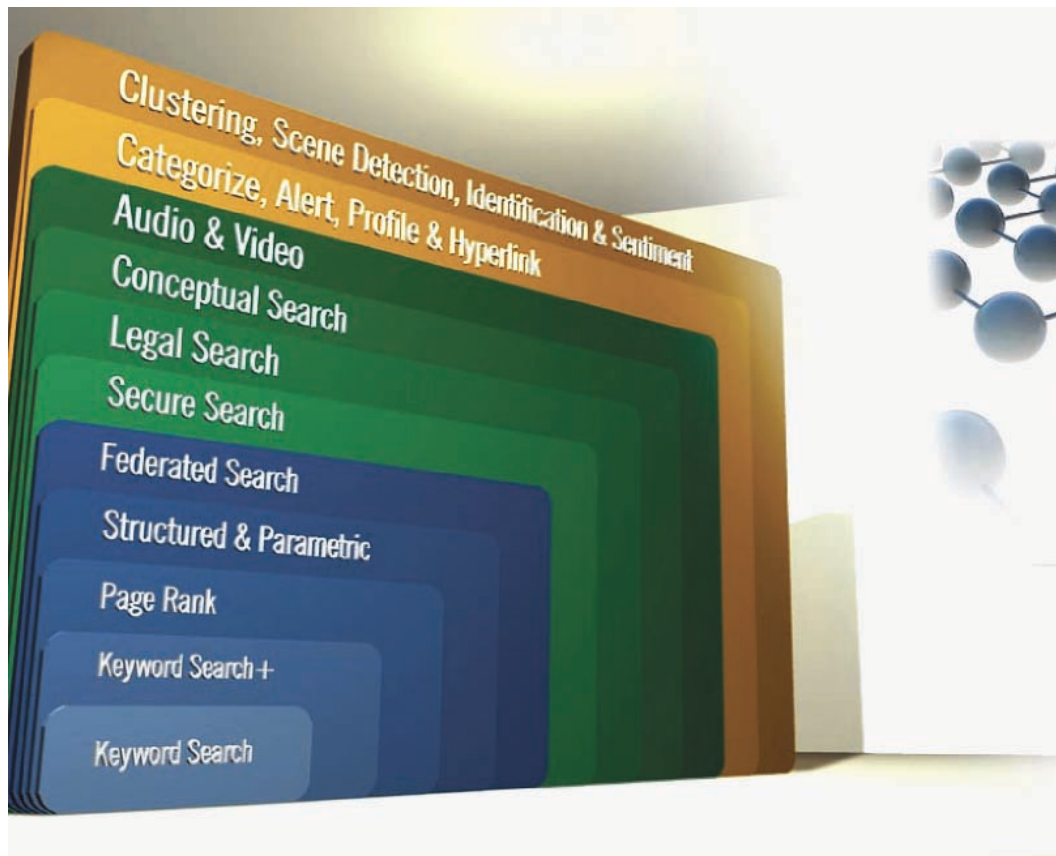


Iskanje po elektronskih arhivih

Podjetja, ki so se specializirala za rešitve inteligentnih iskalnikov, so zelo uspešna

Skupni obseg informacij v elektronskih arhivih se vsako leto podvoji. Če vemo, kaj iščemo, pridemo do želene vsebine zelo hitro, saj je to osnovna funkcionalnost elektronskih arhivov. Kaj pa, če ne vemo točnega podatka o informaciji, po kateri poizvedujemo? Takrat imamo več težav, ki jih lahko precej omilimo s tehnološkim pospeševanjem procesov preiskovanja obsežnih skladišč informacij.

Najbolj znane tehnike za iskanje podatkov uporabljajo ključne besede (ključne) oziroma metodo Boolean (z Boolovimi operatorji AND, OR in NOT logično združujemo iskalne ključne v kompleksen profil iskanja) in kot take zahtevajo vpis besede (iskalnega ključa) v tekstovno polje. Z ukazom »išči« bo iskalnik prikazal seznam dokumentov, ki ustrezajo iskalnim merilom. Čeprav takšno iskanje lahko najde dokumente s ključnimi besedami, iskalnik ne more zaznati, koliko so dokumenti relevantni oziroma kateri iskalni rezultati so v našem interesnem področju. Zato moramo ročno pregledovati najdene rezultate in jih v skladu z našo poizvedbo filtrirati. Če je najdenih ključnih



► Obstaja že mnogo različnih rešitev za konceptualno in kontekstualno iskanje informacij. Rešitve omogočajo iskanje na način, da v iskalniku opišemo našo idejo o iskani vsebini ali pa dodamo kot referenco tekst iz nekega dokumenta.

besed malo, je takšno iskanje dokumentov učinkovito. Zanimivo pa postane, če imamo najdenih

več kot tisoč rezultatov in je naša naloga preveriti in pregledati vse.

Muke pri iskanju

Poglejmo primer: recimo, da nas zanima trenutno stopnje-

vanje napetosti med Hrvaško in Slovenijo. V najbolj znani iskalnik vpišemo omenjeni državi in rezultati prikazanih zadetkov se bodo razvrstili glede na naslov oziroma število ponavljajočih se ključnih besed. Na prvem mestu se nam tako pojavi dokument, v katerem piše: »Republika Slovenija (RS) in Republika Hrvaška (RH) sta od vzpostavitve diplomatskih odnosov 6. februarja 1992 do danes razvili dobre, prijateljske in ...«

Glede na svoje področje interesa moramo iskanje ponoviti in dodati besedo incidenti. V tem primeru nam iskalnik izpiše zadetke, ki niso vezani na politično roketno prvenstvo za moške 2004« Torej moramo iskanje ponoviti in dodati besedi »mejni spori«. Rezultati bodo v tem primeru obsegali dokumente, ki vsebujejo besede »mejni spori«, spregledane pa bodo vsebine, ki se nanašajo na področje, a ne vsebujejo ključnih besed – recimo: »Vladi obeh držav sta se pogovarjali o odprtih problemih med državama. Predsednika vlad sta izjavila, da ...«

Tehnologije vzorčenja in semantike

Tehnologija je na tem področju izjemno napredovala. Podjetja, ki so se specializirala za rešitve na-

prednih oziroma inteligentnih iskalnikov, so zelo uspešna, saj uporabljajo tehnologije vzorčenja in semantike, kar omogoča simulacijo razumevanja iskane vsebine. Ker iskanje temelji na verjetnostnem modeliranju, se za iskanje ne uporablja razumevanje vsebine slovarjev ali besednih zvez (statistični modeli). Besede so obravnavane kot abstraktni simboli.

Obstaja že mnogo različnih rešitev za konceptualno in kontekstualno iskanje informacij. Rešitve omogočajo iskanje na način, da v iskalniku opišemo našo idejo o iskani vsebini ali pa dodamo kot referenco tekst iz nekega dokumenta. Rezultati iskanja se v sodobnih tehnologijah samodejno razvrščajo po tematskih sklopih na način, da so tisti, ki so najpomembnejši glede na iskani niz, na začetku. Tako lahko do želenih informacij pridemo zelo hitro in učinkovito, brez mukotrpnega pregledovanja in ponavljanja iskalnih nizov ali rezultatov. Upamo le lahko, da bo »najboljši« iskalnik na svetu čim hitreje združil eno izmed mnogih tehnologij v svoj brezplačen servis in nam tako precej olajšal naše iskanje po raznovrstnih elektronskih arhivih in svetovnem spletu.