

Metafora včelieho úľa – model vyhľadávania a odporúčania informácií

Pavol Návrat, Martin Kováčik, Anna Bou Ezzeddine, Viera Rozinajová

Ústav informatiky a softvérového inžinierstva FIIT STU

Ilkovičova 3, 842 16 Bratislava

navrat@fiit.stuba.sk, kov.martin@gmail.com, ezzeddine@fiit.stuba.sk, rozinajova@fiit.stuba.sk

Abstrakt

V článku uvádzame zjednodušený opis včelieho úľa. Na jeho základe rozpracúvame model, ktorý zahŕňa tančiaren, hľadisko a dispečerňu. Ukážeme, že náš model je verným modelom včelieho úľa aspoň v tom zmysle, že simuluje viaceré druhy typického správania sa včelieho úľa. To nás oprávňuje nazývať ho modelom včelieho úľa. Motívom našej práce však nie je príspevok k sociobiológii, ale skúmanie možností využitia tejto metafory na opis procesov, ktoré prebiehajú pri vyhľadávaní na webe. Uvedieme výsledky experimentov, ktoré podporujú tento smer výskumu.

1 Úvodná kapitola

Množstvo informácií na Internete je obrovské, niektoré sú veľmi užitočné, iné menej a niektoré vôbec. Podobne ako v Babylonskej knižnici tieto informácie sú úplne zbytočné, ak nevieme, kde ich máme hľadať. Z tohto pohľadu je význam vyhľadávania informácií kľúčový. Vyhľadávanie informácií na Internete nie je pre používateľa ľahká úloha. Existuje veľké množstvo informácií a on-line neexistuje žiadny ľudský sprievodca, ktorý by pomohol odporučiť najvhodnejšiu informáciu pre používateľa s konkrétnymi požiadavkami. Pri takomto pohľade na problematiku by bolo vhodné mať k dispozícii odporúčanie prípadu z bázy prípadov.

V tomto článku predstavíme metódu na vyhľadávanie, respektíve v užšom zmysle odporúčanie informácií (prípadov) používateľovi. Metóda využíva metaforu včiel, ktoré hľadajú najlepší zdroj nektáru. Sociálny hmyz žijúci v kolóniách ako sú včely, termity, mravce a osy sa navzájom rozlišujú svojimi organizačnými schopnosťami bez akéhokoľvek centrálného riadenia [5]. Organizácia vyplýva z interakcie medzi individuami navzájom, medzi individuami a prostredím a zo správania samotných individuí [1]. Samo-organizácia u včiel sa riadi veľmi jednoduchými pravidlami. Kolónia včiel si vyberá najlepší zdroj nektáru na základe jednoduchých pravidiel

správania. Včely vyletujú do okolitej krajiny, kde hľadajú zdroje nektáru. Keď sa včela vráti s nektárom a s informáciou o zdroji nektáru, má tri možnosti [3]:

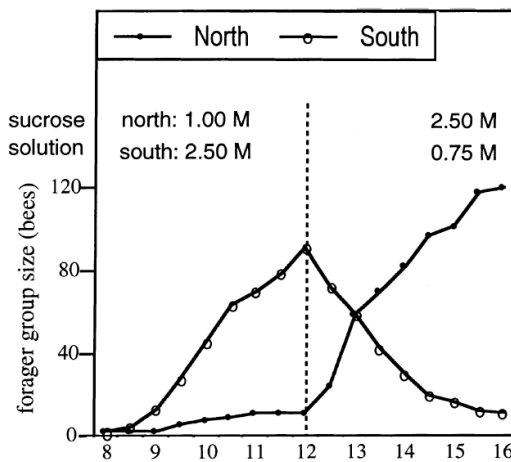
1. podeliť sa o informáciu o zdroji potravy s ostatnými včelami tancovaním, t.j. správaním, kedy včela oznamuje ostatným včelám smer, vzdialenosť a kvalitu zdroja potravy, čím sa snaží na zdroj potravy zlákať ďalšie včely,
2. pokračovať v zháňaní potravy bez zlákania ďalších včiel,
3. opustiť zdroj potravy a ísť do úľa do časti zvanej hľadisko (pozorovateľňa tančiaren), kde bude sledovať tancujúce včely a tak si vyberie svoj ďalší zdroj potravy. Každá včela má len svoj lokálny pohľad na kvalitu a pozíciu zdrojov, avšak tieto jednoduché pravidlá správania umožňujú kolónii včiel vybrať si ten najlepší zdroj nektáru.

Zvyšok článku má tieto časti. V druhej časti opíšeme stručne niektoré súvisiace práce. Samotný modifikovaný model včelieho úľa predstavíme v tretej časti. Nasleduje štvrtá časť, v ktorej podávame správu o experimente, ktorý preukazujeme podobnosť správania sa včiel v našom úli so správaním živých včiel podľa klasického experimentu. V piatej časti uvedieme problematiku ohodnocovania webových stránok a v šiestej naše experimenty preukazujúce istú schopnosť včiel v našom úli ohodnocovať stránky podobne, ako PageRank [12]. Poslednou siedmou časťou je zhodnotenie.

2 Súvisiace práce

Skúmaním správania sa včiel sa zaoberalo viacero výskumných projektov a existuje tiež viacero prístupov k skúmaniu. Sú to väčšinou práce, ktoré sa zaujímajú o modelovanie podstaty správania biologických včiel. Jednou zo základných prác, ktorá sa venuje skúmaniu správania sa včiel je práca [7]. Autori vykonali empirický experiment so včelami, ktorý bol zameraný na preskúmanie, ako si včelie kolónie vyberajú medzi zdrojmi potravy. V blízkosti kolónie boli umiestnené dva zdroje potravy. Prvý bol umiestnený 400m južne od kolónie, druhý 400m severne od kolónie. Na začiatku

experimentu bolo 12 včiel vytrénovaných, aby leteli k severnému zdroju potravy, 15 včiel, aby leteli v južnému zdroju. Zdroje potravy však mali rozdielnu koncentráciu (kvalitu) nektáru. Severný zdroj mal koncentráciu 1.0 jednotiek, južný bol kvalitnejší s 2.5 jednotkami. Experiment prebiehal od 8:00 do 16:00, ale zdroje o 12:00 vymenili. Pozorovanie ukázalo, že počet včiel zháňajúcich potravu z kvalitnejšieho zdroja sa s časom zväčšoval, pričom počet včiel zháňajúcich potravu z horšieho zdroja zostával malý. Po zamenení zdrojov sa zmenila aj situácia. Severný zdroj začalo navštevovať čoraz viac včiel a južný zdroj včely začali opúšťať. Obr. 1. znázorňuje výsledok experimentu. Na horizontálnej osi sa nachádza čas (od 8:00 do 16:00) a na vertikálnej osi sa nachádza počet včiel, ktoré nosia nektár z konkrétneho zdroja potravy.



Obr. 1. Výsledok pozorovania, ako si včelí kolónia vyberá medzi zdrojmi potravy [7].

Ďalšia práca, ktorá sa zaoberá problematikou správania sa včiel, je [8]. Model včely, ktorý autori predstavili, je zložitejší a rozlišuje niekoľko "stavov" včely. Včela svoju "kariéru" začína ako nezamestnaná začiatovníčka, teda nemá žiadne znalosti o zdrojoch potravy v teréne. Včela môže spontánne začať hľadať zdroj potravy, kedy sa z nej stáva prieskumníčka. Druhou možnosťou je začať hľadať zdroj potravy po tom, ako sa včela zúčastnila pozorovania vábného tanca. Vtedy sa z nej stáva odvedená včela a má približnú informáciu o polohe zdroja. Hneď ako včela nájde zdroj potravy, zapamätá si o ňom podrobnosti a začne z tohto zdroja znášať nektár. Týmto sa včela stáva zamestnanou včelou. Včela sa môže opäť stať nezamestnanou včelou a to v prípade, že sa zdroj potravy vyčerpá. Pokiaľ si včela v pamäti uchováva podrobnosti o zdroji, nazýva sa skúsená nezamestnaná včela. V prípade, že včela vykoná prieskumný let ku zdroju, ktorý kedysi opustila, nazýva sa inšpektor. Ak sa nezamestnaná skúsená včela zúčastní pozorovania

vábného tanca, kde odporuje pozíciu zdroja, ktorý sa podobá na ňou zapamätaný zdroj, včelu nazývame reaktivovaná včela. Ak sa nezamestnaná skúsená včela zúčastní pozorovania vábného tanca a odporuje z neho pozíciu zdroja potravy, ktorý jej nie je známy, včela sa stáva odvedenou včelou. Ďalšou možnosťou je, že včela spontánne začne hľadať nové zdroje potravy v prípade, že nebude nasledovať žiadnu tancujúcu včelu. Cieľom autorov v tejto práci bolo vytvoriť simulačný model, napodobňujúci reálne správanie sa včiel. Autorom sa skutočne podarilo nasimulovať veľmi podobné výsledky ako [7].

Na základe matematického modelu z [4] navrhli autorky v [6] prototyp multi-agentového odporúčacieho systému, ktorý využíva správanie sa spoločenstva včiel. Na rozdiel od predchádzajúceho modelu, v ktorom sa autori snažili čo najvernejšie priblížiť k realite, autorky v [6] sa snažili vytvoriť čo najjednoduchší model, ktorý by bol použiteľný v oblasti získavania informácií. Model z [4] rozšírili tak, aby bolo možné použiť viacero zdrojov potravy. Každý agent reprezentuje včelu s týmito vlastnosťami:

pravdepodobnosť opustenia zdroja potravy P_X^i ,

pravdepodobnosť zlákania iných včiel tancovaním P_D^i ,

pravdepodobnosť nasledovania zdroja potravy P_F^i .

Agenty vyhľadávajú vyhovujúce prípady v báze vyriešených prípadov. Použitím tejto terminológie reprezentuje každý prípad v báze prípadov zdroj nektáru. Pri dopyte od používateľa včely-agenty vyrazia za potravou, pričom sa snažia nájsť najlepší zdroj nektáru (prípady, ktorý najlepšie vyhovuje používateľovmu dopytu).

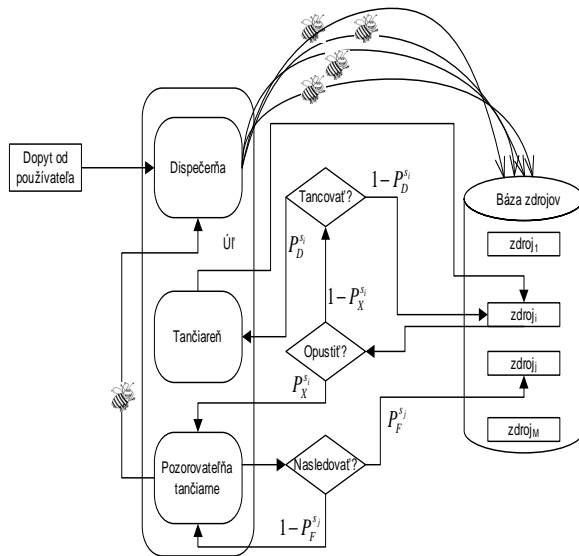
3 Model včelieho úľa

Cieľom práce je perspektívne vytvoriť jednoduchý nástroj na vyhľadávanie a odporúčanie informácií tak, aby používal opísanú metaforu včiel hľadáajúcich najlepší zdroj potravy. Požiadavka jednoduchosti je dôležitá pre použiteľnosť nástroja na vyhľadávanie z veľkého množstva informácií. Bezprostredný cieľ je preto navrhnuť podľa možnosti jednoduchý model

Pri návrhu modelu sme vychádzali z práce [6]. Pôvodný model včelieho úľa pozostával z tanciarne a hľadiska. Tento model bol opísaný v [9].

Oproti pôvodnému modelu sme do úľa pridali oddelenie dispečerňa (Obr. 2). Zavedenie tohto oddelenia prináša do modelu flexibilitu. V pôvodnom modeli pozorujúca včela musela čakať v hľadisku, pokiaľ si nevybrala včelu, ktorú bude nasledovať. Mohla nastať situácia, že pozorujúca včela vyhladovala v hľadisku, či už preto, že ju žiadna

včela v tančiarni nezlákala, alebo preto, že v tančiarni neboli žiadne tancujúce včely. V pôvodnom modeli sa včela nemohla rozhodnúť ísť hľadať nové zdroje potravy, ale musela zostať v hľadisku, pokiaľ ju nezlákala iná včela. Z dispečerne sa včely náhodne rozposielajú do bázy zdrojov, aby hľadali vhodné zdroje potravy. Zavedenie dispečernej sa môže javiť ako vážne koncepčné rozšírenie modelu. Oddelenie dispečerňa podľa biologického modelu u živých včiel neexistuje. Keď sa však lepšie zamyslíme, oddelenie dispečerňa sa dá vnímať virtuálne – včela sa rozhodne ísť hľadať nový zdroj potravy. Zmena, ktorá v modeli nastala, spočíva v tom, že včely, ktoré čakajú určitý čas v hľadisku a nevyberú si včelu, ktorú budú nasledovať, prejdú opäť do dispečernej. Odtiaľ sa znova posielajú k náhodnému zdroju potravy.



Obr. 2. Model včelieho úľa pozostávajúci z dispečernej, tančiarnie a hľadiska (pozorovateľňa tančiarnie).

Významnou zmenou je aj pridanie informačného šumu pri komunikácii včiel v úli. Ako už vieme, včely si informácie o zdrojoch potravy odovzdávajú tancovaním. Tancujúca včela tancom oznámi ostatným včelám pozíciu zdroja potravy. Miera informačného šumu teda určuje, ako nepresne tancujúca včela oznámi polohu zdroja pozorujúcej včele. Keďže včely neuvažujú priestorovú polohu zdrojov, informačný šum bude pravdepodobnosť, s akou tancujúca včela oznámi zlú "adresu" zdroja pozorujúcej včele. V prípade úplne zašumenej adresy sa včela posielajú na náhodný zdroj potravy. Ak sa "adresa" zdroja odovzdá bez chyby, včela sa posielajú na správny zdroj.

Poslednou zmenou je pridanie chyby vyhodnocovania kvality zdroja. Pre zdroj s_i ($i \in \{1, \dots, M\}$) označíme túto chybu ERR_i , a môže nadobúdať hodnoty od 0 do 1. Táto chyba sa berie do úvahy pri výpočte podobnosti používateľovho dopytu a zdroja.

Parametre modelu:

- počet včiel N ($BIOR + BISB$)
- počiatočná distribúcia včiel, teda určenie, koľko včiel zostane v hľadisku $BIOR$ a koľko vyletí do terénu $BISB$
- maximálny čas tancovania včely za určitý zdroj potravy MDT ; konkrétny čas tancovania je však závislý na kvalite zdroja ($MDT * kvalita$)
- maximálny čas, ktorý môže stráviť včela v pozorovateľni OT , tento je nemenný, včela ale môže byť zlákaná inou včelou pred uplynutím tohto času
- informačný šum $NOISE$, teda nepresnosť pri vymieňaní informácie medzi tancujúcou a pozorujúcou včelou
- chyba vyhodnotenia kvality zdroja ERR_i .

Proces vyhľadávania najlepšieho zdroja potravy sa začína dopytom od používateľa. Následne $BISB$ včiel vyletí z úľa, pričom každá z nich sa pošle na náhodný zdroj potravy. Každá včela vyhodnotí kvalitu tohto zdroja vzhľadom k dopytu od používateľa. Čím väčšia je podobnosť medzi používateľovým dopytom a zdrojom, tým vyššia je kvalita zdroja. Kvalita zdroja s_i je označená Q_i a má na ňu vplyv už spomínaná chyba vyhodnotenia kvality zdroja ERR_i .

Včela sa potom rozhoduje, či zdroj potravy opustí alebo nie (rozhodovací blok *Opustiť?*). S pravdepodobnosťou $P_X^{s_i}$ ($P_X^{s_i} = 1 - Q_i$) včela zdroj potravy s_i opustí a s pravdepodobnosťou $1 - P_X^{s_i}$ neopustí. Ak sa včela rozhodne opustiť zdroj potravy, prechádza do hľadiska. Ak sa včela rozhodne neopustiť zdroj potravy, má možnosť zlákať pre zdroj potravy ďalšie včely tancovaním.

Včela sa teda musí rozhodnúť, či pôjde tancovať (rozhodovací blok *Tancovať?*), alebo sa vráti k pôvodnému zdroju potravy s_i . S pravdepodobnosťou $P_D^{s_i}$ ($P_D^{s_i} = 1 - P_X^{s_i}$) včela pôjde tancovať a s pravdepodobnosťou $1 - P_D^{s_i}$ nepôjde. Ak sa rozhodne tancovať, prejde do tančiarnie, kde za svoj zdroj tancuje a to tým dlhšie, čím je zdroj kvalitnejší (max. však čas MDT). Po uplynutí času tancovania sa včela opäť vráti k zdroju potravy, za ktorý tancovala.

Včely, ktoré sú v hľadisku, pozorujú včely v tančiarni a rozhodujú sa, ktoré včely budú nasledovať k zdroju potravy. Včela v hľadisku si náhodne vyberie včelu z tančiarnie. Zdroj, pre ktorý tancuje vybraná včela, označme s_j ($j \in \{1, \dots, M\}$). S pravdepodobnosťou $P_F^{s_j}$

($P_F^{s_j}$ = počet včiel tancujúcich za zdroj s_j / počet všetkých tancujúcich včiel) navštívi zdroj s_j .

S pravdepodobnosťou $1 - P_F^{s_j}$ zdroj s_j nenavštívi a znova si náhodne vyberie tancujúcu včelu. Ak si do stanoveného času nevyberie zdroj, včela z hľadiska ide do dispečernej, odkiaľ je poslaná na náhodný zdroj potravy. Tento cyklus sa podľa uvedených pravidiel stále opakuje.

Opísaný model včelieho spoločenstva a špeciálne správanie sa včely-agenta sme vyjadrili algoritmicky. Implementácia týchto algoritmov je základom simulačného experimentálneho nástroja, ktorý sme implementovali v jazyku Java. Nástroj umožňuje nastaviť jednotlivé parametre. Z množstva vykonaných experimentov uvidíme v ďalšom len niektoré, pomocou ktorých demonštrujeme niektoré vlastnosti nášho modelu. V ďalšej časti opíšeme experiment, ktorý ukazuje, že navrhnutý model skutočne odráža správanie sa včiel pri hľadaní potravy.

4 Experiment: sú to včely?

Základným experimentom v prácach zaoberajúcich sa problematikou včiel je experiment so zamieňaním zdrojov potravy. Výsledok takéhoto experimentu s biologickými včelami znázorňuje obrázok 1. Cieľom tohto experimentu je ukázať, že model dokáže dynamicky reagovať na zmeny. Vykonali sme podobný experiment ako publikovali autori v [7].

Výsledky ukázali, že počet včiel zháňajúcich potravu z kvalitnejšieho zdroja sa s časom zväčšoval, pričom počet včiel zháňajúcich potravu z horšieho zdroja zostával nízky. Po zamenení zdrojov sa zmenila aj situácia. Prvý zdroj začalo navštevovať čoraz viac včiel a druhý zdroj začali včely opúšťať. Obrázok 3 znázorňuje výsledok experimentu. Počty včiel, ktoré navštevujú jednotlivé zdroje, indikujú kvalitu príslušných zdrojov, pričom počet navštevujúcich včiel zodpovedá relatívnej kvalite zdroja vo vzťahu k iným zdrojom. Je tým lepším (jednoznačnejším) indikátorom, čím viac je dosiahnutý stav z hľadiska vývoja v čase "ustálený". Pri porovnaní s pôvodným experimentom v [7] (obrázok 1) zistíme, že počty včiel, ktoré v našom modeli navštevujú jednotlivé zdroje, sa k takémuto stavu blížia rýchlejšie. Je to

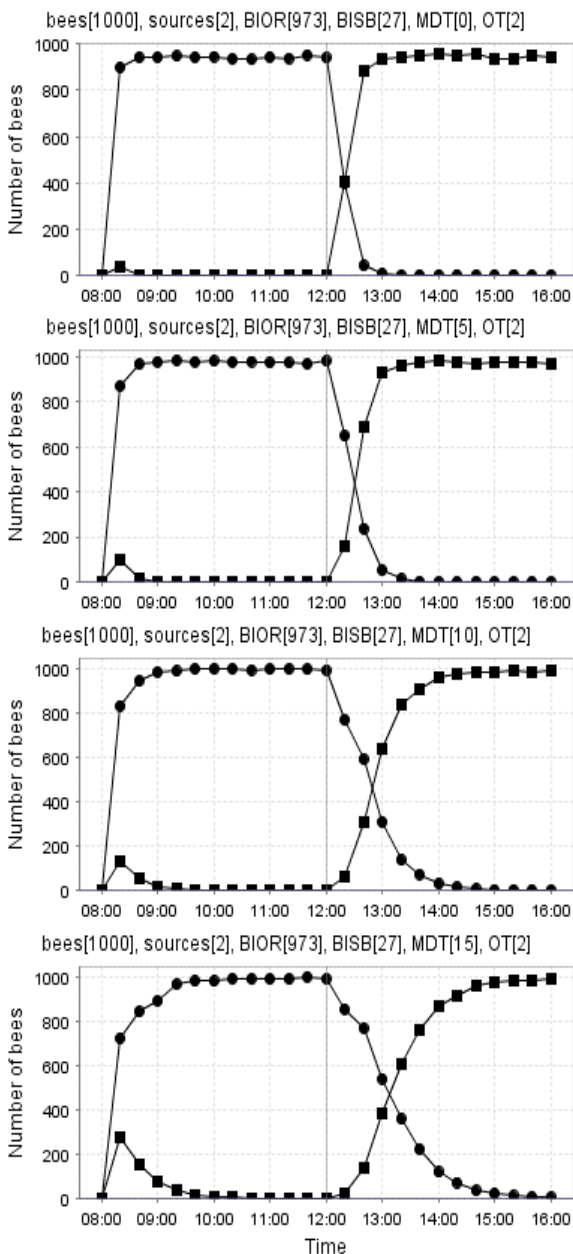
spôsobené najmä tým, že neuvažujeme žiadne časové oneskorenia spôsobené najmä pohybom včiel v priestore. Okrem dĺžky dynamiky modelu obrázok 3 ukazuje aj vplyv dĺžky tancovania na rýchlosť, s akou včely rozpoznávajú kvalitný zdroj. Maximálny čas tancovania sme postupne menili od 0 po 15, ako to postupne vidno na tomto obrázku. Je zaujímavé, že čím dlhšie včely tancujú, tým sa spomaľuje dosiahnutie stavu, v ktorom rozpoznávajú kvalitný zdroj. Je to spôsobené tým, že viacero včiel je zlákaných na zdroj potravy, pretože včela za neho dlhšie tancuje. Ak je zlákaných veľa včiel, následne aj tieto zlákané včely pôjdu tancovať (veľmi pravdepodobne) a tiež budú tancovať dlho - teda rozhodovanie celku je pomalšie.

V [11] bol vyššie opísaný model použitý na vyhľadávanie informácií z danej konštantnej množiny. Model ukázal dostatočnú dynamiku pri vyhľadávaní v tejto množine informácií a súčasne preukázal schopnosť prispôbovať sa meniacej sa kvalite informácií.

5 Ohodnocovanie webových stránok

Dôležitosť webovej stránky je z pohľadu používateľa subjektívna. Každý používateľ má rôzne záujmy a znalosti. Existuje však viacero algoritmov, ktoré zisťujú dôležitosť webovej stránky na základe jej umiestnenia a prepojenia v grafe webu. Vrcholy tohto grafu sú webové stránky a orientované hrany grafu predstavujú prepojenia medzi stránkami. Algoritmus PageRank je vlastne simuláciou imaginárneho používateľa, ktorý náhodne kliká na linky. Po každom kliknutí sa používateľ rozhodne, či klikne znova. Pravdepodobnosť, že bude pokračovať sa nazýva faktor útlmu d . Viaceré štúdie sa zaoberali štúdiom rôznych faktorov útlmu, všeobecne sa však predpokladá faktor útlmu okolo 0.85. Každá webová stránka má unikátny PageRank, ktorý nie je závislý na dopyte používateľa. PageRank teda nevyjadruje relevanciu stránky vzhľadom na zadaný dopyt. Google používa tento algoritmus ako jedno z kritérií pre usporiadanie výsledkov vyhľadávania na dopyt používateľa.

Tento algoritmus využíva štruktúru hypertextových odkazov ako vzájomné "odporúčanie" stránok. Hodnotenie stránky sa nepočíta z prostého počtu odkazov, ktoré na ňu vedú, ale berie sa do úvahy aj hodnotenie odkazujúcich stránok. Ohodnotenie stránky podľa PageRank môže dosiahnuť najviac 10 bodov. Predpokladáme, že na stránku A odkazujú stránky B, C, D, \dots . Ak $L(X)$ je počet odkazov vychádzajúci zo stránky A , vzorec na výpočet PageRank-u stránky, ktorý



Obr. 3. Experiment s výmenou zdrojov. Bolo použitých 1000 včiel a zdroje boli o 12:00 vymenené. Pri jednotlivých simuláciách bol nastavený čas tancovania včiel v tančiarni postupne 0, 5, 10, 15.

uvažuje faktor útlmu d potom bude

$$PR(A) = 1 - d + d \left(\frac{PR(B)}{L(B)} + \frac{PR(C)}{L(C)} + \frac{PR(D)}{L(D)} + \dots \right)$$

Niektoré verzie algoritmu však uvažujú aj so vzorcom

$$PR(A) = \frac{1-d}{N} + d \left(\frac{PR(B)}{L(B)} + \frac{PR(C)}{L(C)} + \frac{PR(D)}{L(D)} + \dots \right)$$

kde N je počet všetkých dokumentov v kolekcii.

Z uvedených vzorcov vyplýva, že PageRank každej stránky je odvodený z veľkej časti z PageRank-u ostatných stránok.

Všeobecný vzorec na výpočet PageRank-u stránky A

$$\text{bude } PR(p_i) = 1 - d + d \sum_{p_j \in M(p_i)} \frac{PR(p_j)}{L(p_j)}$$

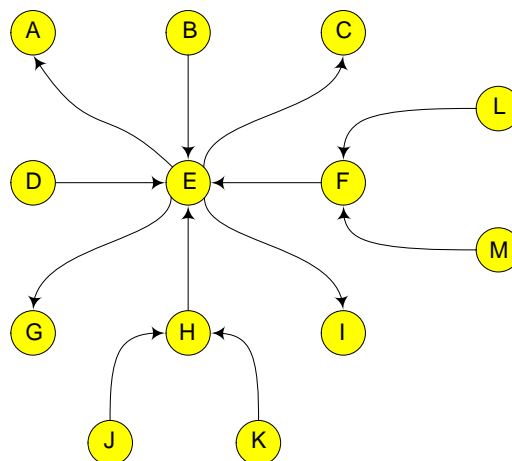
kde p_i je stránka, ktorej PageRank počítame, $M(p_i)$ je

množina stránok, ktoré odkazujú na stránku p_i a $L(p_j)$

je počet vychádzajúcich liniek zo stránky p_j .

Na výpočet dostatočne presnej hodnoty PageRank-u každej stránky je potrebných niekoľko iterácií. Počas každej iterácie sa prepočítava PageRank každej stránky v kolekcii. Počet iterácií závisí od počtu stránok v kolekcii a od zložitosti prepojenia jednotlivých stránok.

Iteratívny prístup k výpočtu PageRank-u neuvažuje „dôležitosť“ vrcholov v grafe, čo je zrejme aj z nasledujúceho príkladu:



Obr 4. Jednoduchý graf prepojenia web stránok.

Pri iteratívnom prístupe sa najskôr vypočíta PR stránky A potom $B, C, \dots M$. Najvýhodnejšie by bolo vypočítať najprv PR stránok, od ktorých sú ostatné stránky závislé. To však nie je možné v prípade, že medzi stránkami existujú cykly (čo v skutočnosti platí). Uvažujme napríklad, že chceme vypočítať PR stránky E . To znamená, že potrebujeme poznať PR stránok B, D, H a F . Na výpočet PR stránky H a F potrebujeme znova poznať PR ďalších stránok. Výhodné by preto bolo postupovať tak, aby sa podľa potreby vypočítal PR závislých stránok.

Iteratívny prístup však ide vrchol za vrcholom nezávisle od ich prepojenia.

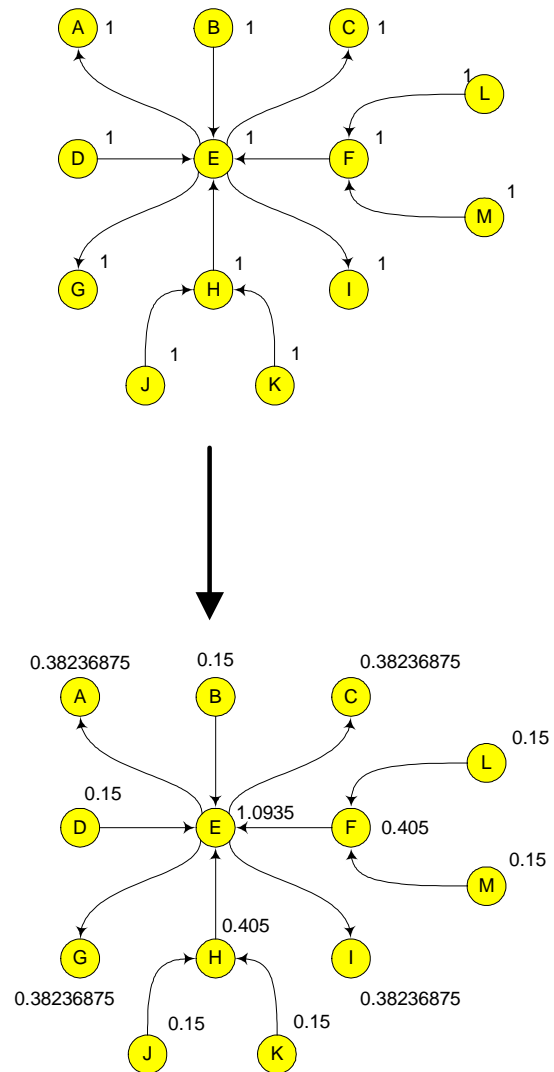
Ukazuje sa, že práve model včelieho úľa by mohol byť s výhodou použitý pre výpočet PageRank-u stránok.

Uvažujme, že by sme použili včely s mierne modifikovaným správaním. Včela by priletela na zdroj E (stránku, alebo vrchol E), vypočítala by $PR(E)$, ale súčasne by vedela, že na presnejšie určenie $PR(E)$ je treba určiť $PR(D)$, $PR(B)$, $PR(F)$ a $PR(H)$. Ak by bol zdroj E kvalitný, včela by išla do tančiarne a tam by tancovala. Netancovala by však za zdroj E , ale za jeden zo zdrojov D , B , F alebo H . Výber zdroja by bol náhodný s rovnomerným rozdelením pravdepodobnosti výberu každého z nich. Takto by sa včela snažila nalákať ostatné včely na zdroj, ktorého PR potrebuje vypočítať. Po skončení tancovania by sa včela opäť vrátila ku zdroju E , pričom ak zlákala včelu na zdroj D , B , F alebo H , môže včela na zdroji E presnejšie určiť $PR(E)$. V prípade, že kvalita zdroja je dostatočne vysoká aj po jeho aktualizácii, existuje vysoká pravdepodobnosť, že včela pôjde znova do tančiarne a tam bude tancovať za jeden zo zdrojov D , B , F , alebo H . Celý cyklus sa teda opakuje až kým nezískame výslednú hodnotu $PR(E)$. Samozrejme, že podobné pravidlá platia pre všetky zdroje. Teda ak včela navštívi zdroj F a ten je dostatočne kvalitný, včela pôjde do tančiarne a bude tancovať za jeden zo zdrojov L alebo M .

Po návšteve zdroja s_i včela vypočíta $PR(s_i)$ a následne určí kvalitu zdroja s_i ako $Q_i = PR(s_i)/MaxPR$, kde $MaxPR$ je doposiaľ najvyššia známa hodnota PR niektorého vrcholu (zdroja) grafu. Ďalšie konanie včely závisí práve na vypočítanej kvalite Q_i . Včela s pravdepodobnosťou $P_x^{s_i} = 1 - Q_i$ opustí zdroj s_i a ide do hľadiska. V prípade, že včela zdroj neopustí, môže sa v závislosti od Q_{s_i} rozhodnúť, že poletí do tančiarne alebo bude pokračovať v nosení potravy zo zdroja s_i . V prípade, že pôjde do tančiarne, bude tancovať za jeden zo zdrojov, ktoré odkazujú na zdroj s_i . Výber tohto zdroja je náhodný s rovnomerným rozdelením pravdepodobnosti výberu ľubovoľného zdroja odkazujúceho na s_i . Nazvime tento zdroj s_j . Včela teda bude v tančiarňi propagovať zdroj s_j . Včely v pozorovateľni si vyberú vždy náhodnú včelu a s pravdepodobnosťou $P_f^{s_j}$ ju budú nasledovať. Pravdepodobnosť $P_f^{s_j}$ je rovná kvalite zdroja propagovaného včelou v tančiarňi $Q_j = PR(s_j)/PR(s_i)$.

V prípade, že si včela v hľadisku po stanovenom čase nevyberie včelu, ktorú bude nasledovať, odlieta do dispečerne, odkiaľ je vyslaná na náhodný zdroj potravy.

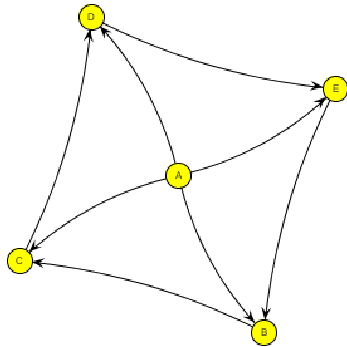
Cyklus výpočtu ohodnotenia vrcholov v grafe pomocou včiel možno opísať takto: Na začiatku je každý vrchol v grafe ohodnotený $PR = 1$. Stanovený počet včiel začne náhodne vyletovať a navštevovať vrcholy grafu. Každá včela sa po návšteve vrcholu (zdroja potravy) následne správa podľa opísaných pravidiel. Cyklus iteratívne pokračuje dovtedy, pokiaľ nie sú vrcholy grafu ohodnotené, alebo po vopred určenom počte iterácií. Začiatkový a konečný stav ohodnocovania grafu je zobrazený na Obr. 5.



Obr. 5. Výpočet PageRank-u jednoduchého grafu.

6 Experiment: vedeli by včely ohodnocovať stránky?

Graf na Obr. 6 je na pohľad veľmi jednoduchý, ale práve jeho cyklická závislosť spôsobuje, že výpočet PageRank-u iteratívnym spôsobom je veľmi zdĺhavý.



Obr. 6. Graf s 5 vrcholmi.

PageRank sa klasickým spôsobom vypočítal po 52 iteráciách. Hodnoty PageRank-u pre uvedené vrcholy sú zobrazené v Tab. 1. Hodnoty dosiahnuté naším modelom zobrazuje Tab. 2.

Vrchol	Page Rank
A	0.15
B	1.2125
C	1.2125
D	1.2125
E	1.2125

Tab. 1. Vypočítaný PageRank vrcholov grafu.

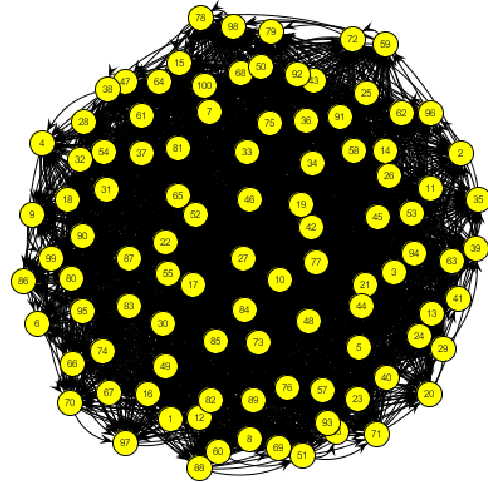
Počet iterácií	Počet včiel
13	100
20	50
40	25
80	10

Tab. 2. Počet potrebných iterácií v závislosti od počtu včiel.

Pri dostatočnom počte včiel sú dosiahnuté dobré výsledky. Klasickým spôsobom bolo potrebných 52 iterácií. Pri použití 100 včiel bolo výpočet ohodnotenia potrebných iba 13 iterácií.

Jeden z ďalších experimentov sme vykonali nad grafom na Obr. 7 s husto prepojenými vrcholmi mnohonásobnými cyklami. PageRank bol klasickým spôsobom vypočítaný po 102 iteráciách. Prehľad o výpočte PageRank-u pomocou včiel poskytuje Tab. 3.

Ako môžeme z dosiahnutých výsledkov vidieť, pri použití väčšieho počtu včiel ako je počet vrcholov v grafe dáva metóda lepšie výsledky ako klasická iteratívna metóda. Ak však znižujeme počet včiel, rapídne rastie aj počet iterácií potrebných na výpočet PR.



Obr. 7. Graf so 100 vrcholmi.

Počet iterácií na úplné	Počet včiel
77	500
174	250
401	100
2011	50

Tab. 3. Počet potrebných iterácií v závislosti od počtu včiel.

7 Zhodnotenie

Tento článok predstavil model a experimentálny prototyp systému na vyhľadávanie informácií, ktorý je inšpirovaný správaním sa včiel v procese hľadania kvalitných zdrojov potravy. Prototyp sa zakladá na pôvodnom modeli včiel z [6], kde sme navrhli niekoľko modifikácií. Napriek modifikáciám model nestratil základné atribúty, ktoré sú vlastné modelovaniu včiel pri hľadaní zdrojov potravy. Modifikácie sa ukázali byť užitočné najmä z hľadiska znižovania počtu včiel potrebných na uskutočnenie kvalitného vyhľadávania informácií.

Skúmali sme tiež použitie včiel na výpočet PageRank-u grafu. Ukázalo sa že použitie včiel na výpočet PageRank-u grafu môže byť veľmi výhodné a v mnohých prípadoch rýchlejšie ako klasický iteratívny prístup. Na rozdiel od iteratívneho prístupu, metóda využívajúca model včielieho

úľa neposkytuje vždy také presné výsledky, t.j. nie je garantované, že včely vždy navštívia všetky zdroje. Táto vlastnosť sa však dá pomerne dobre kompenzovať zvyšovaním počtu včiel, čím sa súčasne znižuje pravdepodobnosť, že včely nenavštívia všetky zdroje.

Ďalšou výhodou včiel je možnosť distribuovaného výpočtu PageRank-u veľkého grafu. Nebolo by zložité vytvoriť distribuovaný systém včiel, kde by včely spoločne komunikovali aj keď by boli rozmiestnené na iných počítačoch. Takto by mohol byť systém včiel použitý aj na výpočet PageRank-u obrovských grafov.

Výhodou je tiež možnosť kontinuálneho výpočtu PageRank-u grafu on-line. Napríklad je známe, že Google vždy za určitý čas prepočítava PageRank-y celého ním indexovaného webu. Ak by sa použil systém včiel, bolo by možné prepočítavať PageRank-y kontinuálne, tak že by nad grafom jednoducho nepretržite lietali včely a aktualizovali by PageRank-y jednotlivých vrcholov (web stránok) grafu.

Pod'akovanie

Prácu na tomto článku čiastočne podporila Vedecká grantová agentúra VEGA v rámci projektu č. VG1/3102/06 a Agentúra na podporu výskumu a vývoja v rámci zmluvy č. APVV-0391-06.

Literatúra

- [1] Bonabeau, E., Thraulaz, G., Dorigo, M.: *Swarm Intelligence: From natural to Artificial Systems*. Oxford University Press, 1999.
- [2] Borges, J. L.: *Ficciones*, edited with an introduction by Anthony Kerrigan. New York: Grove Press, 1962.
- [3] Camazine, S., Deneubourg, D., Franks, N.R., Sneyd, J., Theraulaz, G., Bonabeau, E.: *Self-Organization in Biological Systems*. Princeton, 2003.
- [4] Camazine, S., Sneyd, J.: A model of collective nectar source selection by honey bees: Selforganization through simple rules. *Journal of Theoretical Biology*, 149(1991) 547-571.
- [5] Gordon, D.: The organization of work in social insect colonies. *Nature*, 380 (1996)121-124.
- [6] Lorenzi, F., Sherer dos Santos, D., Bazzan, A.L.C. Negotiation for task allocation among agents in case-based recommender systems: a swarm-intelligence approach. In *IJCAI-05 Workshop on Multi-Agent Information Retrieval and Recommender Systems*. Edinburgh 2005. Pp. 23-27.
- [7] Seeley, D., Camazine, S., Sneyd, J.: Collective decision-making in honey bees: How colonies choose

nectar sources. *Behavioral Ecology and Sociobiology*, 28 (1991) 277-290.

[8] Vries, H., Biesmeijer, J.C.: Modelling Collective Foraging by Means of Individual Behaviour Rules in Honey-Bees. *Behavioral Ecology and Sociobiology*, 44(1998) 109-124.

[9] Návrat, P.: Bee Hive Metaphor for Web Search. In: *CompSysTech 2006*, B. Rachev, A. Smrikarov (Eds.), Veliko Turnovo, Bulgaria, June 2006, IIIA.12-1-7.

[10] Návrat, P., Kováčik, M.: Web Search Engine as a Bee Hive. In: 2006 *IEEE/WIC/ACM International Conference on Web Intelligence (WI 2006 Main Conference Proceedings)* (WI'06), Hong Kong 2006, pp. 694-701.

[11] Návrat, P., Kováčik, M., Bou Ezzeddine, A., Rozinajová, V.: Vyhľadavanie informácií pomocou včiel. In: *Znalosti 2007*, Ostrava 2007, pp. 63-74.

[12] Brin, S., Page, L.: The Anatomy of a Large-Scale Hypertextual Web Search Engine. Computer Science Department, Stanford University, Stanford. <http://www-db.stanford.edu/~backrub/google.html>