

# TraDiCe: Cestovanie v digitálnom priestore

Pavol Návrat, Marián Šimko, Jozef Tvarožek, Alena Kovárová

Ústav informatiky a softvérového inžinierstva  
Fakulta informatiky a informačných technológií  
Slovenská technická univerzita v Bratislave  
Ilkovičova 2, 842 16 Bratislava, Slovensko

{meno.priezvisko}@stuba.sk

**Abstrakt.** Projekt TraDiCe sa zaoberá výskumom získavania a vyhľadávania informácií v rozsiahlych informačných priestoroch. V príspevku prezentujeme základné zameranie a ciele projektu a predstavujeme vybrané výsledky projektu dosiahnuté výskumnými skupinami na Fakulte informatiky a informačných technológií. Prezentované výsledky pokrývajú základné oblasti, ktorým sa v rámci projektu venujeme: prehľadávanie a navigáciu v digitálnom priestore, vyhľadávania a odporúčanie, modelovanie používateľov, skupín, kontextu a domény.

**Kľúčové slová:** digitálny priestor, web, exploratívne prehľadávanie, navigácia, odporúčanie, modelovanie používateľov, modelovanie domény

## 1 Zameranie a ciele projektu

Projekt TraDiCe<sup>1</sup> sa zameriava na základný výskum získavania informácií vyhľadávaním v rozsiahlych informačných priestoroch. Informácie záujemca nachádza alebo odvodzuje v procese jeho interakcie ako s informačným prostredím zahŕňajúcim aj priateľov alebo ľudí s podobným záujmom. Metafora kognitívneho cestovania po digitálnom priestore opisuje (často zvedavého) používateľa, ktorý sa pohybuje po webe alebo nazerá do digitálnych knižníc, či už s vedomím konkrétneho cieľa alebo nie.

Používatelia zanechávajú po sebe v digitálnom priestore stopy, často bez toho, aby si to priamo uvedomovali. Ide napr. o hodnotenia, odporúčania, anotácie, nápisy na virtuálnej stene. Komunikujú s inými, tvoriac spoločenstvá spoločných záujmov. Vyjadrujú však aj svoje názory, píšú blogy či mikroblogy. Takéto a mnohé ďalšie formy implicitnej spätnej väzby sú potenciálne použiteľné na zlepšovanie kognitívnej konštrukcie významu u používateľov – spotrebiteľov informácií, resp. cestovateľov. Potenciál digitálnych knižníc čaká na odkrytie predovšetkým ich integráciou -- či už na úrovni integrácie samotných digitálnych knižníc alebo formou zjednocovania na základe konštrukcie významu naprieč jednotlivými knižnicami distribuovanom prie-

---

<sup>1</sup> <http://tradice.fiit.stuba.sk/>

store webu. V perspektívne „jednotnom“ digitálnom svete informačných zdrojov potom možno využiť pokročilé funkcie vyhľadávania, personalizáciu, sociálne sieťovanie, zohľadnenie kontextu.

V takomto priestore je dôležitý aj spôsob prezentovania a vizualizácie informácie, ktorá predstavuje potenciálnu odpoveď na zadaný dopyt. V tomto smere sa ukazuje ako veľmi perspektívna multiparadigmová vizualizácia, kde sa v závislosti od typu prezentovanej informácie a tiež stavu procesu vyhľadávania využije najvhodnejšia forma vybraná na základe povahy zobrazovaných výsledkov a tiež kontextu.

Na projekte, ktorý trvá od mája 2011 do októbra 2014, participuje viacero výskumných tímov. Na prácu riešiteľského tímu z Fakulty informatiky a informačných technológií STU v Bratislave sa zameriavame v tomto príspevku. Riešiteľský tím z Filozofickej fakulty UK v Bratislave sa venuje otázkam informačného prieskumu, skúmaniu informačného správania a budovania digitálnych repozitárov. Riešiteľské tímy Ústavu informatiky SAV v Bratislave a Fakulty elektrotechniky a informatiky TUKE v Košiciach sa zameriavajú najmä na aspekt modelovania domény a skúmajú podporné metódy pre tvorbu sémantických a konceptuálnych modelov.

## 2 Dosiahnuté výsledky

Najvýznamnejšiu časť výskumu sme realizovali v oblasti nových metód pre *navigáciu* a *vyhľadávanie* v digitálnych knižniciach. Rozpracovali sme systém Annota<sup>2</sup> umožňujúci vytvárať záložky a poznámky k ľubovoľným dokumentom na Webe [6]. Navrhli a zrealizovali sme viacero spôsobov využitia týchto poznámok pri organizácii, navigácii a vyhľadávaní v kolekcii dokumentov, pričom sme sa špeciálne zamerali na doménu digitálnych knižníc. Používaním systému Annota na tvorbu záložiek a poznámkovanie dokumentov v prostredí webu a digitálnych knižníc budujeme dátovú množinu, ktorá prepája doménové dáta (dokumenty a ich metadáta) s obsahom vytvoreným používateľmi (záložky, poznámky, značky).

Konkrétne sme v systéme Annota navrhli a overili nové metódy navigácie. Preskúmali sme metódu navigácie pomocou oblaku dôležitých slov, ktorá využíva históriu navigácie používateľov ako zdroj metadát pre personalizáciu navigácie [4], s. 86. Navrhli sme aj metódu navigácie na podporu prieskumného vyhľadávania pomocou automatickej sumarizácie textu [4], s. 89. Textové súhrny pomáhajú používateľom navigovať sa v informačnom priestore, pretože im umožňujú rýchlo zhodnotiť relevantnosť dokumentov. Skúmali sme metódy použité pri procese vyhľadávania v digitálnych knižniciach a informačných priestoroch. Navrhli sme prístup k prehliadaniu výsledkov vyhľadávania v digitálnych knižniciach pomocou pohľadov založených na zhlukovaní a navigácii využívajúcej priblíženie [5]. Navrhovaný prístup k prehliadaniu výsledkov podporuje poznávanie neznámeho priestoru výsledkov pre používateľov pomocou zobrazenia výsledkov v hierarchických zhlukoch. Vďaka týmto zhlukom môžu používatelia jednoducho a rýchlo identifikovať rozdelenie priestoru výsledkov podľa ich jednotlivých vlastností a nájsť hľadaný výsledok, prípadne sa naučiť nové poznatky o prehliadanom priestore. Po identifikácii hľadaných výsledkov

---

<sup>2</sup> <http://annota.fiit.stuba.sk>

môžu používatelia cestovať priestorom digitálnej knižnice pomocou približovania sa k zaujímavým objektom a ich vlastnostiam v grafovej vizualizácii.

V rámci vyhľadávania sme sa zamerali aj na dopytovanie používateľov. Zaoberali sme sa sémantickým dopytovaním. Navrhli sme metódu na tvorbu dopytov na základe obsahu dokumentu a k nemu pripojených poznámok [4], s. 67. Vytvorený dopyt slúži na vyhľadanie dokumentov súvisiacich s opoznámkovaným dokumentom, pričom poznámky sa používajú ako indikátory záujmu používateľa o konkrétne časti dokumentu a nájdené súvisiace dokumenty súvisia práve s týmito časťami. Na riešenie problému viacznačnosti krátkych dopytov sme navrhli metódu, ktorá dokáže odvodiť dodatočné kľúčové slová dopytu s využitím kontextu implicitnej sociálnej siete používateľa a metódu na vytvorenie tejto sociálnej siete [3].

V rámci výskumu kognitívneho cestovania po digitálnom svete je našim záberom aj podpora *personalizovaných služieb* a *sociálnych sietí*. Vo výskume metód pre odporúčanie sme skúmali možnosti využitia skupinového odporúčania [1], zahrnutie kontextu a grafových algoritmov [4], s. 115. Zamerali sme sa na zlepšenie predikcie hodnotení odporúčaných prvkov, keďže práve táto predikcia je z hľadiska úspešnosti odporúčacieho systému často kľúčová. Aplikáciou princípov skupinového odporúčania a zahrnutím navrhnutých virtuálnych skupín sa nám podarilo zvýšiť presnosť odporúčania oproti štandardným prístupom, ktoré sa dnes široko používajú.

Zaoberali sme sa aj modelovaním domény a používateľov, ktoré je v kontexte personalizácie nevyhnutné. Dôležité je ich efektívne budovanie a udržiavanie. V rámci modelovania domény sme navrhli nové metódy pre extrakciu relevantných doménových pojmov založených na využití dodatočných zdrojov informácií ako sú používateľské anotácie, vizuálne charakteristiky [4], s. 106, ale tiež pokročilé metódy zhľukovania alebo metódy inšpirované prírodou [4], s. 4.

Z pohľadu podpory kognitívneho cestovania využívajúce sociálne siete sme skúmali potenciál mikrobloggerov ako atraktívneho sociálneho informačného prostredia, kde však používatelia trpia syndrómom preťaženia množstvom informácií a načrtli sme metódy pre podporu prieskumného vyhľadávania využívajúce multiperspektívnu analýzu a spätnú väzbu používateľov [7] alebo analýzu sentimentu [2].

Okrem návrhu samotných metód sa v projekte venujeme aj tvorbe nástrojov pre platformu pre vyhodnocovanie metód. Jedným zo zámerov bola efektívna reprezentácia modelu používateľa v rámci webového prehliadača, čo tvorí jadro platformy modelovania používateľa a personalizácie vo forme rozšírenia do prehliadača. Nástroje umožňujú škálovateľnú a rozšíriteľnú architektúru platformy v prehliadači a bohaté možnosti začínajúce zberom dát, ich spracovaním a uchovávaním až po samotné scenáre prispôsobovania používateľovi. Umožňuje ďalšie možnosti personalizácie, ktoré ponúka prostredníctvom svojich personalizačných rozšírení. Vedľajší cieľ je aj na celoživotné modelovanie používateľa a navrhujeme efektívnu metódu indexovania celoživotných záznamov a modelovania rozličných charakteristík používateľa webového prehliadača.

### 3 Zhodnotenie

V článku sme predstavili zameranie a ciele výskumného projektu TraDiCe, v ktorom sa zaoberáme výskumom nových metód pre získavanie a vyhľadávanie informácií

v rozsiahlych informačných priestoroch. Viaceré z navrhnutých metód boli rozpracované a overené v rámci systému Annota, ktorý slúži na podporu práce výskumníkov. V súčasnosti záznamy zahŕňajú údaje o cca 200 používateľoch (výskumníkoch), cca 98000 dokumentov, cca 8500 záložiek od používateľov, cca 2800 rôznych tagov a cca 6000 priradení tagov. Zozbierané dáta poskytujú dátovú bazu pre ďalší výskum metód a overovanie čiastkových výsledkov. Dáta je možné získať na webovom sídle projektu<sup>3</sup>. S perspektívou narastajúceho objemu dát medzi ďalšie výzvy patrí nasadenie navrhnutých metód vo veľkých dátových korpusoch.

*PodĎakovanie.* Táto práca vznikla s podporou APVV-0208-10 Kognitívne cestovanie po digitálnom svete webu a knižníc s podporou personalizovaných služieb a sociálnych sietí.

## English summary

### *TraDiCe: Travelling in Digital Space*

Project TraDiCe focuses on fundamental research in the area of information retrieval based on search in large information spaces. In this paper, we present the goals of the project and selected results achieved by research groups at Faculty of Informatics and Information Technologies. The results include the areas: navigation in digital information space, search and recommendation, methods for modelling of users, groups, context and domain.

## Literatúra

1. Kompan, M., Bieliková, M. Personalized recommendation for individual users based on the group recommendation principles. In: Studies in Informatics and Control. Vol. 22, No. 3 (2013), pp. 331-341
2. Korenek, P., Šimko, M. Sentiment Analysis on Microblog Utilizing Appraisal Theory. World Wide Web Journal. World Wide Web (2014) 17:847–867, Springer.
3. Kramár, T., Barla, M., Bieliková, M. Personalizing Search Using Socially Enhanced Interest Model, Built from the Stream of User's Activity. Journal of Web Engineering. - ISSN 1540-9589. - Vo.12, No. 1-2 (2013), pp. 65-92
4. Návrat, P., Bieliková, M., Láclavík, M., Paralič, J., Steinerová, J. Cognitive Traveling in Digital Space of the Web and Digital Libraries. STU Press: Bratislava, 2013, ISBN: 978-80-227-4089-0.
5. Rástočný, K., Tvarožek, M., Bieliková, M. Web Search Results Exploration via Cluster-Based Views and Zoom-Based Navigation. Journal of Universal Computer Science. Vol. 19. No. 15, 2320-2346.
6. Ševcech, J., Móro, R., Holub, M., Bieliková, M. User Annotations as a Context for Related Document Search on the Web and Digital Libraries. Informatica 38 (2014) 21–30.
7. Žilínčík, M., Návrat, P., Kosková, G. Exploratory search on Twitter utilizing user feedback and multi-perspective microblog analysis. PLoS ONE 8(11):e78857. doi:10.1371/journal.pone.0078857.

---

<sup>3</sup> <http://annota.fiit.stuba.sk/dataset>