

# System AWA v podpore adaptivity webu

Kristína Machová, Jana Modrovičová

Katedra kybernetiky a umelej inteligencie, FEI,  
Technická Univerzita Košice, Letná 9, 042 00, Košice  
Kristina.Machova@tuke.sk, Jana.Modrovicova@tuke.sk

**Abstrakt.** Článok sa venuje riešeniu problému znižovania kognitívnej záťaže používateľa internetu. Popisuje systém AWA, ktorý je zameraný na odporúčanie internetových stránok používateľovi na základe odhadu jeho záujmov. Záujmy používateľa ako aj primárne skupiny webových stránok sa získavajú pomocou zhlučovacích metód. Implementácia systému je navrhnutá v prostredí JSP a pracuje v on-line režime. Článok obsahuje aj popis vykonaných experimentov a ich následný rozbor.

**Kľúčové slova:** adaptívny web, systém AWA, zhlučovanie

## 1 Úvod

Bez Internetu si dnes nevieme ani predstaviť náš profesijný ale aj súkromný život. Väčšinu informácií, ktoré potrebujeme si oveľa skôr zaobstaráme prostredníctvom Internetu ako klasickým spôsobom: telefonovaním, návštevou informačných centier, úradov práce apod. Avšak niekedy je návratnosť informácií nízka. Tento problém by mohol vyriešiť sémantický web, čo nie je jednoduché. Zdá sa, že skôr realizovateľnou úlohou je dať internetu adaptívny rozmer. Preto sme sa pokúsili znížiť kognitívnu záťaž používateľa Internetu tým, že sme vyvinuli systém AWA, ktorého úlohou je ponúkať mu stránky, ktoré spadajú do jeho oblasti záujmu. Modelovanie záujmov používateľa sme uskutočnili pomocou notoricky známeho zhlučovacieho algoritmu K-means [3]. Bolo by možné použiť aj iný zhlučovací algoritmus. Avšak, K-means je implementačne jednoduchý a efektívny algoritmus. V minulosti sme sa pokúšali riešiť túto úlohu generovaním modelu používateľa klasifikačným algoritmom HGS [4].

Implementácia systému AWA je navrhnutá v prostredí JSP a pracuje v on-line režime. To jej umožňuje prispôbovať sa podmienkam okolia. Na vývoj aplikácie bol použitý jazyk Java. Táto implementácia bola testovaná na serveri Wikipédia.

## 2 System AWA

System AWA (Adaptívna Webová Aplikácia) z hľadiska štruktúry pozostáva z dvoch rozhraní - z administrátorského a používateľského ako aj samostatnej databázy. Spolupráca sa uskutočňuje prostredníctvom databázy, ktorá je umiestnená na externom serveri. Ukladajú sa tam všetky údaje získané od používateľov od registrácie až po informácie o spôsobe používania aplikácie. Zatiaľ čo administrátor používa databázu ako zdroj informácií, u používateľa je to obojstranné tzn., že nielen

používateľ využíva databázu ako zdroj informácií ale aj databáza získava informácie o zmenách v jeho rozhraní. Celá aplikácia je navrhnutá v prostredí JSP a algoritmus ako aj práca s databázou a s webovými stránkami sú vytvorené v prostredí Javy. Databáza bola umiestnená na externom serveri hlavne z dôvodu prístupnosti počas testovania aplikácie. Je navrhnutá pomocou MySQL a je tvorená z viacerých tabuliek a to: tabuľka používateľov, hodnotení, jednotlivých stránok. Po spustení aplikácie sa zobrazí všetkým používateľom hlavná stránka, kde majú možnosť “prihlásenia” sa alebo “registrácie” do systému AWA.

## 2.1 Administrátorské rozhranie

Prihlasovanie sa administrátora je podobné ako u bežných používateľov. Odlišné je pracovné rozhranie, do ktorého sa administrátor dostane po správnom zadaní login-u a hesla. Administrátor plní úlohu kontrolóra funkčnosti systému. Vo svojom rozhraní má k dispozícii údaje o zaregistrovaných používateľoch (vidí ich mená, priezviská, login-y, e-mail, záluby). Tieto údaje je možné použiť v prípade, že niektorý z používateľov zabudne svoje prihlasovacie údaje. Môžu sa mu obratom poslať na mail, tak ako je to v mnohých internetových obchodoch. Ďalšou možnosťou je “Štatistika”, kde je viditeľné hodnotenie aplikácie z pohľadu používateľov formou ankety dostupnej v používateľskom rozhraní. Túto anketu môže každý používateľ vyplniť iba raz. Výsledky ankety sú uvedené v druhej tabuľke administrátorského rozhrania. Poslednou možnosťou je “Odhlásenie”. Pokiaľ používateľ nie je zaregistrovaný (či už používateľ administrátor alebo iný), klikne na odkaz *Registrácia*. Zobrazí sa mu registračný formulár, kde vyplní povinné údaje (meno, priezvisko, heslo, e-mail, login) a nepovinné údaje (záluby). Týmito údajmi sa naplní tabuľka používateľov v databáze, pričom každému používateľovi je priradené identifikačné číslo, ktoré spolu s login-om slúži k jeho identifikácii.

## 2.2 Používateľské rozhranie

Po zaregistrovaní a následnom prihlásení sa do systému, dostane sa používateľ do používateľského rozhrania. Tam si môže vybrať spôsob vyhľadávania:

- *Vyhľadávanie podľa kľúčového slova.* Tento spôsob vyhľadávania je podobný ako u viacerých internetových vyhľadávačov (napr.: Google, Morfeo, Altavista a pod. ).
- *Vyhľadávanie pomocou zhlukov stránok.* Používateľovi bude ponúknutý zhluk stránok, do ktorého prezeraná a vyžiadaná stránka patrí. V pravom hornom pod - okne sa zobrazia url adresy stránok z daného zhluku. Po kliknutí na danú url adresu sa stránka z hluku objaví v hlavnom pod - okne aplikácie.
- *Vyhľadávanie pomocou podobných používateľov.* Používateľ si môže prezrieť aj stránky, ktoré v minulosti prezerali jemu podobní používatelia. Podobnosť používateľov je určená tak podľa prezeraných stránok, ako aj podľa záľub, ktoré definovali pri registrácii [2].

Používateľ je taktiež požiadaný aby ohodnotil danú aplikáciu prostredníctvom dotazníka. Dotazník pozostáva z piatich otázok, ktoré sú zamerané na vizuálne hodnotenie aplikácie, jej prehľadnosť ako aj na samotnú adaptivitu. Výsledky sú štatisticky spracované a uvedené v % v administrátorskom rozhraní.

### 3 Experimenty

Experimenty spočívali v spúšťaní aplikácie AWA zaregistrovanými používateľmi v on-line režime. Po prihlásení používateľa do systému AWA bolo monitorované jeho chovanie počas vyhľadávania stránok na serveri Wikipedia, ktorý sa spúšťal priamo v aplikácii. Url adresy prezeraných stránok sa stiahli aj s odkazmi, upravili sa a boli uložené do databázy. Po načítaní požadovaného dokumentu zo servera bol tento najprv parsovaný a až potom ukladaný do databázy. Pred parsovaním bol každý dokument doplnený o XML špecifikáciu [2].

Zhlukovací algoritmus K-means bol testovaný pre rôzne hodnoty K. V Tabuľke 1. sú uvedené získané výsledky pre  $k = 6$ . Celkový počet prezretých stránok pri tomto experimente bol 160 aj s opakujúcimi sa stránkami.

**Tabuľka 1.** Popis získaných zhlukov pomocou K-means pre  $K=6$ .

Zhluk č.	Kľúčové slová	Nesprávne zaradené stránky	Celkový počet (bez opakovania)	Úspešnosť v %
1	G8, G8 nations, Summit, United States, Japan	4	11	63,6
2	Pluto, Space shuttle, Eris (dwarf planet)	0	2	100
3	Processor, Audio processor, Processing	2	8	75
4	Bob, Jana Kirschner, 1978,1996	2	2	0
5	Motor racing, Motorcycle, Racing	2	6	66,7
6	Wiki, Protection Policy, Coffe	34	67	49,2

### 4 Záver

V rámci riešenia problému znižovania kognitívnej záťaže Internetu bola navrhnutá a implementovaná webová aplikácia AWA, ktorá odporúča webové stránky používateľovi na základe sledovania histórie jeho prístupov na Internet. Systém AWA využíva aj informácie o záujmoch používateľa získané od používateľa pri prihlásení. V počiatočných spolupráce s používateľom nemá systém AWA k dispozícii informácie o (mnohých) predchádzajúcich prístupoch používateľa na Internet, preto si musí jeho profil zostaviť z informácií získaných priamo od neho. Neskôr môže tento profil

korigovať a modifikovať na základe jeho reálneho chovania sa pri prehľadávaní Internetu. Testy ukázali, že pri menšom počte zhlukov je väčšie množstvo zle zaradených stránok. Pre budúcnosť sa črtajú aj možnosti použitia tohto systému vo vzdelávacom procese [1], [5].

Tento príspevok vznikol s podporou VEGA grantu MŠ SR č. 1/4074/07 „Metódy anotovania, vyhľadávania, tvorby a sprístupňovania znalostí s využitím metadát pre sémantický popis znalostí“.

## Referencie

1. Babiuch, M.: The Usage of Laboratory for Mobile, Wireless and Virtual Computer at the Education. In *Proceedings of XXXII. Seminary ASR '07 "Instruments and Control"*. Ostrava: Katedra ATR, VŠB-TU Ostrava, 27. 4. 2007, s. 9-13. ISBN 978-80-248-1272-4..
2. Harman, J.: *Súťaživý princíp strojového učenia a zhlukovanie v podpore používateľa internetu*. Diplomová práca. Technická univerzita, Košice. 2007, 50s.
3. Kanungo, T. at all.: An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24, 2002, 881-892.
4. Machová, K.: An Application of Machine Learning for Internet Users. In. *Emerging Solution for Future Manufacturing Systems*. Springer Science+Business Media, New York, USA, 2005, 459-466, ISBN (HB) 0-387-22828-4 / e-ISBN 0-387-22829-2 (BASYS'04, Vienna, AUSTRIA, 27-29 September 2004).
5. Schmotzer M.: Zefektívnenie výučby. *Zborník prednášok a prezentácií: HTML ako nástroj pre tvorbu výučbových prezentácií II*. Katedra matematickej informatiky PF UPJŠ, Košice, Dec 2001, 9-16, PP. KEGA 255/2000

## Annotation:

*The AWA system - a support for adaptive web*

This paper focuses on solving a problem of decrease of cognitive load of Internet users. It describes the system AWA – a system trying to suggest web pages to a user based on the estimation of his/her interests. The interests of users as well as primary groups of web pages are obtained by means of clustering methods. The system AWA has been designed to utilize the server Wikipedia. It is implemented in the JSP environment and works in on-line mode. As a result, it reacts on user behavior in real time. The paper describes experiments and comments them.