

Využití preferencí zájemců při obchodování s nemovitostmi

Radek Strnad (radek.strnad@gmail.com)
Vedoucí práce: RNDr. Michal Kopecký, Ph.D.

Motivace

Internetový trh s realitami je v současné době přesycen prezentacemi různého druhu. Typická realitní kancelář nabízí nemovitosti hned na několika místech. To vede ke značné duplicitě dat a k prodloužení doby při nalézání vhodné nemovitosti zákazníkem.

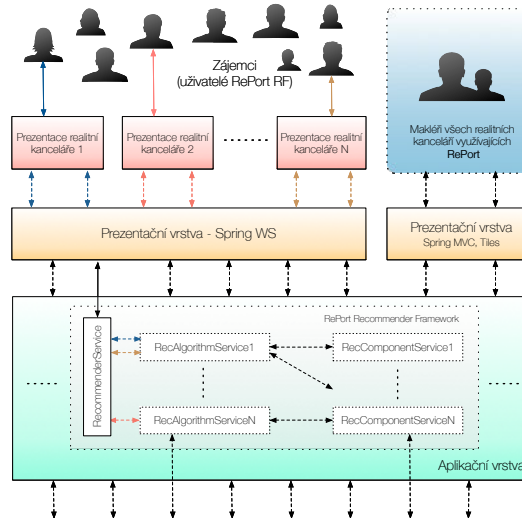
Nasazení doporučovacího systému může být řešením výše uvedených nedostatků. Správně navržený doporučovací systém sníží dobu potřebnou k vyhledání požadované nemovitosti a zvýší komfort uživatele.

Doporučovací systémy

- Sbírají data o předchozím chování uživatelů a na základě analýzy jejich podobnosti s návštěvníkem, dohromady s jeho preferencemi a nastavenými omezeními, doporučí produkty.
- Byly zkoumány algoritmy systémů založených na analýze obsahu, kolaborativní filtrování, hybridní doporučovací systémy a systémy zohledňující kontext.
- Poznatky byly využity k vytvoření obecného frameworku nad informačním systémem pro správu realit, který byl vyvinut v rámci dřívější školní práce (Informační systém RePort).
- V rámci frameworku byl implementován vzorový algoritmus, který byl odladěn nad reálnými daty.

Vlastnosti frameworku pro DS

- Využití existující architektury realitního informačního systému RePort
- Vlastní algoritmus pro každou realitní kancelář
- Logické oddělení jednotlivých algoritmů
- Automatický a transparentní sběr dat
- Definovatelné proměnné
- Automatická identifikace uživatele
- Standardizované rozhraní algoritmů (top-N, setřídění, ohodnocení absolutní oblíbenosti nemovitosti)
- Sdílení funkčních komponent i celých algoritmů
- Napojení algoritmů na události
- Kontext vyhledávání



Vzorový algoritmus

- Vychází z kolaborativního filtrování (memory-based, item-based)
- Content-boosted pro variantu studeného startu (93% návštěvníků nikdy neohodnotí víc jak 5 nemovitostí)
- Nasazen na reálné prezentaci realitní kanceláře střední velikosti
- Malá náročnost na explicitní uživatelské interakce (pouze ohodnocení 1 - 5 hvězdami), zbývající data posbírány z 5 různých implicitních interakcí (čas strávený na stránce, počet otevření detailu, ...)
- Pouze 0,1% návštěvníků ohodnotí nemovitost explicitně. Algoritmus proto musí záviset na implicitních interakcích, zejména na délce času stráveného prohlížením nabídky, jehož význam je diskutabilní.
- Nebylo využito existujících knihoven, ale veškerý kód výpočetní kód vznikl v rámci práce.
- Kvůli rozumné odezvě prováděny částečné výpočty doporučení a materializace vektoru ohodnocení, na testovacím stroji (Core i5 - 750) dosaženo až 7 vláken s odezvou < 1s, tedy cca 25 000 doporučení / hod.
- Vyhodnocení algoritmu provedeno offline metodou rozdělením dat na učící a testovací množinu a následnou optimalizací vah vektorů uživatelských interakcí
- Naměřená normalizovaná střední hodnota chyby ~ 0,09 - 0,12