

Databázový systém Matylda

Návrh softwarového projektu

Vývojový tým

Předpokládaný počet řešitelů: 5

Vedoucí: Mgr. Martin Nečaský Ph.D.

Motivace

V současné době se mnoho nákupů odehrává v internetových obchodech. Přestože lze touto cestou nakoupit téměř cokoliv, faktem zůstává, že hlavním prodejním místem potravin a drogerie jsou stále kamenné obchody. Pokud by však existovala rozsáhlá a dostatečně detailní databáze těchto produktů napříč prodejními řetězci, pak se otevírá mnoho příležitostí, jak ulehčit a pomoci zákazníkům s nákupy v kamenných obchodech.

Existují servery, např. [1] nebo [2], které se ale zaměřují především na slevy a neposkytují data s dostatečnou mírou detailu. Produkty zde lze vyhledávat maximálně podle názvu, ale už ne podle např. složení, historie cen, místa nákupu, výrobce, země původu atd. Kdyby byly takovéto informace k dispozici online, daly by se použít k podrobnému vyhledávání zboží a výrazně tak optimalizovat běžné nákupy.

Cíle projektu

Cílem projektu je návrh a implementace systému, který umožní uživatelům vyhledávat potraviny a drogerii (dále jen produkty) podle libovolné kombinace parametrů. To zahrnuje především vývoj databázového systému, který bude:

- škálovatelný (je třeba počítat s jednotkami milionů produktů a desítkami tisíc uživatelů)
- rozšiřitelný (jednotlivé komponenty systému by měly být maximálně nezávislé a komunikovat pouze přes dohodnuté rozhraní)
- dostatečně robustní (musí být možné uložit jakýkoli produkt z oblasti potravin a drogerie)
- rychlý (desítky vyhledávacích dotazů za sekundu)

Dalším úkolem bude vývoj webové aplikace, která bude zprostředkovávat služby databáze uživatelům. Nyní bude následovat podrobnější specifikace jednotlivých částí.

Požadavky na databázový systém Systém musí být natolik dynamický, aby dokázal uložit libovolný produkt beze změny schématu, přičemž pro každý produkt platí následující:

- Produkty mají jisté vlastnosti společné: ID, název, kategorie, cena, historie cen, dovozce, výrobce, prodejce
- V jiných vlastnostech se můžou lišit: složení
- Produkt může mít několik variant (např. stejné máslo lze koupit v balení po 125g nebo 250g)
- Produkty jsou strukturovány do kategorií (které tvoří orientovaný acyklický graf)

Při reprezentaci produktů bude nutné počítat s velkou variabilitou dat. Např. máslo má určitě jiné složení než salám nebo olivový olej. V databázovém systému půjdou také ukládat slevy, které se mohou týkat jednoho i více produktů (princip dárkových košů).

Dále lze do databáze uložit uživatele a jejich nákupní košíky. Nákupní košík bude seznam produktů (resp. jejich variant), které uživatel nakupuje pravidelně. V tomto košíku bude moci sledovat aktuální ceny a případně slevy týkající se jednotlivých položek.

Databázový systém bude "read-intensive", předpokládá se tedy řádově více vyhledávacích dotazů než aktualizací. Aktualizační dotazy provádí služba pro import dat (viz níže) a proto se těmto dotazům dá přiřadit menší priorita. Kvůli potenciálnímu širokému využití bude nutné, aby šly na systém pokládat nejrůznější vyhledávací dotazy. Tím je myšleno zejména vyhledávání produktů podle názvu, ceny, místa nákupu a slevových akcí, např.:

- Najdi vepřovou šunku s obsahem masa alespoň 80 %, českého výrobce a bez E123.
- Kde koupím v Praze nejlevněji nízkotučné mléko Madeta?
- Kde v Praze nakoupím nejlevněji celý svůj nákupní košík?

Požadavky na import dat Kvůli efektivnímu plnění databázového systému bude implementována následující funkcionality:

- Systém automatického hromadného importu dat z různých zdrojů (zpočátku XML feed s možností rozšířit o další formáty)
- Samostatná aplikace pro ruční zadávání jednotlivých výrobků a slev

Požadavky na webovou aplikaci Bude vyvinuta webová aplikace, která bude využívat služeb databáze a spravovat uživatele. Budou rozlišení přihlášení a nepřihlášení uživatelé.

Nepřihlášení uživatelé budou moci vyhledávat produkty podle libovolné kombinace parametrů. U každého produktu budou zřetelné klíčové informace jako místo prodeje, cena, složení a historie cen produktu. Dále bude možné prohledávat slevy.

Přihlášení uživatelé si budou moci navíc sestavovat vlastní košíky a získávat tak notifikace o slevách, produktech vyhledávaných jinými uživateli atd.

Další požadavky

Z důvodu absence pevného schématu u složení výrobků by mělo být zváženo použití nějakého typu NoSQL databáze [3]. Při plném provozu musí databázový systém zvládnout množství produktů v řádech jednotek milionů, přičemž produkt může mít až několik stovek atributů. Databázový systém musí umožňovat dostatečně rychlé vyhledávání, tzn. dotaz musí trvat maximálně v řádech sekund. Dále musí být schopen zpracovat desítky dotazů za jednu minutu. Systém půjde zprovoznit na PC s OS Windows.

Architektura databázového systému bude poskytovat API, které bude zpřístupňovat funkce systému a bude moci být potenciálně využíváno dalšími webovými (mobilními) aplikacemi.

Webová aplikace bude hostovatelná na IIS, bude uživatelsky přívětivá a nebude vyžadovat instalaci žádného speciálního softwaru ze strany uživatele.

Předpokládaný průběh práce

Projekt je dimenzován pro 5 lidí na 9 měsíců. Práce bude probíhat podle následujícího schématu:

1. V prvním kroku budou analyzovány a otestovány existující technologie a přístupy potřebné k implementaci databázového systému. To zahrnuje návrh datového modelu, otestování modelu na několika databázích s testovacími daty a výběr nejvhodnějšího řešení.
2. V dalším kroku proběhne podrobná specifikace funkcí databázového systému, architektury, rozhraní mezi jednotlivými moduly a webové aplikace.
3. Během implementace budou průběžně vytvářeny automatické testy.
4. Bude vytvořena uživatelská i programátorská dokumentace.

Reference

- [1] www.akcniceny.cz
- [2] www.kupi.cz
- [3] <http://nosql-database.org>