

# Personalized search: Graph vs. OLAP Database

## Onderzoeksvoorstel Bachelorproef 2019-2020

Shauni Van de Velde<sup>1</sup>

### Samenvatting

**Context** De zoekfunctie van de meeste e-commerce websites en platformen is beperkt tot de productcatalogus. Rekeninghoudend met de algemene trend rond personalisering van de customer experience zou het ook aangewezen zijn om ook de search op de e-commerce site te personaliseren. In dit onderzoek wordt bekeken welke databasetechnologie hier het meest geschikt voor is.; **Nood** Om de user experience te garanderen, is het belangrijk dat de meest performante technologie gekozen wordt, hiermee beperken we laadtijden.; **Taak** In dit onderzoek worden twee databanktechnologieën vergeleken, het besluit hiervan zal aantonen welke technologie de beste keuze is voor Personalized Search **Object** In dit document staat het hele proces, de resultaten en de analyse beschreven; **Resultaat** Er wordt verwacht een duidelijk antwoord te kunnen bieden op de vraag over welke technologie het meest gepast is voor Personalized Search **Conclusie** Er zal ofwel een duidelijk antwoord zijn, dit wil zeggen dat één van de technologieën beter presteert dan de andere op elk gebied, ofwel zal er moeten afgewogen worden wat de gebruiker het belangrijkste vindt; **Perspectief** Toekomstgewijs is dit een goede investering van tijd en middelen. Dit is een zeer courant onderwerp dat voor vele e-commerce websites van belang is.

### Sleutelwoorden

Databanken en big data. Graph — OLAP — Personalized Search

### Co-promotor

Nicolas Lierman<sup>2</sup> (Multiminds)

Contact: <sup>1</sup> shauni.vandevelde@student.hogent.be; <sup>2</sup> nicolas.lierman@multiminds.eu;

### Inhoudsopgave

1	Introductie	1
2	State-of-the-art	1
3	Methodologie	2
4	Verwachte resultaten	2
5	Verwachte conclusies	2
	Referenties	2

### 1. Introductie

De meeste e-commerce online platformen hebben reeds een zoekfunctie ingebouwd, maar deze is vaak beperkt tot de enkel de productcatalogus. Rekening houdend met de algemene trend rond personalisering van de customer experience zou het ook aangewezen zijn om ook de search op de e-commerce site te personaliseren. Dit zou resulteren in een betere experience voor de klant en een hogere conversie voor het bedrijf. In deze bachelorproef wordt onderzocht in hoeverre de persoonlijke data van een specifieke gebruiker en contextuele data kan ingeschakeld en gecombineerd worden met de data uit de productcatalogus om persoonlijke en relevante zoekresultaten te genereren. Met deze persoonlijke data wordt bedoeld de historische aankoopdata van de persoon zelf, maar ook bepaalde demografische data zoals geslacht, leeftijd, gezinssamenstelling, etc. De onderzoeksvraag kan dus als volgt geformuleerd

worden: Welk databankmodel presteert het best om gepersonaliseerde zoekresultaten te bekomen?

### 2. State-of-the-art

Personalized Search (Pitkow e.a., 2002) verwijst naar het zoeken op het web waarbij de resultaten afhankelijk zijn van de interesses en voorkeuren van de gebruiker die verder gaan dan de query zelf.

Personalisatie in e-commerce toepassingen biedt een groot voordeel aan bedrijven. De loyaliteit van klanten wordt veel sterker als deze gebruik maken van gepersonaliseerde features. (Telang & Mukhopadhyay, 2005) Een zoekfunctie is een voorbeeld van zo een feature. Recommender Systems (Resnick & Varian, 1997) zijn aanbevelingen vanuit het systeem die rekening houden met de beschikbare informatie van gebruikers en hun voorkeuren om zo een filter te plaatsen op de informatie die weergegeven wordt.

De studie van Diehl (2003) onderzocht het effect van gepersonaliseerde zoekresultaten op de kwaliteit van keuzes die klanten maken, en vond een positieve correlatie. De studie ontdekte dat het verlagen van search cost (Smith, Venkatraman & Dholakia, 1999) leidde tot minder kwaliteitsvolle keuzes. De reden daarvoor is dat klanten slechtere beslissingen maken als de search cost lager ligt omdat zij minder ideale opties aangeboden krijgen. Personalized Search en Recommender System zorgen voor een enorme verbetering in de kwaliteit van de keuzes die de klant maakt, en verminderen het aantal producten die deze klant bekijkt

Maak - OK

2x keuk:  
1x ...  
2x ...

relaties  
correlaties

meta-data

Databank!  
Verbond?

OK

Quid AE?



alvorens hij/zij gevonden heeft wat hij/zij nodig heeft.

Een gevolg van gepersonaliseerde zoekopdrachten is dat we een Filter Bubble (Pariser, 2011) creëren. Het verlaagt de kans dat nieuwe informatie gevonden wordt doordat de resultaten van een zoekopdracht partijdig zijn en eerder wijzen naar dingen die de gebruiker reeds gezien heeft. Dit concept wordt een Filter Bubble genoemd omdat gebruikers eigenlijk geïsoleerd worden in hun eigen wereldje, waar ze enkel de informatie te zien krijgen die ze willen zien. Als we deze gebruikers met hun bubbels in groepen opdelen, verkrijgen we wel het probleem dat deze een vertekend beeld op de realiteit krijgen, zij krijgen bijvoorbeeld in het nieuws slechts het deel te zien dat voor hun interessant is. Als we deze lijn doortrekken naar de klanten van e-commerce websites, zullen zij ook slechts de merken te zien krijgen waar hun voorkeur naar uit gaat. Hierdoor verminder je de kans dat ze een nieuw merk ontdekken of een ander product uitproberen. Als we terug refereren naar de studie van Diehl (2003), dan kunnen we afleiden dat dit een positief effect zal hebben op de klanttevredenheid.

### 3. Methodologie

Om de onderzoeksvraag te beantwoorden wordt er een simple webapplicatie opgezet waar een zoekterm kan ingevoerd worden. Ook worden er twee databankmodellen ontworpen, een model dat gebruik maakt van Neo4j (Graph databank platform), en een model dat gebruik maakt van Elasticsearch (OLAP databank platform) en Kibana om de resultaten te visualiseren. Deze zullen elk met hun eigen API communiceren, en in beide modellen wordt dezelfde gebruiker- en productdata ingevoerd. Bij het opstellen van de modellen wordt mogelijk al duidelijk of één van de modellen niet in staat zal zijn om dezelfde functionaliteiten te hebben als het andere, en dan zal moeten afgewogen worden of de voordelen van het ene model opwegen tegenover het andere model. Als beide modellen dezelfde functionaliteit kunnen bereiken, wordt er bekeken welke het meest performante is. Mogelijks zou het ook haalbaar zijn om beide manieren te combineren op voorwaarde dat de responstijd binnen de acceptabele norm valt.

### 4. Verwachte resultaten

Er wordt verwacht dat er ofwel een duidelijk verschil merkbaar is in performantie tussen de twee verschillende modellen. Mogelijk is dat één van de twee modellen totaal niet haalbaar is om efficiënt een link mee te leggen tussen bijvoorbeeld familieleden, in dit geval bekijken we of het mogelijk is deze twee modellen samen uit te voeren, als dit een resultaat biedt dat binnen de norm valt qua performantie, dan is dit ook een mogelijke oplossing. Het kan zich ook voordoen dat beide modellen vrij performant en efficiënt de queries kunnen verwerken, in dit geval worden de voor- en nadelen alsook de moeilijkheidsgraad van implementatie afgewogen.

### 5. Verwachte conclusies

Er wordt verwacht dat het Graph model makkelijker en beter zal presteren als we rekening willen houden met het

aankoopgedrag van vrienden en familie. Indien dit ook mogelijk is bij een OLAP-model, verwachten we dat Graph nog steeds beter zal presteren. Indien we hier geen rekening mee houden zal het OLAP-model beter presteren, aangezien dit model aangepast is aan grote hoeveelheden data waar complexe zoekopdrachten op kunnen uitgevoerd worden.

### Referenties

- Diehl, K. (2003). Personalization and decision support tools: Effects on search and consumer decision making. *ADVANCES IN CONSUMER RESEARCH*, VOL 30, 30, 166-167.
- Pariser, E. (2011). *The Filter Bubble: What the Internet Is Hiding from You*. Penguin Group, The.
- Pitkow, J., Schütze, H., Cass, T., Cooley, R., Turnbull, D., Edmonds, A., ... Breuel, T. (2002). Personalized search. *Communications of the ACM*, 45(9). doi:10.1145/567498.567526
- Resnick, P. & Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, 40(3), 56-58. doi:10.1145/245108.245121
- Smith, G. E., Venkatraman, M. P. & Dholakia, R. R. (1999). Diagnosing the search cost effect: Waiting time and the moderating impact of prior category knowledge. *Journal of Economic Psychology*, 20(3), 285-314. doi:10.1016/s0167-4870(99)00010-0
- Telang, R. & Mukhopadhyay, T. (2005). Drivers of Web portal use. *Electronic Commerce Research and Applications*, 4(1), 49-65. doi:10.1016/j.elerap.2004.10.004

Dit goed uitleggen in de BP!

Niet te simpel te maken en te standaard

waarom gr leg uit wat OLTP? API - OLAP (data warehouse) OLAP is?

reference cases? vgl. tabellen met zegt de markt? indicatoren vld data + ETL + DWHoutlet. - invloed van markt? - invloed van kennis + documentatie