



Universiteit  
Leiden  
The Netherlands

## Enterprise Search is geen Google

Feenstra, R.J.

### Citation

Feenstra, R. J. (2010). Enterprise Search is geen Google. *Automatisering Gids*. Retrieved from <https://hdl.handle.net/1887/111492>

Version: Not Applicable (or Unknown)

License: [Leiden University Non-exclusive license](#)

Downloaded from: <https://hdl.handle.net/1887/111492>

**Note:** To cite this publication please use the final published version (if applicable).

# Enterprise Search is geen Google

16 APRIL 2010

De aanschaf van een Enterprise Search-systeem begint met een keuze tussen federated search en integrated search. Alle zoek-varianten, zegt Rob Feenstra, voorzien in een behoefte, maar er is niet één systeem dat binnen afzienbare tijd alle digitale bronnen ontsluit. Een misverstand dat nog steeds bestaat.

Het moet nog maar eens gezegd: Enterprise Search is geen Google voor bedrijven. De inrichting van een Enterprise Search-systeem hangt af van de aard van de informatievraag, de omvang en de homogeniteit/heterogeniteit van de gegevens, de openbaarheid van de informatie enzovoorts.

Wie een Enterprise Search-systeem wil aanschaffen, moet zich er dus rekenschap van geven dat hij vooraf veel keuzes moet maken. Een heel elementaire keus daarbij is die voor federated search of integrated search. Laten we eens kijken hoe die keuze is uitgevallen in een wereld waarin het zoeken naar informatie van wezenlijk belang is: die van de universiteitsbibliotheken. Het aardige daarvan is dat het gaat om instellingen waarvan de taak en het informatieaanbod voor een groot deel overeenkomen. Niettemin zijn er verschillen in de gekozen oplossingen.

**Toen het aantal digitale wetenschappelijke bronnen begon toe te nemen**, zagen de bibliotheken zich geplaatst voor de taak om vereenvoudiging aan te brengen in het aantal zoekingen. Aanvankelijk gebeurde dat met de gedachte om alle bronnen via één systeem te ontsluiten. Maar toen er een paar jaar later knopen werden doorgehakt, bleek dat ideaalbeeld een illusie. De Universiteit Utrecht koos voor integrated search. Er werd een systeem ontwikkeld op basis van Autonomy waarmee een groot deel van de

tijdschriften (maar niet alle) via één enkele index ontsloten kon worden. Het resultaat was een zoekfunctie die op dit moment bijna 20 miljoen artikelen ontsluit.

Een groot aantal andere universiteiten koos voor een commercieel federated search-systeem. De voordelen: het was aanmerkelijk goedkoper, de gebruiker had meer invloed op de keuze van de bronnen die hij wilde doorzoeken en het was minder onderhoudsgevoelig. Bovendien was er de mogelijkheid om de zoekresultaten gemerged of per database te bekijken. Nadelen waren er ook. Het belangrijkste nadeel was de snelheid, of beter, het gebrek daaraan. Gebruikers vergeleken het systeem wat dat betreft met Google en keken niet naar de tijd die het kost om verschillende databases separaat te doorzoeken. Een nadeel dat direct samenhang met het gebrek aan snelheid was het beperkte aantal databases dat gelijktijdig kon worden ontsloten. Enkele andere problemen waren ontdubbeling en de gebrekkige relevance ranking.

Een interessante variant werd ontwikkeld in Groningen, waar de resultaten van zoekacties via het federated search-systeem worden geïndexeerd en gebruikt om bij latere zoekacties te kunnen bepalen welke bronbestanden het beste doorzocht kunnen worden.

Inmiddels zijn een paar universiteiten overgegaan op integrated search-systemen die zijn gebouwd met open-sourcesoftware. Uiteraard veel goedkoper in aanschaf, maar duurder in ontwikkeling. De leverancier van het federated search-systeem heeft al een paar jaar een product op de markt gebracht dat federated en indexed search combineert, maar de ontwikkeling daarvan verloopt traag en tot nu toe hebben de Nederlandse universiteitsbibliotheken zich er nog niet aan gewaagd. De laatste ontwikkeling is die waarbij de leverancier van de software ook de metadata levert. De

Erasmus Universiteit heeft onlangs voor een dergelijke oplossing gekozen.

Alle genoemde zoekvarianten voorzien weliswaar in een behoefte, maar met name bij wetenschappers is er nog altijd een grote behoefte om rechtstreeks in de databases te zoeken. Ten eerste omdat die vaak toch betere en/of meer op de specifieke inhoud toegespitste zoekmogelijkheden bieden en omdat de experts wel weten waar ze moeten zoeken en geen behoefte hebben aan een grote hoeveelheid voor hen zinloze zoekresultaten. Bovendien is bij het direct zoeken in de bronbestanden de kans groter dat de wetenschapper stuit op een resultaat waar hij niet naar op zoek was, maar dat niettemin heel nuttig is (serendipiteit).

Bij geen van de universiteitsbibliotheken is het mogelijk gebleken om via één zoekvak of één systeem binnen afzienbare tijd alle digitale bronnen te ontsluiten. Het misverstand dat zoiets mogelijk is in de eigen omgeving leeft nog steeds volop bij veel (potentiële) gebruikers van Enterprise Search-systemen. Het is belangrijk om je daar bewust van te zijn.

Een beperkende factor die hierboven nog niet is genoemd is beveiliging. Je kunt je niet alleen richten op het ophalen van data uit andere bestanden, maar moet je ook buigen over de vraag wie welke data mag benaderen. Federation kan daarbij geschikter zijn dan integrated search omdat er in het eerste geval direct gebruikgemaakt kan worden van de authentication- en authorisationstructuur van de desbetreffende bron, door met de query de user/password-gegevens mee te sturen. Ook prijsoverwegingen kunnen een rol spelen bij de keuze voor federated search. Als er grote hoeveelheden gegevens moeten worden ontsloten, dan kan indexering duur zijn. Het is daarbij belangrijk om niet alleen te kijken naar het moment van aanschaf maar ook om in kaart te brengen welke toekomstverwachtingen er zijn. Wanneer dit

onvoldoende in ogenschouw wordt genomen (en dat komt nogal eens voor), dan kan het gevolg zijn dat er een nieuwe, aanmerkelijk duurder licentie moet worden aangeschaft.

Uit financieel en beheerstechnisch oogpunt lijkt er dus wat te zeggen voor federated search. Echter, wanneer die keuze wordt gemaakt, wordt er nogal eens voorbijgegaan aan een aantal zaken. We noemden hierboven al snelheid, ontdebbling en relevance ranking, maar er is meer. Bij federated search moet gebruikgemaakt worden van de grootste gemene deler als het gaat om het vastleggen van zoektermen. Logische operatoren worden niet in elke database op dezelfde wijze gebruikt. Het gebruik van taxonomieën en thesauri is veel lastiger bij federated search dan bij indexed search en datzelfde geldt voor allerlei presentatie-elementen als clustering en faceting en andere zaken die steeds belangrijker worden naarmate het informatieaanbod toeneemt.

**De keuze tussen federated search, integrated search of een combinatie daarvan** is maar een van de vele die gemaakt moet worden als een organisatie overweegt een Enterprise Search-systeem aan te schaffen. Denk bijvoorbeeld aan de beslissing om alleen metadata te indexeren of ook de tekst van documenten. De vraag moet worden beantwoord hoelang je de geïndexeerde data wilt bewaren (ongebreedeld bewaren kan tot gigantische, en kostbare, indexen leiden). Schaf je iets aan dat voldoet aan de huidige eisen of anticipeer je op toekomstige ontwikkelingen? Het moge duidelijk zijn dat het bij die overwegingen gaat om factoren van verschillende aard (technisch, functioneel, organisatorisch, financieel enzovoorts). Keuzes kunnen allerlei kanten uitgaan en het ideale resultaat zal ongetwijfeld niet bereikt worden, maar om tot een goed resultaat te komen, is het in ieder geval noodzakelijk dat alle factoren in kaart worden gebracht en in

ogenschouw worden genomen en dat alle belanghebbende partijen hierbij worden betrokken.

*Rob Feenstra* is adviseur bij M&I/Partners.