

Tigran Spaan en
Mart Trautwein

Beter zoeken in een open architectuur

VEEL ORGANISATIES WERKEN MET VERSCHILLENDE APPLICATIES DIE ONDERLING NIET OF BEPERKT INFORMATIE KUNNEN UITWISSELEN. DAT LEVERT BEPERKINGEN OP ALS HET GAAT OM ZOEKEN EN VINDEN VAN GEGEVENS.

EEN OPEN ARCHITECTUUR ZORGT VOOR TOEKOMSTBESTENDIGE APPLICATIES DIE DOORZOEKBAAR ZIJN MET ÉÉN ZOEKMACHINE.

Stel, uw organisatie mag meedoen met een samenwerkingsportal. U bereidt alles goed voor, u verbetert de collectieregistratie, u maakt selecties, bedenkt voorbeelden, doet mee in een klankbordgroep. Iedereen is enthousiast en er is alleen nog een technisch stapje nodig. En dan blijkt, dat u helemaal niet kunt meedoen, want het programma dat u gebruikt, voldoet niet aan de afgesproken standaarden. U kunt uw informatie dus niet aanbieden aan het portal. En in de roadmap van uw leverancier komt de module die dit oplost, pas over twee jaar beschikbaar.

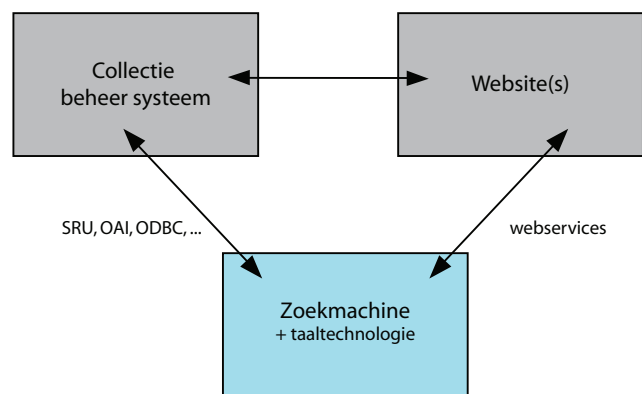
U zit in een *vendor lock-in*: u bent afhankelijk geworden van uw IT-leverancier en zijn pakket. En u bent niet de enige. Weet u nog hoe vijf tot tien jaar terug grote bedrijven al hun informatie en administratie in één systeem gingen onderbrengen? Veel organisaties proberen daar nu weer vanaf te komen.

De oplossing voor dit probleem is een *open architectuur*. Met een open architectuur kunt u voor iedere toepassing het beste programma kiezen, kunt u meedoen aan alle samenwerkingsportals en hebt u geen verdubbeling van toepassingen. Weet u op hoeveel plaatsen in een gemiddeld bedrijf adresgegevens worden opgeslagen, vaak op totaal verschillende manieren maar wel met veel overlap? In dit artikel geven wij u praktische handvatten om te komen tot een open architectuur. We doen dit aan de hand van een concreet voorbeeld, namelijk zoekmachines.

Beter zoeken in een open architectuur

Bedrijven en organisaties hebben, vaak zonder zich ervan bewust te zijn, vele verschillende zoekmachines in huis: voor de website(s), het intranet, voor email, de desktop-omgeving, het DMS, databases, bibliotheek, archief, HRM-systeem, CRM-systeem, beeldbank, wiki's en het collectieregistratiesysteem. Dat is niet handig voor eindgebruikers. Deze moeten op al die verschillende manieren zoeken in de verschillende bronnen. Bovendien is het niet efficiënt. Als u één van de zoek-

machines verbetert, moet u hetzelfde ook nog eens doen voor de andere. De oplossing voor deze problemen is gebruikmaken van een open architectuur.



Figuur 1. Open architectuur en de zoekmachine

Binnen een open architectuur is het mogelijk om het zoeken in te richten als onderdeel van uw IT-architectuur. Deze manier van denken is al jaren standaard bij grote organisaties. Doordat er steeds meer goede betaalbare zoekoplossingen zijn, kiezen nu ook kleinere en middelgrote organisaties voor deze open architectuur voor zoekmachines. Binnen organisaties is dit concept bekend als *Enterprise Search*. Datzelfde open architectuur is ook van toepassing voor zoekmachines buiten de eigen organisatie, bijvoorbeeld portals.

Open architectuur, open standaarden, open source. Informatici praten graag over verschillende openheidsbegrippen, maar wat bedoelen ze nu eigenlijk?

Open source-programma's zijn programma's waarvan de broncode openbaar is. Als u in een project open source software gebruikt, kan een andere leverancier de code dus ook lezen. Daardoor kan hij het project beheren of overnemen als uw huidige leverancier dat niet kan of als u van leverancier wilt wisselen. Lagere kosten kunnen een bijkomend



voordeel zijn van open source. Het is belangrijk dat achter een open source-programma zich een grote community bevindt – aan een open source-product dat door één leverancier wordt ondersteund, hebt u niets.

Open architectuur betekent dat u voor iedere toepassing een specifieke oplossing hebt waarbij al deze programma's gegevens met elkaar kunnen uitwisselen. Ieder programma voldoet aan open standaarden en is daardoor vervangbaar zonder dat de rest van uw applicaties op de schop moet. Dat heeft als voordelen dat u voor iedere toepassing het beste programma kunt kiezen en dat u nooit afhankelijk wordt van de leverancier van een programma. De leverancier van het programma moet dit garanderen. Bovendien kunt u ge-

Open standaarden van het juiste niveau

Er zijn veel onafhankelijke organisaties die voorlichting geven over het gebruik van open standaarden, bijvoorbeeld DEN (www.den.nl) en NoiV (www.noiv.nl). Daardoor heeft iedereen inmiddels wel een beeld van wat open standaarden zijn. Maar niet alle open standaarden stellen u in staat om tot een open architectuur te komen. XML wordt vaak genoemd als om een voorbeeld van een open standaard wordt gevraagd, maar juist XML is niet voldoende voor het bereiken van een open architectuur. XML is een eerste stap op weg naar een open architectuur. In XML is de *vorm* van informatie vastgelegd, maar er wordt niets gezegd over structuur, inhoud en informatie-uitwisseling.

Standaarden kunnen, enigszins schematisch en versimpeld, als volgt worden ingedeeld:

Niveau 1 Afspraken over *vorm*; bijvoorbeeld: XML, SQL, HTML, RDF, ODF,...

Niveau 2 Afspraken over vorm én *uitwisseling*; bijvoorbeeld: http, SOAP, REST,...

Niveau 3 Afspraken over vorm en/of uitwisseling én veldbenamingen en *structuur*; bijvoorbeeld: DC, SKOS, SRU, OAI-PMH, EAD, GedCom,...

Niveau 4 Afspraken over (een deel van) het bovenstaande plus de *inhoud* van velden, bijvoorbeeld MARC21, ISAD(G),...

makkelijk nieuwe programma's koppelen aan uw bestaande open architectuur. U voorkomt verdubbeling van functies, zoals adresbeheer op vijf plaatsen.

Open standaarden zijn afspraken over vorm, uitwisseling, structuur, velden en invulling van gegevens. Open standaarden zijn een noodzakelijke voorwaarde om te komen tot een open architectuur. Bovendien zijn open standaarden nodig om mee te doen aan samenwerkingsplatforms en portals. Een open architectuur is mogelijk als programma's voldoen aan open standaarden op niveau 3 of 4. Ondanks de voordelen van een open architectuur kiezen veel organisaties voor het doen van zaken met één leverancier. Dit komt vaak

voor in bijvoorbeeld de erfgoedsector, waar veel organisaties klein van omvang zijn. Deze organisaties krijgen van een leverancier het aanbod voor een collectieregistratiesysteem of bibliotheekstelsel inclusief website en een zoektoepassing in één totaalpakket. Dat lijkt overzichtelijk. Toch zien veel organisaties de potentiële problemen die deze keuze met zich meebrengt. De afhankelijkheid van de leverancier is groot, en als één onderdeel van het pakket niet bevalt, dan kan dat niet worden vervangen door iets beters. Soms is ook deelname aan samenwerkingsportals problematisch.

Open source in de erfgoedsector

De afgelopen jaren zijn er veel goede open sourcepakketten opgekomen waar de erfgoedsector zijn voordeel mee kan doen. Daarbij zijn de kosten voor licenties, hosting en beheer vaak aanzienlijk lager dan bij gesloten totaalpakketten en is er sprake van keuzevrijheid en een hogere kwaliteit van losse programma's.

Koha (www.koha.org) is een wijdverbreid open source bibliotheekstelsel dat in Nederland door onder meer het Rijksmuseum is gekozen voor de bibliotheek. Koha ondersteunt standaard MARC21 en SRU. Voor archiefsystemen zijn **Collective Access** (www.collectiveaccess.org) en **ICA AtOM** (www.ica-atom.org) goede keuzes; de laatste ondersteunt standaard EAD en ISAD(G), de eerste zal dit binnen afzienbare tijd ook doen.

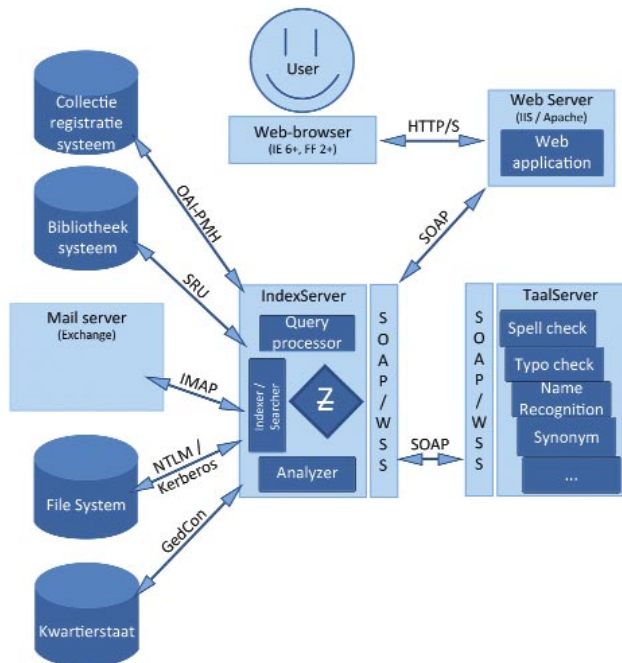
Bij het *zoeken* ligt het iets subtieler. **Lucene** (lucene.apache.org) is zelf geen zoekmachine maar een zeer hoogwaardige verzameling modules waarmee een zoekmachine kan worden gebouwd. Op Lucene zijn volledige zoekmachinepakketten gebouwd zoals Attivio AIE en ons eigen pakket, de GridWalker Zoekmachine voor het Nederlands. Maar ook maatwerkoplossingen zijn goed mogelijk met Lucene.

Trends in zoeken

Vijftien jaar geleden stond zoeken gelijk aan bibliotheekzoeken. Je moest een formulier invullen met veel velden en kreeg dan het juiste antwoord. De tweede generatie zoekmachines die volgde, was van het type *Google*: één veld invullen en als resultaat veel goede en minder goede zoekresultaten. De derde generatie zoekmachines – op dit moment veel voorkomend – heeft een interface met in eerste instantie één zoekveld, waarna door klikken op zogeheten *navigatoren* het zoekresultaat wordt verkleind. Navigatoren, ook wel facetten genoemd, zijn lijsten met velden als auteur, datum, trefwoorden, categorieën, waarop kan worden doorgeklikt. Derde generatie zoekmachines zien er vaak niet uit als zoekmachine maar als een gewone website. Denk bijvoorbeeld aan LinkedIn en YouTube.

De vierde generatie zoekmachines maakt gebruik van kennis van *taal en semantiek*, zoals ondersteuning voor spelling,

typo's, historische spellingswijzigingen, morfologie, fonologie, samenstellingen, woordgroepen, verbanden, naamherkenning, synoniemen, afkortingen, thesauri, enzovoorts.



Figuur 2. Architectuur vierde generatie zoekmachines met taalondersteuning

Kennis van taal bij het zoeken speelt bij veel processen in organisaties een rol. Denk bijvoorbeeld aan automatische classificatie, het automatisch inlezen van facturen in een betalingssysteem, het matchen van adressenbestanden of het combineren van thesauri. Een recente ontwikkeling is *opinion mining*: een taaltoepassing waarbij bedrijven of politieke partijen kunnen zien hoe er over hen wordt gesproken op internet, in blogs en op fora.

Waar wilt u over vijf jaar staan?

Wat is belangrijk voor u: de beste applicaties voor ieder werkproces, zodat u meer gedaan kunt krijgen? Deelname aan samenwerkingsportals? Eén goede zoekmachine voor al uw collecties, applicaties en informatie in plaats van verschillende zoekmachines voor alles? Het verlagen van kosten van uw IT-architectuur op langere termijn?

Als u kiest voor een open architectuur dan is het belangrijk om nu al de eerste stappen te zetten. Het gaat er vooral om dat u weet waar u naartoe wilt en dat u bij keuzes die u toch al moet maken, open standaarden voorop stelt. Om te komen tot een open architectuur is het van belang dit als doelstelling te definiëren, zodat dit voor iedereen in de organisatie duidelijk is: over vijf jaar hebben wij een open architectuur. Geef er ook de redenen bij waarom u dit belangrijk vindt – dat helpt bij het krijgen van draagvlak.

Bepaal voorts welke toepassingen in uw huidige architec-

tuur verdubbeld zijn. Hier valt winst te halen. Wij noemden hierboven het zoeken en adresbeheer als klassieke voorbeelden, maar wellicht spelen er bij u andere verdubbelingen die u nu echt weg wilt hebben.

Het is belangrijk te bepalen welke programma's u wilt behouden. De reden hiervoor kan zijn dat het een prettig werkend pakket is of dat de prijs/kwaliteitverhouding gunstig is. Het kan daarbij natuurlijk zo zijn dat u een totaalpakket hebt waarvan u een deel wilt behouden en een deel niet. Vraag bij uw leverancier de specificaties van de uitwisselingsstandaarden. Als deze niet op niveau 3 of 4 zijn, dan hebt u nu nog de tijd om hier duidelijke afspraken over te maken met de leverancier, of op zoek te gaan naar alternatieven. U hebt waarschijnlijk al een idee welke programma's u de komende jaren toch al wilde aanschaffen of vervangen. Het enige wat u hier hoeft te doen is dat u het ondersteunen van open standaarden op niveau 3 of 4, bovenaan de lijst met criteria zet.

Roadmap naar open source

De roadmap op weg naar een open architectuur bestaat uit vier groepen programma's:

1. De programma's die u toch al wilde vervangen;
2. de programma's die nu verdubbeld zijn;
3. de programma's die niet voldoen, en
4. de programma's die niet aan open standaarden voldoen.

Maak hiervan een nieuwe lijst en zet bovenaan de lijst die u had en de verdubbelers (groepen 1 en 2) en geef deze zelf een volgorde. Daaronder zet u de programma's die niet meer voldoen (groep 3) en helemaal onderaan de programma's die niet open standaarden ondersteunen op niveau 3 (groep 4). U hebt nu uw nieuwe roadmap voor de komende vijf jaar. Door deze volgorde aan te houden geeft u uw huidige leveranciers bovendien extra tijd om aan de standaarden te voldoen die u van hen vraagt. Misschien is vervanging tegen die tijd dan toch niet nodig. En als u meer tijd nodig hebt dan vijf jaar, dan kan dat natuurlijk ook.

Conclusie

Een open architectuur is belangrijk om verschillende redenen. Eén daarvan is dat het zoeken in een open architectuur in één keer goed kan worden geregeld. Het bereiken van een open architectuur is gemakkelijker dan het lijkt, mits u het gefaseerd aanpakt en een duidelijke roadmap maakt. **IK**

Tigran Spaan is informaticus, logicus en taalkundige en eigenaar/directeur van GridLine BV. Mart Trautwein is senior architect bij GridLine BV en gepromoveerd in de informatica en taaltechnologie. Hij ontwikkelde reeds in de jaren 90 zoekmachines en definieerde de architectuur voor WieWasWie, een genealogisch zoekportaal waaraan 120 archieven meewerken.