

Zoekoplossingen voor de onderneming

Integratie van de zoekfunctie in uw IT- architectuur

Zoekoplossingen voor de onderneming kunnen een zegen zijn voor de bedrijfsproductiviteit en -prestaties, mits deze effectief en efficiënt worden ingezet.

INLEIDING: HET BELANG VAN ARCHITECTUUR

Bedrijfszoekoplossingen ontwikkelen zich steeds meer als krachtige hulpmiddelen die bedrijven in staat stellen een universeel doel te bereiken — beter gebruik maken van informatie.

Multi-tiered zoekoplossingen die de computer, afdelingen, de onderneming en het Internet uitkammen op zoek naar informatie, zullen de productiviteit bevorderen en de effectiviteit binnen de onderneming verhogen. Deze oplossingen bieden een pad naar de oplossing van belangrijke problemen waarmee bedrijven tegenwoordig worden geconfronteerd — overvloedige opslagplaatsen van identieke gegevens, verschillende methoden voor het ophalen van informatie en een gebrek aan consistente semantische definities binnen de onderneming. Ze kunnen ook helpen tegemoet te komen aan de groeiende verwachtingen van werknemers dat bedrijfssystemen even gemakkelijk te doorzoeken en even gebruiksvriendelijk zijn als Internetsoftware en Internetdiensten.

Om te kunnen profiteren van de voordelen van een bedrijfszoekfunctie moet de zoekoplossing effectief geïntegreerd worden in andere technologische bedrijfssystemen. BearingPoint heeft de essentiële principes van een bedrijfszoekfunctie vastgesteld en op basis daarvan een aanbevolen architectuur ontwikkeld voor de succesvolle implementatie ervan.

DE PRINCIPES VAN EEN BEDRIJFSZOEKFUNCTIE

Er gelden verschillende algemene principes voor de ontwikkeling van een zoekoplossingarchitectuur:

Eenvoud. Het werkingsprincipe van een bedrijfszoekfunctie is simpele systemen en oplossingen. Complexiteit remt de productiviteit van softwareontwikkelaars; bemoeilijkt de planning, bouw en het testen van producten; leidt tot beveiligingsproblemen en veroorzaakt frustratie bij de eindgebruiker en beheerder. De computer- en communicatietechnologieën hebben zich dusdanig ontwikkeld dat een servicegeoriënteerd model nu realiseerbaar is. Servicegerichte software en een simpele ontwikkeling van zoekoplossingen vormen het antwoord op de behoefte van de gebruiker aan een onweerstaanbare, geïntegreerde zoekoplossing die 'gewoon werkt'.

Functionaliteit. De prioriteiten van de zoekoplossingarchitectuur zijn (in volgorde van prioriteit): bruikbaar, betrouwbaar, veilig, schaalbaar (wat betreft gebruikers, berichtvolumes en services) en in staat om onafhankelijk te groeien door middel van toevoeging van nieuwe services afkomstig van de zoekoplossingstructuur.

IN DIT PERSPECTIEF:

INLEIDING: HET BELANG VAN ARCHITECTUUR	1
DE PRINCIPES VAN ZOEKEN BINNEN DE ONDERNEMING	1
RICHTLIJNEN VOOR DE ONTWIKKELING VAN OPLOSSINGEN	2
DE ZOEKOPLOSSING-ARCHITECTUUR	3
Conceptuele architectuur	4
Voorbeeld van zoekoplossing-architectuur	4
DE MOGELIJKHEID VAN DE ZOEKOPLOSSING	8

Betrouwbaarheid. Het zoekoplossingsplatform moet een solide infrastructuur zijn. Daarom moeten geavanceerde technologieën worden vermeden. Als vuistregel geldt dat de oplossing in de kernarchitectuur gebruik maakt van een technologie die:

- maximale overdraagbaarheid biedt naar iedere omgeving.
- tot tevredenheid getest is in vergelijkbare gebruiksscenario's elders.

Essentieel onderdeel. Een bedrijfszoekoplossing moet worden beschouwd als een essentieel bedrijfs onderdeel, dat uiterst veilig en schaalbaar is en 7 dagen per week 24 uur per dag beschikbaar is. De zoekoplossingstructuur moet van de grond af met deze kenmerken in gedachten worden opgebouwd.

Toegankelijkheid. Om door velen te kunnen worden gebruikt, moet de architectuur van de zoekoplossing transparant en toegankelijk zijn. De toegankelijkheid moet van het begin af in de architectuur worden ingebouwd, want deze functionaliteit kan niet later worden toegevoegd.

Flexibiliteit. De idee van flexibele services vormt de sleutel tot een succesvolle zoekoplossing. Services niet in de zin van een eindgebruikersservice, maar services als afzonderlijke functie-eenheden, die op het zoekplatform via gedefinieerde interfaces beschikbaar zijn.

RICHTLIJNEN VOOR DE ONTWIKKELING VAN OPLOSSINGEN

Op basis van de hierboven genoemde principes, moet de zoekoplossingstructuur voldoen aan de volgende richtlijnen:

Open standaards, open technologie en algemene interfaces. De zoekoplossing moet zijn gebaseerd op open standaards die technologisch onafhankelijk zijn, door gebruikmaking van standaards als J2EE, .NET en webservices. Zo wordt de grootst mogelijke flexibiliteit verzekerd, wordt voorkomen dat de interface moet worden gewijzigd wanneer een andere toepassingsinterface wordt gewijzigd en wordt de complexiteit van API-connectiviteit tussen verwante systemen en toepassingen op de achtergrond gehouden.

Modulaire architectuur. De zoekoplossing is gebaseerd op een n-tier, componentgebaseerde architectuur. Deze benadering biedt abstractie van elke laag en voor iedere component, waardoor componentonafhankelijkheid en logisch opgedeelde functionaliteit wordt gecreëerd. De architectuur moet op losse wijze zijn gekoppeld en gestructureerd.

Interoperabiliteit. Het gebruik van XML (extensible markup language) XSLT (extensible style sheet language transformation) zorgt voor interoperabiliteit die een abstractie verschaft waarbij het presentatiedomein wordt gescheiden van de onderliggende logica. Door de presentatie onder te verdelen in een presentatie-sublaag en een presentatie-interfacesublaag wordt de presentatielogica geïsoleerd. De architectuur steunt voor haar interoperabiliteit op de onafhankelijkheid van apparatuur, platformen en technologieën.

Metagegevensbeheer. Metagegevensbeheer omvat het maken, opslaan, openen, bewaren en onderhouden van metagegevensinhoud, met inbegrip van het beheer en de synchronisatie van gedeelde metagegevens. Metagegevensbeheer binnen de onderneming biedt een bedrijf de mogelijkheid om:

- de toewijzing van metagegevens aan bestaande inhoud te automatiseren. Dit maakt het zoeken effectiever omdat de juiste inhoud kan worden gevonden ongeacht waar deze zich bevindt.
- de capaciteit te ontwikkelen om metagegevens zo snel mogelijk na de het tijdstip van creatie van inhoud aan deze inhoud te koppelen.
- als bedrijf duidelijk stewardship te verwerven voor de ontwikkeling en het voortdurende beheer van taxonomieën en authority-lijsten.
- processen te implementeren voor het actueel houden van taxonomieën en de controle van links voor metagegevens.
- verschillende weergaven van dezelfde inhoud te krijgen, zoals op basis van rol, tijd of proces.
- inhoudbronnen te integreren — de zoekoplossing moet een open architectuur hebben ter ondersteuning van integratie van meerdere inhoudbronnen. Open standaards zijn vereist voor een soepel functioneren van het platform, vooral wanneer het gaat om de interoperabiliteit van meerdere inhoudbronnen.

Multi-tiered zoekoplossingen kunnen een antwoord bieden op de grote problemen waarmee bedrijven momenteel geconfronteerd worden.

Beveiliging. Een zoekoplossing voor de onderneming moet de beveiligingsprincipes die verbonden zijn met de verschillende inhoudsbronnen binnen de onderneming navolgen en versterken. De principes van verificatie, autorisatie, controle en identiteitsbeheer moeten worden toegepast, opdat uitsluitend bevoegde gebruikers toegang hebben tot vertrouwelijke informatie. Zoekoplossingsbeveiliging moet voldoen aan het beveiligingsbeleid van de onderneming en de vereiste regels. In sommige gevallen is beveiliging op documentniveau vereist voor een effectief gebruik van ongestructureerde inhoud.

Netwerkinfrastructuur. Bij de implementatie van de oplossing moet rekening worden gehouden met het effect op de netwerkinfrastructuur en het bedrijf dat wordt ondersteund door deze infrastructuur. In het geval van een zoekoplossing voor een bedrijf moet er rekening worden gehouden met waar de te doorzoeken gegevens zich bevinden, of mensen en bedrijven buiten de onderneming ook van de zoekfunctie gebruik maken, en met de beveiliging en algehele prestaties.

Overwegingen en richtlijnen met betrekking tot het algemene netwerkontwerp zijn:

- Hoeveel verkeer wordt er gegenereerd op het bedrijfsnetwerk en de daaraan gekoppelde segmenten? Het aantal doorzochte gegevensbronnen, de locatie van de gegevensbronnen en het gebruikte type netwerkstructuur zijn van bepalende invloed op deze ontwerpoverweging.
- Wat is de locatie van de gebruikers van de zoekoplossing, in hetzelfde gebouw, op een ander terrein of binnen een WAN (wide area network)?
- Heeft er netwerkverkeerclassificatie plaatsgevonden? Verkeerclassificatie is een belangrijke overweging in een omgeving waarin gebruik wordt gemaakt van SAN-gegevensreplicatie, zoals Symmetrix Remote Data Facility (SRDF). Het bedrijf dient rekening te houden met het effect van zoeken op het netwerk en moet het juiste netwerksegmentatiebeleid en de servicekwaliteit (QoS) implementeren.

- Wat voor type toegangscontrole is binnen het netwerk aanwezig? Bij veel implementaties wordt de zoekoplossing gebruikt voor toegang tot en het presenteren van bedrijfsgegevens of anderszins vertrouwelijke gegevens. Een juiste implementatie is dan ook van cruciaal belang. Correcte netwerksegmentatie, een VLAN (virtual local area network)-ontwerp, een firewall en netwerkregistratie zijn allemaal zaken die in overweging moeten worden genomen.
- Welk type netwerktolerantie is vereist? Omdat ze wordt ingezet op de financiële afdeling, de personeelsafdeling en in andere beheerfuncties, wordt een zoekoplossing voor de onderneming vaak beschouwd als een belangrijke toepassing, die een hoge beschikbaarheid vereist. Om dit te bereiken, wordt de zoekoplossing vaak geclusterd geïmplementeerd, met single- of multi-site overnameopties.

Noodherstel en bedrijfscontinuïteit. Om te kunnen voldoen aan de principes van noodherstel en bedrijfscontinuïteit is voor het minimaliseren van de single points of failure hardware- en software-redundantie vereist. De algemene ontwerpfilosofie is dat redundante componenten worden gebruikt voor het creëren van een redundante infrastructuur. Hierdoor wordt de fouttolerantie van afzonderlijke componenten verhoogd en de systeembeschikbaarheid voor interne en externe gebruikers en processen verbeterd.

DE ARCHITECTUUR VAN DE ZOEKOPLOSSING

De architectuur van een zoekoplossing voor bedrijven moet robuust zijn en de doelstellingen van de onderneming ondersteunen. In realiteit betekent dit een modulaire, n-tier bedrijfsarchitectuur die voldoet aan de hierboven genoemde standaards en richtlijnen. Deze is in staat een zoekoplossing te ondersteunen die tegemoet komt aan de huidige bedrijfsbehoeften, de groei voor de toekomst stimuleert, de time to market versnelt en de gebruikskosten drukt.

Conceptuele architectuur

In afbeelding 1 wordt de conceptuele architectuur van de zoekoplossing weergegeven.

De componenten van de conceptuele architectuur zijn:

Inhoudsbronnen. Tot de inhoudsbronnen behoren het bedrijfsintranet, websites, bestandsservers, inhoudsbeheersystemen, zoals Documentum en FileNet en bedrijfstoepassingen voor het hele scala van transactionele systemen, waaronder CRM (customer relationship management), SCM (supply chain management) en ERP (enterprise resource planning). In principe vertegenwoordigt dit blok de verschillende bedrijfssystemen die als inhoudsbronnen kunnen dienen die gebruikers mogelijk willen doorzoeken.

Connectiviteitservices. De laag connectiviteitservices levert de algemene adapters voor de onderliggende inhoudsbronnen. Deze laag bevat twee soorten adapters: connectoren en federatieve toegang. Connectoren zijn een uitbouw van de verschillende technologie-adapters die vereist zijn voor verbinding met de inhoudsbronnen. Connectoren voor federatieve toegang verzorgen de verbinding met inhoud van derden.

Metagegevensservices.

De laag metagegevensservices biedt aanvullende services met toegevoegde waarde aan de inhoud. Ze biedt twee typen services: metagegevens-toewijzing en gegevenswoordenboek-toewijzing. Metagegevenstoewijzing-services helpen bij het definiëren van de bedrijfsgegevens in de verschillende systemen. Ze verzorgen tevens enkelvoudige toewijzing tussen algemene elementen via de opbouw van een bedrijfsgegevenswoordenboek voor meerdere systemen. Het bedrijfsgegevenswoordenboek is een reeks metagegevensdefinities en representaties van gegevenselementen.

Federatieve toegang. De laag federatieve toegang verzorgt de opbouw van zoekopdracht-brokers die informatie uit onderliggende systemen kunnen verkrijgen. Voor toegang tot de bedrijfsgegevens is een universele index voor verwante semantiek vereist.

Voorbeeld van zoekoplossingarchitectuur

In afbeelding 2 wordt een voorbeeld van een zoekoplossingarchitectuur getoond, die is afgeleid van de conceptuele architectuur die hierboven is weergegeven.

De architectuurlagen zijn:

Gebruikersinterface

Dit is de laag waarmee de gebruikersinteracties worden bestuurd via een webbrowser of zelfs een mobiel apparaat. Deze laag verzorgt de presentatie van schermen en gekoppelde gegevens aan zoekopdrachtgevers.

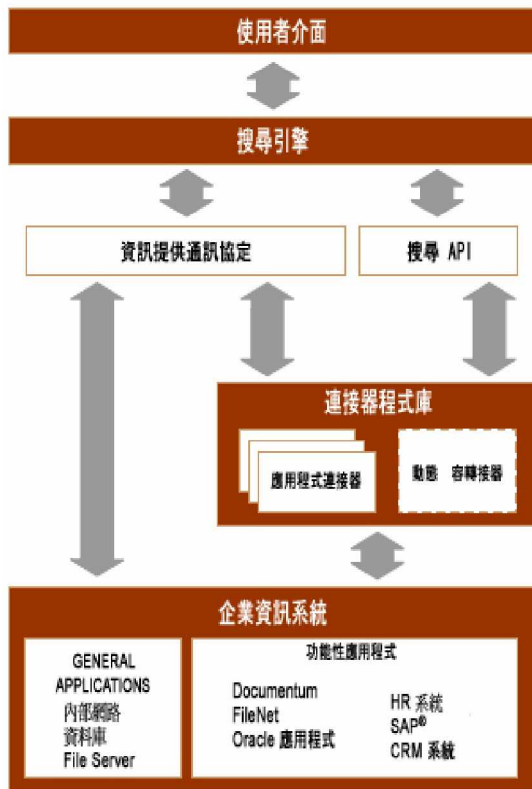
De gebruikersinterface is een XML-gebaseerde interface, die dynamisch en snel door de toepassingslaag wordt gecreëerd. Afhankelijk van het apparaat waarmee de zoekopdracht wordt uitgevoerd, levert een XSLT-bestand de juiste presentatie. Zo wordt presentatie van inhoud gescheiden en worden de verantwoordelijkheden duidelijk verdeeld. Deze laag is ook nauw verbonden met de beveiligingscomponenten autorisatie en verificatie, om ervoor te zorgen dat de zoekresultaten alleen aan de personen worden getoond die hiervoor inzagebevoegdheid hebben.

Belangrijke opties die de bruikbaarheid van de zoekoplossing vergroten, zijn functiegerichte presentatie, modi voor weergave/bewerken en contextuele hulp en paginering van de zoekresultaten.

Afbeelding 1. conceptuele architectuur van zoekoplossing



Afbeelding 2. Voorbeeld van zoekoplossing-architectuur



de beveiligingskenmerken die door de beveiligingsgegevens-agents worden vereist. In plaats daarvan is de aanbevolen implementatiemethode adapters te gebruiken die de zoekopdracht namens de zoekopdrachtgever doorstuurt naar de base content catalogus.

- **Resultaataggregator.** Dit domein verzamelt de resultaten die overeenkomen met de zoekopdracht uit de verschillende catalogi en presenteert deze aan de gebruiker. Daarnaast moeten ook de mogelijkheden worden geboden van feedback op de geleverde resultaten, een alternatieve zoekopdracht en de mogelijkheid om een gecategoriseerde zoekopdracht terug te geven aan de zoekmachine.

Zoekmachine

De zoekmachine is een hardware en/of softwareproduct dat via een vak voorziet in een bedrijfszoekfunctie. Deze kan naar verwachting honderden verschillende bestandsindelingen in elke gewenste taal en miljoenen documenten doorzoeken indexeren. De zoekmachine moet over beveiligingsfuncties beschikken om ervoor te zorgen dat gebruikers alleen die documenten te zien krijgen waarvoor zij inzagebevoegdheid hebben.

De belangrijkste functies waarover een zoekmachine moet beschikken, zijn (zie afbeelding 3):

Een normale bedrijfsomgeving heeft binnen zijn het intranet verschillende opslagplaatsen van webinhoud en documenten. In deze opslagplaatsen bevinden zich verschillende technologieën, die worden gebruikt voor de feitelijke hosting van inhoud. Hosting omvat inherente beveiligingsmechanismen, individuele indexcatalogi en backend-opslagplaatsen van de inhoud zelf. De architectuur van de zoekoplossing moet functionaliteit bieden in drie domeinen:

- **Zoekopdracht-preprocessor.** De zoekopdracht-preprocessor is een centrale service die vanaf een aantal locaties kan worden aangeroepen. De preprocessor splitst de zoekopdracht op, controleert de spelling, maakt vertalingen in andere talen, identificeert synoniemen, concepten en het meest geschikte type zoekopdracht voor de gekoppelde inhoudsopslagplaats; bijvoorbeeld zoekwoord, concept, Booleaans en volledige tekst.
- **Catalogusadapter.** Catalogusadapters verzorgen het integratiecontract tussen de oorspronkelijke zoekopdracht en de backend inhoudsindexcatalogi. Een enkelvoudige-indexfunctie voor alle inhoudsopslagplaatsen voorziet niet in een

- **Presenteren.** Voor de presentatie wordt een standaard zoekinterface geleverd, die standaard vanaf de zoekmachine kan worden gehost. Van deze interface moet het onderliggende XSLT-opmaakprofiel worden aangepast. Met aanvullende zoekfuncties als KeyMatch, ondersteunen synoniemen en filters specifieke webpagina's als onderdeel van de zoekresultaten. Synoniemen zijn voorgestelde alternatieve woorden of zinsdelen voor de zoekopdracht. Met aanvullende functies worden logische informatiegroepen gecreëerd, de zogeheten 'collecties', die helpen tegemoet te komen aan specifieke zoekbehoeften van de gebruiker.

Afbeelding 3. zoekmachinefunctionaliteit



- **Crawlen.** Crawlers zijn agents die documenten van webserver opvragen en ophalen voor automatische indexering. Er zijn twee methoden van gegevenscrawling: gegevens ontdekken en gegevens ophalen. Met een configuratie-interface kan in verschillende gegevensbronnen worden gecrawled. Alle inhoud wordt bijeengebracht in een hoofdindex, die iedere keer dat er weer wordt gecrawled wordt bijgewerkt.

Crawling is een proces van drie stappen: oorspronkelijke ontdekking (nieuwe url's om te doorzoeken), indexering en onafgebroken zoeken. Elk nieuw document dat de crawler aantreft, wordt gescand op links. De links worden ofwel onmiddellijk onderzocht of later opgehaald. De crawler is bedreven in het omgaan met beveiligde inhoud en verwerkt S-HTTP-communicatie. De zoekmachine-crawler kan voorzien in basisverificatie, NTLM-verificatie en aangepaste -cookie- en formulier , NTLM-verificatie en aangepaste-cookietoegang en formuliertoeegang. De zoekmachine crawlt inhoud uit databases als Oracle, SQL Server, mySQL, IBM DB2 en Sybase. Een gegevenstype dat de crawler niet kan doorzoeken, kan direct in XML-indeling in de zoekmachine worden ingevoerd.

- **Indexeren.** Er kunnen verschillende methoden worden gebruikt om de gegevens op voorhand te indexeren:

– Een feed naar het zoekstelsel met behulp van standaard- of aangepaste adapters.

– Directe toegang via het zoekstelsel (alleen met de meegeleverde standaardadapter heeft men directe toegang tot gegevens).

– HTTP of non-HTTP gegevensfeeds.

- **Feeds.** Sommige typen documenten kunnen het beste met behulp van feeds in de zoekmachine worden opgenomen, in plaats van te worden gevonden via links op doorzochte webpagina's. Bijvoorbeeld:

– Documenten die niet met behulp van de crawler kunnen worden opgehaald. Bijvoorbeeld, records in een database of bestanden in een systeem dat niet geschikt is voor het web.

– Documenten die kunnen worden doorzocht, maar het beste opnieuw kunnen worden doorzocht op andere tijdstippen dat die welke zijn ingesteld door de automatische crawlplanner van de zoekmachine.

- Documenten die kunnen worden doorzocht, maar geen links op de website bevatten, waardoor de crawler ze opnieuw kan ontdekken tijdens een nieuwe crawl.
- Documenten die kunnen worden doorzocht, maar sneller geüpload kunnen worden met behulp van feeds vanwege webserver- of netwerkproblemen.

Mogelijke feedtypen zijn:

- Inhoudsfeeds—inhoudsfeeds bevatten zowel de url als de inhoud. Mogelijk bevatten zij eveneens enkele metagegevens, zoals de datum waarop het record voor het laatst werd gewijzigd. Voor een inhoudsfeed kan elke gewenste gegevensbronnaam worden opgegeven.
- Webfeeds—In een webfeed bevatten alle records url's, maar geen inhoud. Deze url's worden normaal gecrawled. Voor dergelijke feeds wordt de term 'web' gebruikt als de naam van de gegevensbron. Alle url's moeten een volledig gekwalificeerde domeinnaam (FQDN) bevatten in het host-gedeelte van de url.
- Databasefeeds —Aangepaste connectors kunnen worden geschreven om records uit een database te 'pushen' voor de aanmaak van een index.

Het instellingsproces van zoekmachinefeeds wordt in afbeelding 4 weergegeven.

Afbeelding 4. zoekinhoud-feeder API



- 1 Script exporteert gegevens in XML-indeling om het feedbestand te genereren
- 2 Script plaatst het XML-bestand op de zoekmachine
- 3 Zoekmachine importeert het XML-bestand en indexeert de inhoud

- **Beveiliging binnen de zoekmachine.** De zoekmachine moet het bedrijfsbeveiligingssysteem versterken. Alle services moeten achter een firewall worden uitgevoerd en slechts bij uitzondering vindt er communicatie plaats met de zoekmachine van buiten de firewall. De zoekmachine crawlt en indexeert zowel openbare als vertrouwelijke documenten. Via een optioneel beveiligingspakket wordt het bedrijfsbeveiligingsbeleid op documentniveau afgedwongen.

De presentatie van een correcte reeks zoekresultaten aan een gebruiker is gebaseerd op het filteren van de resultaten van de doorzochte index. De inhoud wordt opgesplitst in twee categorieën: openbare inhoud en alle inhoud. Gebruikers kunnen alleen in openbare inhoud of in openbare inhoud en beveiligde documenten in de index zoeken, zoals aangegeven door de indexbeheerder. Standaard blijft inhoud volledig beveiligd, zodat gebruikers zonder toegang geen inhoud te zien zullen krijgen waarvoor ze geen inzagebevoegdheid hebben.

De zoekmachine moet basisverificatie, NTLM (NT LAN Manager)-verificatie en formulierverificatie ondersteunen. Elk van deze verificatieprotocollen wordt verschillend verwerkt door de zoekmachine en vereist een andere instelling door de beheerder. Via alle drie methoden kunnen beveiligde documenten op intranetsites worden doorzocht en geïndexeerd en kan verificatie voor het doorzoeken van die documenten worden uitgevoerd.

Een zoekautorisatie-API staat een webservice toe de communicatie te regelen tussen de autorisatie-API van de zoekmachine en de bedrijfsserver die de toegangscontroleservices levert, de zogeheten toegangsconnector (AC, access connector). Deze AC verzorgt een laag tussen de zoekmachine en het toegangscontrolesysteem van het bedrijf.

Connectorbibliotheek

Deze softwarelaag vormt de berichten- en/of toegangslaag tot de verschillende gegevensbronnen in een onderneming, van gestructureerde gegevensbronnen tot bedrijfstoepassingen. Meestal worden een integratieplatform of aangepaste connectors of adapters gebouwd om de gegevensbronnen te integreren.

Aangepaste connectors kunnen ook worden gebouwd door de volgende drie stappen uit te voeren:

- Converteer de gegevens naar XML inde indeling die in de zoekmachinefeed is opgegeven.

- Upload de XML naar de machine met behulp van het HTTP-protocol.

Deze laag vormt de onderliggende structuur voor het beschikbaar maken van informatie in een onderneming, ongeacht de bron of de opslagindeling. Deze laag staat een open verbinding toe tussen services en integratieadapters die uit verschillende gegevensbronnen zijn samengesteld.

De metagegevens die in deze laag worden gedefinieerd kunnen ook in de infrastructuur voorzien die nodig is voor het consolideren van informatie uit ongelijksoortige bronnen in een onderneming. Doorgaans wordt hierdoor ook mogelijk dat items die semantisch overeenkomen snel worden ontdekt en opgehaald. Zo worden met een verenigd metamodel consistente en contextueel correcte metagegevensdefinities opgesteld voor de consolidatie van ongelijksoortige informatie. Hierdoor worden tevens de zoekresultaten uitgebreid met informatie van contextgedreven links naar gestructureerde gegevens en bedrijfstoepassingen.

Informatiebronnen en -typen

Informatie bevindt zich tegenwoordig overal in ondernemingen. Deze is beschikbaar in gestructureerde en ongestructureerde indeling, in eigen of niet-eigen opslagplaatsen (zie tabel 1). Zoekmachines kunnen informatie doorzoeken die geïndexeerd kan worden, wat het ophalen vergemakkelijkt en relevante zoekresultaten oplevert. Het is daarom essentieel dat gegevensbronnen dienovereenkomstig worden geclassificeerd en dat vervolgens de juiste ophaalmechanismen worden gebruikt.

Tabel 1. Informatiebronnen en -typen

	EIGEN	NIET-EIGEN
Gestructureerd	Databasebeheersystemen (Oracle, SQL Server, Sybase, Informix) Documentbeheersystemen (Documentum, FileNet, etc.) Bedrijfstoepassingen, zoals ERP en CRM	Databasebeheersystemen (mySQL)
Ongestructureerd	Netwerkbestandssystemen/bestandssystemen Microsoft Office-documenten Afbeeldings-, video- en audiobestanden	Open office-documenten Bestandssystemen/bestands server Tekstbestanden E-mail

Gestructureerd versus ongestructureerd. Het grootste verschil tussen ongestructureerde en gestructureerde informatie is een vooraf gedefinieerd informatiemodel. Gestructureerde informatie is informatie die in een vooraf gedefinieerde indeling, zoals HTML, kan worden geplaatst, of in een tabel, zoals in het geval van databasetechnologie. Een andere vorm van gestructureerde informatie zijn metagegevens —informatie over informatie. Ongestructureerde gegevens daarentegen, kennen geen vooraf gedefinieerde indeling en dus geen structuur. Voorbeelden van dergelijke gegevens zijn tekstbestanden en PDF-documenten. In het geval van ongestructureerde informatie, vormt het ophalen van informatie een speciale uitdaging voor de zoekoplossing.

Eigen versus niet-eigen opslagplaatsen.

Bedrijfsinformatie is tegenwoordig beschikbaar ERP, CRM en kennisbeheersystemen en in statische en dynamische portals. Sommige van deze systemen bieden bij het ophalen van informatie eigen indelingen waarin de informatie wordt gepresenteerd. Weer andere maken gebruik van niet-eigen indelingen, zoals XML, wat steeds populairder wordt. In de meeste toepassingen wordt nog altijd met tekst gewerkt, maar andere mediatypen, zoals afbeeldingen, audio en video zijn sterk in opkomst. Daarom is een doeltreffende combinatie van automatisch tekst ophalen, ophalen in metagegevens (gewoonlijk handmatig) en inhoudsgericht ophalen van multimediagegevens vereist.

DE MOGELIJKHEID VAN DE ZOEKOPLOSSING

Een zoekoplossing voor de onderneming kan een katalysator zijn voor de productiviteit van de medewerkers en weloverwogen besluitvorming. Bij de implementatie van een zoekoplossing in de bedrijfsomgeving moet echter terdege rekening worden gehouden met de vereisten en mogelijkheden van de bedrijfsarchitectuur. Zorgvuldige architecturale planning en implementatie zijn essentiële voorwaarden voor de verwezenlijking van de grote voordelen die een zoekoplossing biedt.

[Neem contact met ons op](#) voor meer informatie over de wijze waarop onze oplossingen uw bedrijf kunnen helpen.

**WERELDWIJD BEHEER EN TECHNOLOGIEADVIES
VOOR DE MODERNE BEDRIJFSOMGEVING**

BearingPoint is een toonaangevend consultancybureau dat wereldwijd beheer- en technologieadvies levert aan bedrijven uit de Global 2000 en veel van 's werelds grootste publieke bedrijven. Onze ervaren professionals helpen bedrijven rond de wereld hun doelstellingen te bereiken en waarde voor de onderneming te genereren. Door hun bedrijfsprocessen en informatiesystemen af te stemmen, helpen we onze klanten aan een concurrentievoordeel omdat ze informatie sneller kunnen leveren. Neem voor meer informatie contact met ons op via 1.866.661.FIND (+1-603-589-4089 buiten de Verenigde Staten en Canada) of ga naar onze website op www.bearingpoint.com.

BearingPoint biedt strategisch advies,
toepassingservices, technologieoplossingen en
beheerde services aan Global 2000-bedrijven en
overheidsinstellingen.



Bearingpoint
1676 International Drive
McLean, VA 22102
www.bearingpoint.com