



Een contentfilter tegen niet-relevante of onjuist geclassificeerde informatie

STRUCTUUR BRENGEN IN DARK DATA

De transformatie naar een volledig digitale organisatie vraagt het uiterste van alle betrokkenen, zegt Hans Kaashoek. De toenemende diversiteit van informatiestromen en de explosieve groei van berichtenverkeer stellen hoge eisen aan de kwaliteit van informatieverwerking. Hoe brengt de IT-manager structuur in de groeiende hoeveelheid dark data?

door: HANS KAASHOEK beeld: SHUTTERSTOCK

Een organisatie loopt een groot risico dicht te slibben met informatie waarvan de betekenis en waarde voor de organisatie niet duidelijk zijn. Continue aandacht is vereist voor de diversiteit, het volume en de kwaliteit van de te verwerken informatie. Om wildgroei van dark of 'rot'-data (redundant, oud en triviaal) te voorkomen zal de onderneming een informatie-analyse en -classificatieschil tussen de inkomende informatiestromen en de onderneming en tussen de informatiebronnen en haar medewerkers moeten implementeren. Deze schil dient als een permanent contentfilter tegen niet-relevante of onjuist geclassificeerde informatie. Alleen wat relevant en juist is, wordt toegelaten en kan direct verwerkt of gepresenteerd worden: clean data waarmee processen en klantinteractie in hoge mate kunnen worden geautomatiseerd – ongeacht het soort informatiekanaal, de informatiebron of de structuur van de onderneming.

Diversiteit

Het verleden was zeker niet perfect, maar soms wel eenvoudig en gestructureerd. Een voorbeeld. Bij iedere autoverzekering, ongeacht de

maatschappij en het type verzekering, kreeg je een standaardaanrijdingsformulier. In geval van een calamiteit vulden beide partijen alle gevraagde gegevens in, maakten een situatieschets en verzonden het ondertekende formulier naar hun maatschappij. Daar werd het ontvangen in de postkamer, middels scanning gedigitaliseerd en gerouteerd naar de schadeafdeling voor verdere verwerking. Bij iedere maatschappij waren het proces en de daarbij behorende gegevens identiek. Verdere interne procesversnelling en

verbetering van de klantenservice waren mogelijk door gebruik te maken van document- en werkstroomautomatisering. Door het formulier zelf te uploaden of te versturen als e-document kwam het direct in de behandelstroom. Met het gebruik van de smartphone als gegevensinvoerdevice is de uniformiteit van de informatiestromen en inhoud verleden tijd. Twee jaar geleden introduceerde de overkoepelende verzekeringsinstantie de app mobielschademen.nl. Deze combineerde directe gegevensin-

ÉÉN KLANTSISTEEM

Om te komen tot één centraal klantinformatiesysteem analyseert en consolideert creditkaardienstverlener ICS al zijn klantgerelateerde content. Zo verzorgt ICS naast de eigen Visa en MasterCard de gehele creditkaardienstverlening voor ABN AMRO, ANWB, BMW en de Bijenkorf. In het consolidatieproces van de diverse informatiesystemen heeft ICS nog een AS/400 elektronische archiefapplicatie met daarin essentiële klantinformatie van de afgelopen 20 jaar. Om deze klantgegevens en bijbehorende content toe te kunnen voegen aan het centrale klantinformatiesysteem voert ICS een Content ETL-project uit. Een eerste essentiële stap is de analyse van de inhoud van het AS/400-archief en de waardebevestiging van de content. Dit in het kader van nieuwe (privacy)wetgeving en de beoogde optimalisatie van de informatiehuishouding. Als tweede stap worden de metadata van de content verrijkt zodat deze aansluiten bij de gegevensstructuur van het nieuwe systeem. Hierna vindt de daadwerkelijke migratie van de content plaats en kan het AS/400-systeem worden uitgefaseerd.

voer en -validatie met de optie om foto's mee te sturen. Het centrale orgaan stuurde de gegevens door naar de betreffende maatschappijen. Het verwerken van al deze verschillende inkomende informatiestromen wordt het domein van omnichannel-inputmanagement.

Om nog klantvriendelijker te zijn heeft nu iedere verzekeringsmaatschappij haar eigen schade-melding en registratieproces: als smartphone-app of via Twitter, WhatsApp of FaceBook. Alleen een serie foto's of een videofilm is soms al voldoende. De vereiste gegevens worden on-the-fly uit een grote verscheidenheid van in- en externe bronnen verzameld. Echter, de vereiste kanaalflexibiliteit en vergaande integratie met de gegevensverwerkende systemen voor optimale klantenselfservice staat vaak in schril contrast met de flexibiliteit van de in gebruik zijnde applicaties en systemen. Een volledig digitale verzekeraar maakt bij voorkeur alleen gebruik van social media-gebaseerde communicatiekanalen. Maar een gevestigde verzekeraar met een groot kantenbestand zal een brede variëteit aan communicatiekanalen en berichtsoorten open moeten houden. Met voor ieder kanaal een andere berichtvorm en -inhoud en registratievereisten. Het vormen en onderhouden van een eenduidig klantbeeld of zaakdossier wordt een grote uitdaging.

De hiervoor geschetste berichtendiversiteit geldt voor iedere organisatie met een bestaand klantenbestand en enige vorm van klantinteractie. Lokale overheden, banken, retailers, zorgverleners, woningcorporaties maar ook productiebedrijven worden dagelijks geconfronteerd met weer een nieuw communicatiekanaal en een daarbij behorend ander berichtentype. In contact blijven en communiceren met je klanten betekent soms dat een organisatie zich ondergeschikt moet maken aan de communicatievoorkeur van haar doelgroep. Hiervoor moet de organisatie wel beschikken over de juiste strategie, kennis, processen en technologie om deze ontwikkeling te ondersteunen. Iets waar het bij veel gevestigde organisaties ernstig aan schort. Het zijn niet alleen de flexibiliteit ten aanzien van de externe informatiestromen en de unifor-

me verwerking daarvan, die de positie en het functioneren van een organisatie in de digitale wereld bepalen. Ook toegang tot interne informatie en de daarbij behorende informatiestromen spelen een belangrijke rol. De creatie, verwerking en opslag van ongestructureerde informatie zijn van ouds een gebruikergedreven taak. Met als gevolg dat deze informatie over de jaren heen vaak in verschillende systemen of op verschillende locaties is opgeslagen: opslaglocaties zoals een document- of contentmanagementsysteem, een SharePoint-projectsite, het e-mailsysteem, in- of externe FileShare-locaties, de eigen lokale schijf of als objecten binnen een gegevensverwerkende applicatie. Waarbij voor iedere opslaglocatie andere, of zelfs geen, registratie- en ontsluitingscriteria van toepassing zijn. Het niet via één zoekopdracht kunnen vinden en samenvoegen van alle relevante informatie om een zaak snel en adequaat te behandelen, is nog steeds een van de top-ergermissen van de kenniswerker. En vanuit kosten, beheer en complianceperspectief een van de pijnpunten voor de IT-manager.

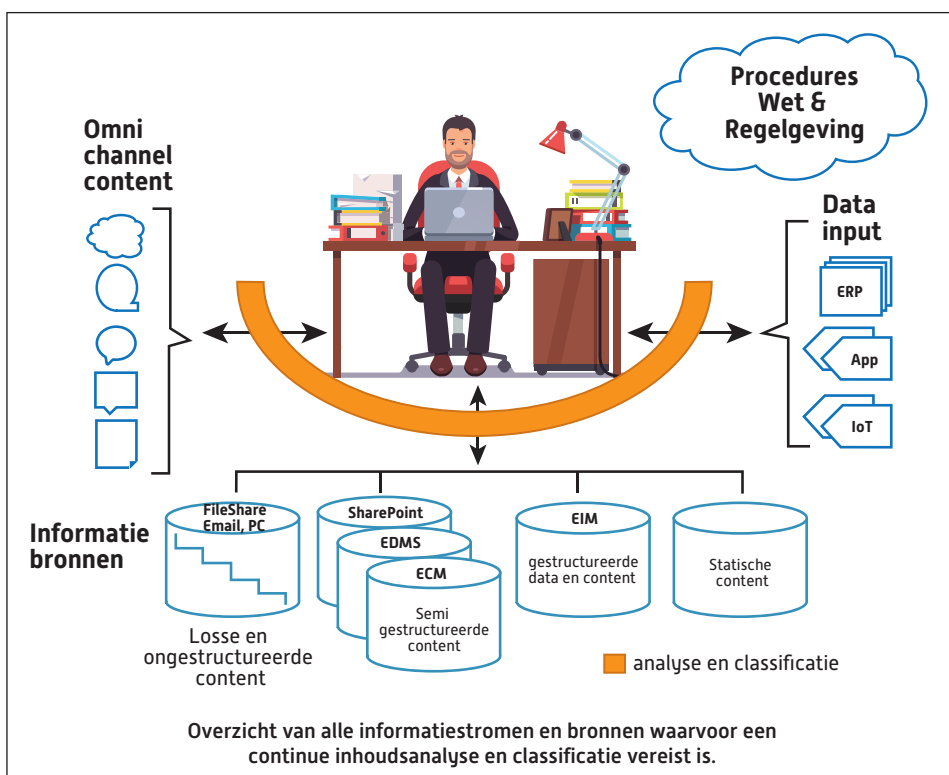
Explosie datavolume

De hiervoor geschetste informatieuitwisseling is in volume en omvang nog maar gering in het licht van actuele ontwikkelingen. Video gaat een steeds grotere rol spelen in de onderlinge communicatie. De steeds hogere resolutie van dit medium en de vastleggingsvrijheid van de gebruiker resulteren al snel in explosief grote bestanden. Liever tien filmpjes om zeker te zijn dat het totaalbeeld goed is, dan één goed gestructureerde omschrijving. En met de doorontwikkeling van Internet of Things krijgt de berichtenstroom een continu karakter. Het voortdurend meten en registreren van data via camera's, sensoren, smartwatches of mobiele telefoons vergroten de hoeveelheid te beheren informatie. Waarbij het van tevoren nooit echt duidelijk is wat de inhoud van de datastromen is en de waarde daarvan voor de organisatie. Deze onzekerheid zal een belangrijk argument zijn voor het blijvend opslaan van big en dark data. Indien deze data niet rechtstreeks door een or-

ganisatie worden ontvangen, dan kunnen ze ook uit steeds meer openbare bronnen worden ontleend. Open data maken voorheen gesloten bronnen eenvoudiger benaderbaar en de inhoud toegankelijk. Het extraheren van relevante informatie uit alle verschillende in- en externe bronnen ter ondersteuning van de analyse en afhandeling van een klantgestuurd bericht of verzoek, stelt hoge eisen aan de informatie-ontsluiting en ordening. Een generiek enterprisezoekstelsel is voor de kenniswerker hierbij een goed hulpmiddel, maar biedt zeker nog niet de ideale oplossing. Deze oplossing bestaat uit een combinatie van on-demand contextuele informatie-analyse en zoekfuncties, én een gestructureerde informatieclassificatie van de diverse informatie-objecten vooraf.

Kwaliteit en uniformiteit

Iedere organisatie moet voldoen aan specifieke bedrijfsregels en condities. Regels die de verschillende producten en diensten definiëren, en bepalen hoe zaken in processen worden behandeld, geaccepteerd of afgewezen. Bij deze regels en processen horen informatie-elementen die ter ondersteuning van of als aanvulling op gegevensverwerkende applicaties en processtappen beschikbaar moeten zijn. Om tot één universele informatiehuishouding te komen zullen alle informatie-elementen in lijn met een bedrijfsprocesgericht begripkader moeten worden geclassificeerd. Masterdata en metadata-kwaliteitsmanagement zijn vereist voor alle gestructureerde en ongestructureerde informatie-elementen. Deze, liefst automatische, analyse en classificatie van alle informatie-elementen zullen in de informatieclassificatie-gateways moeten plaatsvinden. Dit voordat de informatie wordt gepresenteerd, vastgelegd en opgeslagen. Hierdoor is de informatie in de organisatie altijd actueel geclassificeerd en eenduidig door een medewerker te benaderen. Hetzelfde geldt voor de in de organisatie aanwezige verzamelingen ongestructureerde informatie (content). De ontsluiting en toegankelijkheid hiervan zijn nu afhankelijk van de kwaliteit van de metadata, zoals die ooit door de gebruiker zijn meegegeven. De analyse van deze content en de herclassificatie ervan, al dan niet als onderdeel van een contentrationalisatie-, consolidatie- en migratieproject, zijn essentiële succesfactoren bij het adequaat realiseren van één digitale informatiehuishouding. Een dergelijk contentoptimalisatieproces wordt ook wel Content ETL genoemd (Extract, Transform en Load) en kent dezelfde stappen als een database- en informatie-ETL-project. Content-ETL-projecten hebben doorgaans een positief resultaat. Een kostenbesparing voor opslag en beheer door deduplicatie, schoning en consolidatie van 30 procent is geen uitzondering. Een 45 procent reductie in zoektijd en een verhoging van de informatierelevantie leveren de organisatie het echte gewin. <<



Hans Kaashoek is Managing Partner bij Strategy Partners Nederland B.V. en verantwoordelijk voor marktonderzoek en strategie-adviesdiensten (hans.kaashoek@strategy-partners.nl).



Voor reacties en nieuwe bijdragen van IT-experts: Henk Ester, 020 235 6415 h.ester@automatise-ringgids.nl