

Business Intelligence software wordt eindelijk intelligent

# PROCESS MINING

Business Intelligence is een begrip dat verwijst naar software die gebruikt kan worden om gegevens over operationele bedrijfsprocessen te verzamelen en deze vervolgens te analyseren. Het doel van BI software is het verkrijgen van meer kennis en inzicht, welke gebruikt kunnen worden om processen beter te besturen en in te richten.

Door Wil van der Aalst

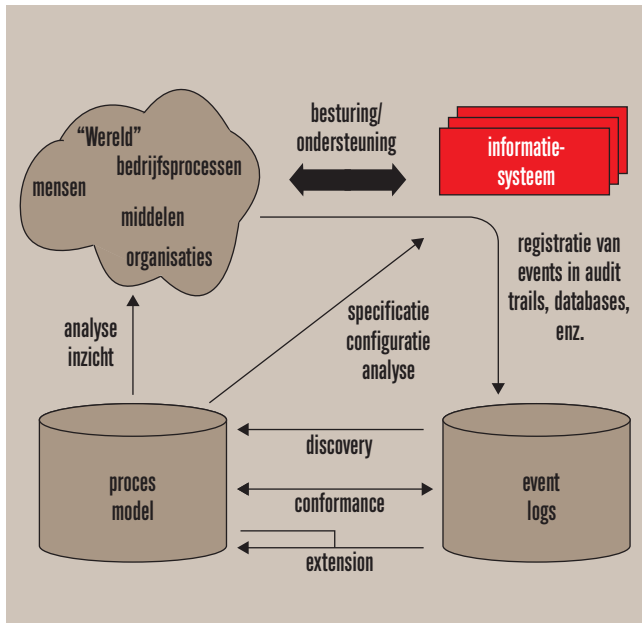
**D**e focus ligt hierbij vaak op performance indicatoren die betrekking hebben op kosten, doorlooptijden, service-niveaus, en bezettingsgraden. De 'I' in BI suggereert dat deze systemen intelligent zijn. Niets is minder waar. Vaak gaat het om vrij triviale analyses van gestructureerde data. Veel processen zijn lastig te karakteriseren in termen van performance indicatoren omdat de werkelijkheid niet te vangen is in een paar getallen. Ook veronderstellen BI-tools gestructureerde processen en gestructureerde gegevens waarin geen afwijkingen mogelijk zijn.

Neem bijvoorbeeld de behandeling van hartpatiënten in een ziekenhuis. Met behulp van de huidige generatie BI-tools is het mogelijk om vast te stellen wat het aantal operaties is geweest in een bepaalde periode en hoelang de gemiddelde behandeling van de patiënten na de operatie heeft geduurd. Ook is het mogelijk om één individuele patiënt te volgen. Helaas is het niet mogelijk af te leiden hoe de processen daadwerkelijk verlopen zijn en waar en waarom er afwijkingen en knelpunten zijn ontstaan. De reden hiervoor is dat de werkelijkheid wordt gevangen in getallen zonder dat het proces adequaat zichtbaar gemaakt wordt. De 'management dashboards' laten meters en grafieken zien, maar niet het onderliggende proces. De meer geavanceerde BI-tools beelden performance indicatoren af op een procesmodel. Het gaat dan echter altijd om een vooraf gedefinieerd procesmodel dat niet noodzakelijk de werkelijkheid hoeft weer te geven. Vaak laten met de hand gemaakte procesmodellen een sterk geïdealiseerde wereld zien waardoor de daadwerkelijke verbeterpunten verborgen blijven. Kortom, het is de hoogste tijd dat BI-tools echt intelligent worden. Gelukkig zijn er recent belangrijke doorbraken gerealiseerd op het gebied van process mining, die kunnen helpen bij de analyse van processen zoals deze daadwerkelijk plaatsvinden.

## Afspiegeling van de werkelijkheid

Om het idee van proces mining duidelijk te maken is het van belang om te realiseren dat informatiesystemen vaak een afspiegeling van de werkelijkheid vormen. Enerzijds worden informatiesystemen gedreven door modellen van de bedrijfsprocessen die ze ondersteunen. Anderzijds is er een voortdurende interactie tussen het bedrijfsproces en het informatiesysteem. Neem bijvoorbeeld een ERP-systeem als SAP R/3. SAP R/3 veronderstelt dat er binnen organisaties bepaalde processen plaatsvinden, bijvoorbeeld het plaatsen van bestellingen, en dat er voor deze processen bepaalde activiteiten in een zinvolle volgorde worden uitgevoerd. Dit zijn de processen die vastgelegd zijn in het zogenaamde SAP referentiemodel. De ontwikkeling en configuratie van SAP R/3 wordt dus gedreven door deze veronderstelde processen. In werkelijkheid laat het systeem echter ook andere processen en procesvolgordes toe om de benodigde flexibiliteit te realiseren. Onafhankelijk van de wijze waarop het systeem gebruikt wordt, worden zogenaamde events gelogd. Een event is een gebeurtenis die overeenkomt met een bepaalde actie of handeling. In de context van informatiesystemen wordt ook wel gesproken over (database) transacties of audit trail entry's. De event logs van systemen als SAP R/3 tonen een afbeelding van de werkelijkheid en vormen het startpunt voor process mining.

Het is van belang om in te zien dat in toenemende mate gebeurtenissen in bedrijfsprocessen geregistreerd worden in event logs. Pakketten voor het realiseren van informatiesystemen, bijvoorbeeld Enterprise Resource Planning (ERP), Customer Relationship Management (CRM), Product Data Management (PDM), en Workflow Management (WfM) systemen, bieden standaard de functionaliteit om gedetailleerde event logs te genereren. Ook hebben steeds meer systemen ingebouwde logging-functionaliteit. Complexe professionele systemen zoals medische apparatuur (bijvoorbeeld de röntgensystemen van Philips), high-end kopieermachines, lithografiesystemen (bijvoorbeeld de



Afbeelding 1: Drie soorten process mining: (1) discovery, (2) conformance, en (3) extension.

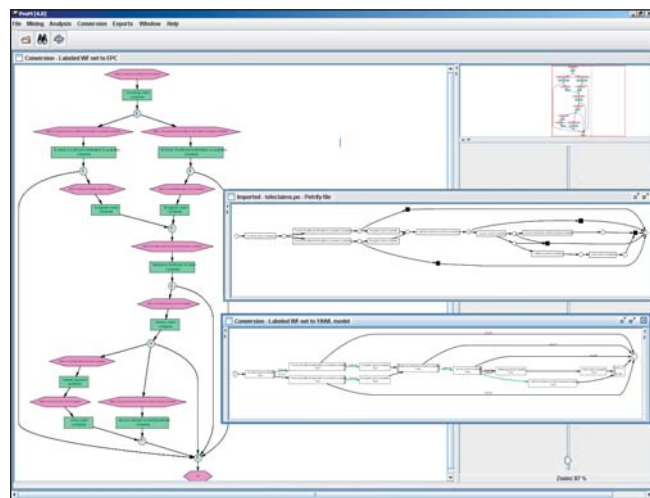
wafer stepper van ASML) zijn in staat alle handelingen te registreren en te distribueren via het Internet. Ook het toenemende gebruik van webservices leidt tot steeds meer en betere event logs. Door bijvoorbeeld SOAP messages te volgen kan de communicatie tussen partijen in een Service Oriented Architecture in kaart gebracht worden. De brede beschikbaarheid van event logs en de koppeling tussen bedrijfsprocessen en informatiesystemen maakt process mining mogelijk. Zoals aangegeven in afbeelding 1 beschouwen we drie soorten process mining: (1) discovery, (2) conformance, en (3) extension.

## Technieken

Process mining-technieken gericht op discovery leiden automatisch modellen af uit de event logs. Op basis van ruwe informatie kunnen bijvoorbeeld procesmodellen afgeleid worden die weergeven wat er daadwerkelijk heeft plaatsgevonden. Het is dus niet de analist die met de hand een (veelal subjectief) model van de werkelijkheid maakt. In plaats daarvan wordt de event log onderzocht op mogelijke causale verbanden waarna de gevonden verbanden worden samengevat in een grafisch model (bijvoorbeeld een Petri-net-achtige notatie zoals gebruikt in tools als Protos, COSA, en YAWL of de EPC-notatie gebruikt in tools als ARIS en SAP R/3). Dit klinkt utopisch. Recent zijn er echter krachtige technieken ontwikkeld die dit mogelijk maken. De discovery-technieken beperken zich niet tot het procesaspect (dat wil zeggen de volgorde van activiteiten). Indien events ook informatie bevatten over resources (wie heeft de activiteit uitgevoerd en welke middelen zijn hierbij gebruikt), data (wat waren de karakteristieke kenmerken), tijd (wan-

neer vond de activiteit plaats) kunnen ook andere aspecten afgedekt worden. Het is bijvoorbeeld mogelijk om op basis van event logs een zogenaamd sociaal netwerk af te leiden. In een sociaal netwerk worden de personen en/of organisatieonderdelen weergegeven met hun onderlinge relaties. Hierdoor is het mogelijk weer te geven dat sommige personen een centrale plaats innemen in de organisatie en dat er veel werk van de ene naar de andere afdeling stroomt.

Een tweede categorie van process mining-technieken is gericht op conformance. Het doel is objectief te toetsen of de werkelijkheid overeenkomt met de verwachting of vooraf opgestelde eisen. Conformance checking kan helpen bij het concretiseren van compliance en regelgeving als Sarbanes-Oxley (SOX). Schandalen binnen organisaties (WorldCom, Enron) hebben laten zien dat de werkelijke procesgang sterk kan afwijken van veronderstelde processen. Binnen veel organisaties bestaan procesbeschrijvingen die weergeven hoe processen uitgevoerd (dienen te) worden. Deze procesmodellen worden vaak gemaakt met eenvoudige teken-tools zoals Visio of met meer geavanceerde gereedschappen als Protos en ARIS. Er is echter geen enkele garantie dat de processen ook daadwerkelijk deze modellen volgen. In veel gevallen beschrijven procesmodellen een geïdealiseerde 'Powerpoint werkelijkheid', die ver afstaat van de dagelijkse realiteit. Daarom is het zinvol om op basis van event logs deze modellen te toetsen. Process mining-technieken zijn in staat om aan te geven hoe goed de *fit* tussen realiteit en model is en kunnen ook weergeven waar in het proces de belangrijkste afwijkingen plaatsvinden. Vervolgens kunnen de gevallen die afwijken verder onderzocht worden met de eerder genoemde discovery-technieken. Ook is het mogelijk business rules te toetsen. Het is bijvoorbeeld interessant te weten in hoeveel procent van de gevallen een business rule daadwerkelijk gevolgd wordt.



Afbeelding 2: ProM met daarin een proces verkregen door process mining.

De laatste categorie (in afbeelding 1 *extension* genoemd) pakt een al bestaand model en vult dit aan met gegevens uit de event log. Het initiële model kan met de hand gemaakt zijn maar kan ook ontdekt zijn door middel van process mining. Een voorbeeld is het projecteren van knelpunten op een procesmodel op basis van in de log geregistreeerde tijdstempels. Ook kan op basis van gegevens in de log bepaald

## Met ProM is het mogelijk verschillende modeltypen door elkaar te gebruiken

worden wanneer er een bepaald pad in het procesmodel gevolgd wordt. Door middel van klassieke data mining-technieken kan bijvoorbeeld ontdekt worden dat een bepaalde categorie klanten vaak een bepaalde tak van het proces doorloopt. Het ultieme doel van process mining is een model te creëren dat goed overeenkomt met de werkelijkheid en dat alle aspecten afdekt (activiteiten, resources, tijd, data, enzovoort).

### De weerbarstige werkelijkheid

Aan de Technische Universiteit Eindhoven wordt sinds 2000 onderzoek gedaan naar process mining. Daarvoor lag de nadruk sterk op het modelleren van processen voor work-

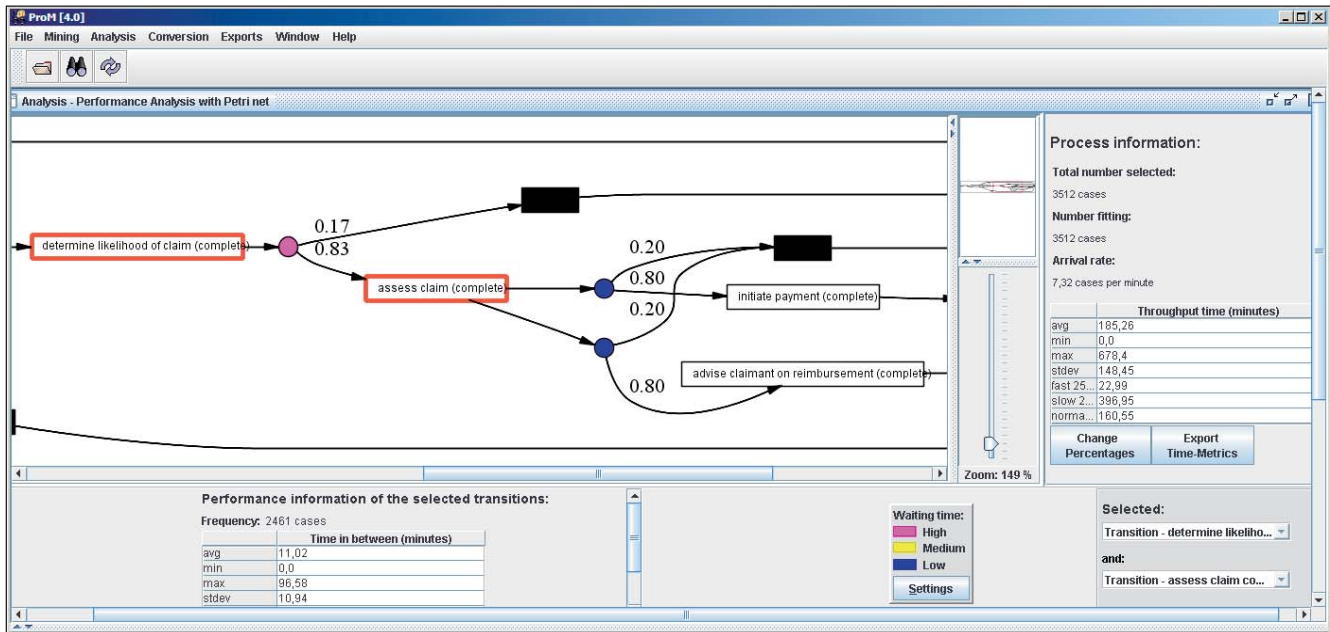
flow-systemen en de analyse van bedrijfsprocessen door middel van simulatie. Door de ervaringen op het gebied van simulatie en het realiseren van workflow-systemen werd het echter steeds duidelijker dat er meestal een grote discrepantie bestaat tussen model en werkelijkheid. Helaas hebben de meeste consultants en leveranciers vooral aandacht voor (her)ontwerp en wordt er te weinig energie besteed aan een gedegen analyse van de werkelijkheid waarin uitzonderingen vaak de regel zijn. Dit wordt bijvoorbeeld duidelijk in simulatiestudies waarin vaak aan parameterwaarden moet worden 'gesleuteld' om doorlooptijden overeen te laten komen met de werkelijkheid. Ook kan het falen van veel workflow-projecten verklaard worden door naïeve procesmodellen. Ervaringen met de analyse van patiëntstromen in ziekenhuizen laten bijvoorbeeld zien dat zelfs voor homogene groepen patiënten de processen minder structuur hebben dan verwacht.

Ook hebben de betrokken partijen vaak uiteenlopende beelden van het proces. Managers zijn vaak niet op de hoogte van de feitelijke procesgang (helikopterview) en uitvoeren zien vaak maar een beperkt deel van het proces (tunnelvisie). Het sterk vertekende beeld leidt tot systemen die onbruikbaar zijn en gemiste kansen om de processen te verbeteren. Zoals eerder aangegeven schieten BI-tools hier tekort doordat ze ook uitgaan van een geïdealiseerd en erg versimpeld beeld van de werkelijkheid. Performance indicatoren op een sterk geaggregeerd niveau zullen ziekenhuizen bijvoorbeeld niet helpen bij het in kaart brengen van en grip krijgen op patiëntstromen.



Performance Management Strategy	Data warehouse development	Reporting en analysis
<ul style="list-style-type: none"><li>- Design strategic management model</li><li>- Metrics definition</li><li>- Performance Management Roadmap</li></ul>	<ul style="list-style-type: none"><li>- Architecture design</li><li>- ETL design</li><li>- Data Quality Assurance</li><li>- ETL Development</li></ul>	<ul style="list-style-type: none"><li>- Reporting</li><li>- Dashboard development</li><li>- Data mining and analysis</li></ul>
		

www.quintica.nl



Afbeelding 3: ProM kan ook de bottlenecks zichtbaar maken in afgeleide modellen.

## ProM

Gedreven door de geschetste problemen wordt aan de Technische Universiteit Eindhoven door een vijftiental onderzoekers gewerkt aan de ontwikkeling van ProM; dit is een raamwerk waarbinnen gereedschappen zijn ontwikkeld die de eerder genoemde vormen van process mining ondersteunen. Er zijn bijvoorbeeld meer dan 20 verschillende discovery-technieken ontwikkeld. Het gebruikt van ProM start met een zogenaamde MXML log. Dit is een log waarin informatie over events op een gestructureerde manier wordt opgeslagen. Omdat de meeste informatiesystemen hun informatie niet opslaan in MXML formaat maar in leveranciersspecifieke databases, audit trails en transactie logs is het gereedschap ProMimport ontwikkeld. Met behulp van ProMimport kunnen logs vanuit systemen als FLOWer, Staffware, Apache, CPN Tools, CVS, Eastman, PeopleSoft, Subversion, WebSphere en Adept geladen worden. ProMimport maakt het eenvoudig om de benodigde import-functionaliteit te ontwikkelen en ondersteunt ook het anonimiseren van vertrouwelijke gegevens (bijvoorbeeld namen van medewerkers).

Met ProM is het ook mogelijk vele verschillende modeltypen door elkaar te gebruiken. De toolset ondersteunt bijvoorbeeld niet alleen diverse Petri-net formaten maar ook EPC-modellen en BPEL-specificaties. ProM kan ook modellen uitwisselen met systemen als Protos, ARIS, Websphere, YAWL, EPC Tools, enzovoort. Daarnaast zijn modelconversies mogelijk. Indien met behulp van process mining bijvoorbeeld een Petri-net ontdekt wordt, kan dit model vertaald worden naar BPEL, EPC's en YAWL.

Afbeelding 2 toont een schermafbeelding van ProM waarin op basis van een log een procesmodel is afgeleid. Het proces-

model is afgebeeld als een EPC die ingeladen kan worden in ARIS, een Petri-net dat als invoer kan dienen voor een scala van analysegereedschappen, en een YAWL model dat meteen uitgevoerd kan worden door het open source workflow-systeem YAWL.

ProM ondersteunt niet alleen discovery maar ook de twee andere vormen van process mining: conformance en extension. De conformance checker kan bijvoorbeeld in een procesmodel grafisch laten zien waar de grootste afwijkingen plaatsvinden en deze afwijkingen ook kwantificeren. Afbeelding 3 geeft een voorbeeld van extension; hier is in een procesmodel aangegeven wat de doorlooptijd is, waar de knelpunten zitten en hoeveel tijd er gemiddeld tussen twee activiteiten verstrijkt. Deze informatie kan weer gebruikt worden om volledig automatisch simulatiemodellen te genereren.

ProM is ontwikkeld onder een open source licentie (CPL) en is te verkrijgen via [www.processmining.org](http://www.processmining.org). De ontwikkeling van ProM wordt geleid vanuit de Technische Universiteit Eindhoven maar maakt gebruik van een breed internationaal netwerk van samenwerkende onderzoeksgroepen op het gebied van process mining. Naast het aanbieden van een zeer toepasbaar gereedschap is het doel van ProM ook om leveranciers van BI tools te stimuleren meer 'werkelijk intelligente' functionaliteit te bieden.

### Wil van der Aalst

Prof.dr.ir. W.M.P. van der Aalst is hoogleraar Informatiesystemen aan de faculteit Wiskunde & Informatica van de Technische Universiteit Eindhoven. Ook is hij verbonden aan de faculteit Technologie Management van dezelfde universiteit en de Business Process Management groep van Queensland University of Technology in Australië.