

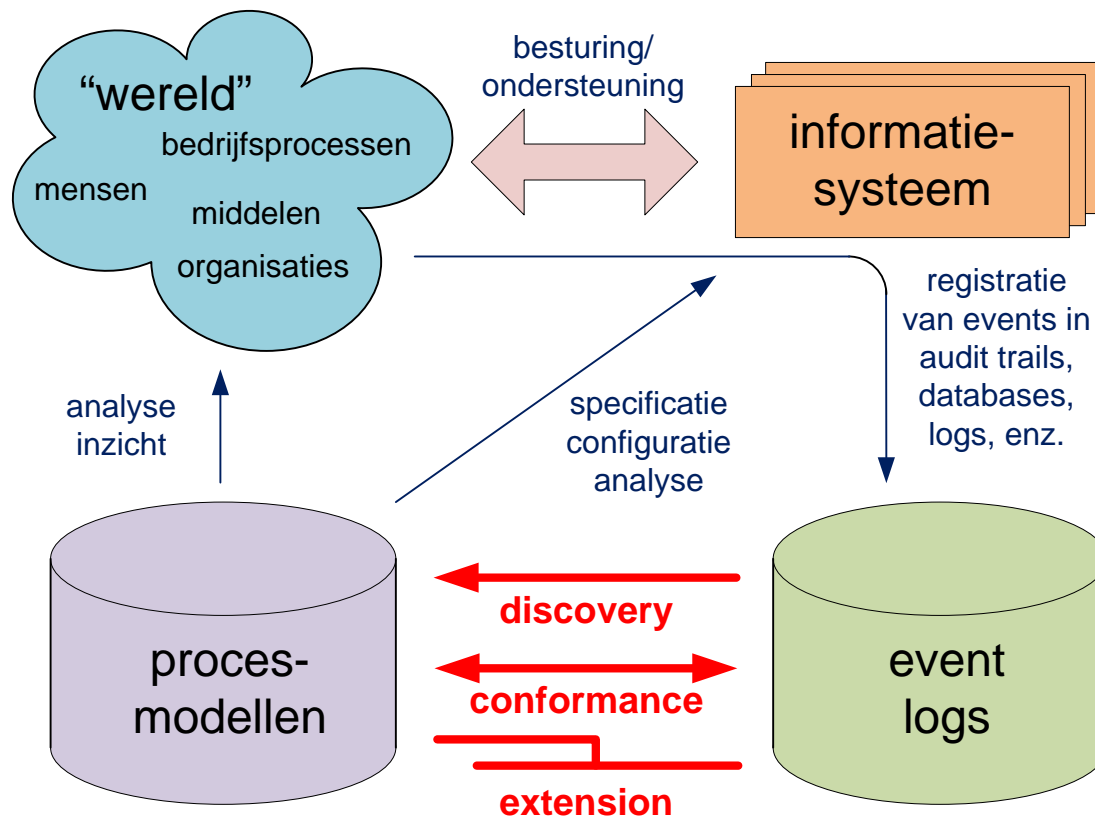
Process Mining: Wat gebeurt er nu echt? ... en hoe kan het beter?

Organisaties registreren steeds meer gegevens die gebruikt kunnen worden om processen te analyseren. Helaas schieten traditionele data mining and BI (Business Intelligence) technieken tekort vanwege de afwezigheid van een duidelijke procesoriëntatie. Process mining daarentegen slaat een brug tussen data mining en modelgebaseerde procesanalyse. Hierdoor is het mogelijk te laten zien wat er daadwerkelijk gebeurt. Door het afspelen van historische gegevens is het ook mogelijk te laten zien waar de knelpunten zitten en waar processen afwijken van het ideale pad. Deze inzichten zijn uitermate relevant voor zowel de IT professional als de business consultant.

Process Mining

De exponentiële groei van gedigitaliseerde data en rekenkracht maakt het mogelijk om processen steeds nauwkeuriger te volgen en te analyseren. Steeds meer gebeurtenissen worden immers geregistreerd en opgeslagen in zogenaamde *event logs*. Het boeken van een reis via het Internet leidt tot een lawine van events die opgeslagen worden in de databases van de betrokken organisaties. Het is in veel gevallen nu al mogelijk een product (bijvoorbeeld een bestelling, pakket, of container) op afstand te volgen. Steeds meer professionele apparaten (bijvoorbeeld CT/MRI scanners in een ziekenhuis) zijn verbonden met de informatiesystemen van leveranciers die inzicht willen krijgen in het gebruik van hun producten en storingen willen begrijpen en voorkomen. In toenemende mate wordt ook software op afstand in de gaten gehouden. Dankzij deze event logs en moderne process en data mining technieken is het mogelijk inzicht te krijgen in wat er nu werkelijk gebeurt.

Dankzij slimme analysetechnieken die zoeken naar veelvuldig voorkomende patronen is het mogelijk automatisch procesmodellen af te leiden ("discovery"). Deze modellen geven inzicht in wat er nu echt gebeurt binnen een proces of organisatie. Vaak wijken de gevonden procesmodellen sterk af van de normatieve procesbeschrijvingen die vaak uitgaan van een ideale situatie die weinig met de werkelijkheid te maken heeft. Om afwijkingen tussen een dergelijke geïdealiseerde procesbeschrijving en de daadwerkelijke gang van zaken in kaart te brengen worden er zogenaamde "conformance checking" technieken ingezet. Deze kunnen laten zien wat de mate van compliance is (bijvoorbeeld "80% van de events verlopen volgens plan") en waar in het proces de grootste afwijkingen te vinden zijn (bijvoorbeeld "deze controleactiviteit wordt veelvuldig overgeslagen"). Indien procesmodel en event log redelijk overeenstemmen is het ook mogelijk het procesmodel te verrijken met additionele informatie uit de log. Het is bijvoorbeeld mogelijk in het procesmodel te laten zien waar de knelpunten zitten (dit op basis van tijdstempels in de log).



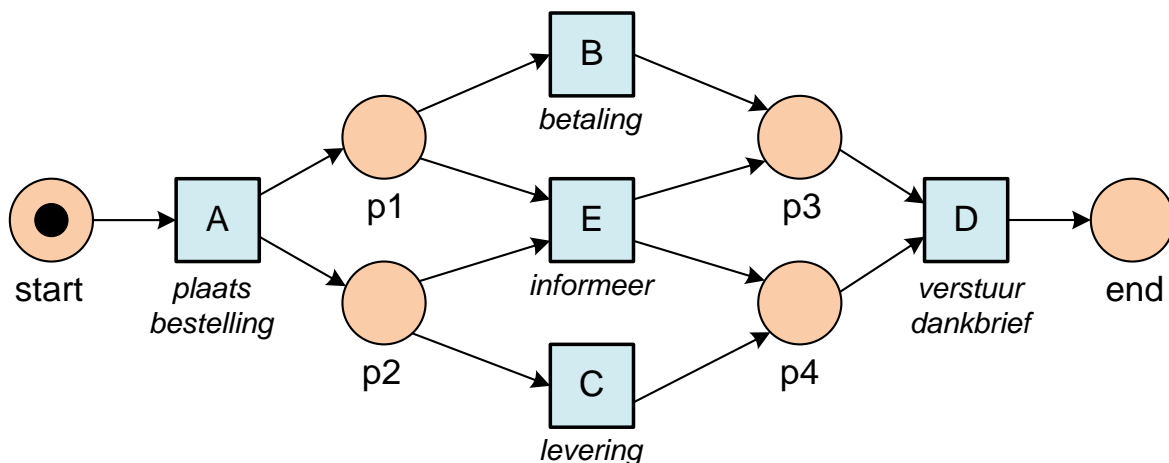
Bijchrift: Informatiesystemen besturen en ondersteunen in toenemende mate processen en organisaties. Als een gevolg hiervan worden steeds meer events geregistreerd. Deze kunnen gebruikt worden voor drie vormen van process mining: (a) “discovery”, (b) “conformance”, en (c) “extension”. “Discovery” is gericht op het automatisch construeren van een model op basis van een event log. “Conformance” daarentegen gaat uit van een event log en een model en laat zien waar de grootste afwijkingen tussen model en werkelijkheid zitten. “Extension” probeert het model te verrijken met behulp van kennis afgeleid uit de log met historische informatie.

Essentieel voor process mining is de koppeling tussen procesmodel en event log waardoor het mogelijk is de werkelijkheid af te spelen op het model. Afwijkingen tussen beide worden hierdoor meteen zichtbaar. Dit is belangrijk voor compliance en auditing. Tijdstempels van events kunnen gebruikt worden om knelpunten op te sporen. Dit levert vaak ideeën op voor procesverbetering. Ook kunnen voorspellingen gedaan worden op basis van de historische gegevens in de event log.

Process Discovery (kader)

Het doel van process discovery is het afleiden van een procesmodel op basis van patronen in een event log. Laten we voor het gemak even uitgaan van een eenvoudig bestelproces met vijf mogelijke activiteiten: *A*, *B*, *C*, *D*, *E*. Stel dat *A* het plaatsen van een bestelling is, *B* de betaling van deze bestelling, *C*

de levering van de bestelling, en *D* het versturen van een dankbrief. Het kan echter ook zijn dat de bestelling niet leverbaar is, in dit geval wordt via activiteit *E* de klant op de hoogte gebracht. Stel dat in de database van het betreffende bedrijf over langere tijd informatie over het bestelproces is opgeslagen. Hieruit blijkt dat er drie mogelijke procespaden zijn: *ABCD*, *ACBD*, en *AED*. Het pad *ABCD* is 543 keer gevolgd, het pad *ACBD* is 378 keer gevolgd, en *AED* is 45 keer gevolgd. Het is eenvoudig in te zien dat alle bestellingen eerst worden geregistreerd (activiteit *A*) en dat er altijd aan het einde een dankbrief wordt verstuurd (activiteit *D*). Tussen *A* en *D* worden ofwel *B* en *C* uitgevoerd of alleen *E*. De volgorde tussen betaling (*B*) en levering (*C*) lijkt er niet toe te doen omdat beide sequenties voorkomen. Met behulp van process mining algoritmen is het mogelijk automatisch een procesmodel af te leiden dat als uitgangspunt voor allerlei analyses gebruikt kan worden. Het afgebeelde model laat een Petri-net zien dat ontdekt is op basis van een event log waarin alleen de procespaden *ABCD*, *ACBD*, en *AED* voorkomen. De representatie van het ontdekte proces kan eenvoudig vertaald te worden naar de gewenste modelleringtaal (bijvoorbeeld BPMN).



Process discovery is verre van eenvoudig: slimme analysetechnieken zijn nodig omdat een event log in veel gevallen slechts een fractie van de mogelijke procespaden bevat. Ook is het essentieel dat informatie uit de log op het model geprojecteerd kan worden. Het model moet knelpunten en afwijkingen kunnen laten zien, ook als deze zich verplaatsen. Op deze manier is het mogelijk te laten zien dat de gemiddelde wachttijd voor activiteit *B* erg lang is doordat betalingen vaak op zich laten wachten of dat er soms leveringen plaatsvinden (activiteit *C*) zonder een corresponderende betaling gedaan wordt (activiteit *B*).

Spaghetti en Lasagne processen

In navolging van het open-source gereedschap ProM (www.processmining.org) zijn er recent diverse softwaretools beschikbaar gekomen die process mining ondersteunen, bijvoorbeeld BPM|one (Pallas Athena), Reflect (Futura Process Intelligence), ARIS Process Performance Manager (Software AG), Interstage BPME (Fujitsu), en Enterprise Visualization Suite (Businesscape). De toepassing van dit soort gereedschappen laat zien dat er twee typerende klassen van processen zijn: *Spaghetti* processen en

Lasagne processen. Lasagne processen zijn gestructureerd waardoor het eenvoudig is het onderliggende proces te vinden. De toegevoegde waarde zit niet in het ontdekken van het proces, maar in de koppeling tussen event log en model. Hierdoor is het mogelijk afwijkingen op te sporen en te analyseren. Ook is het mogelijk de ontdekte informatie te projecteren het model. Op deze manier kunnen knelpunten, kosten en onderbezetting snel zichtbaar gemaakt worden. Spaghetti processen zijn, zoals de naam al aangeeft, minder gestructureerd, waardoor het niet eenvoudig is om het onderliggende proces te ontdekken. In veel gevallen is het echter mogelijk Spaghetti processen enorm te verbeteren door simpelweg zichtbaar te maken wat er nu echt gebeurt. Merk op dat het mogelijk is de events van een bepaalde periode (bijvoorbeeld het afgelopen jaar) af te spelen op het ontdekte model. Dit kan erg confronterend zijn omdat problemen en inefficiënties direct zichtbaar worden. Toch is dit het beste startpunt voor herontwerp of procesverbetering.

Cartografie en navigatie

Process mining kan vergeleken worden met cartografie en navigatie. Procesmodellen brengen processen in kaart net zoals een landkaart een gebied in kaart brengt. Gegeven een bepaald gebied zijn er vele kaarten denkbaar: kaarten voor de automobilist, fietser of watersporter, kadasterkaarten, en kaarten met water- en elektriciteitsvoorzieningen. Processen veranderen over het algemeen sneller dan de geografische werkelijkheid. Daarom is het van groot belang steeds up-to-date “proceskaarten” (d.w.z. procesmodellen) te hebben voor de verschillende doelgroepen. Dankzij process mining is het mogelijk deze modellen te genereren op basis van events uit de daadwerkelijke processen. Nauwkeurige landkaarten kunnen gebruikt worden om informatie op te projecteren. Denk bijvoorbeeld aan mashups gebaseerd op Google Maps: kaarten die de te boeken hotels laten zien, kaarten die de herkomst van bezoekers weergeven, enz. Een ander voorbeeld is het zichtbaar maken file-informatie op een kaart. De koppeling tussen event logs en gegenereerde procesmodellen maakt het mogelijk zowel actuele als historische informatie te visualiseren door de gegevens op het procesmodel te projecteren (bijvoorbeeld file-informatie in een bedrijfsproces). Process mining helpt ook bij het navigeren binnen en tussen bedrijfsprocessen, net zoals een navigatiesysteem van TomTom, Garmin of Navigon dit doet voor de automobilist. Door historische informatie, modellen, en actuele informatie te combineren is het mogelijk voorspellingen te doen en aanwijzingen te geven. Een navigatiesysteem laat bijvoorbeeld steeds de verwachte aankomsttijd zien. In processen kan men op dezelfde manier voorspellingen doen (bijvoorbeeld de verwachte doorlooptijd of kans op succes). Op basis van deze voorspellingen kan een informatiesysteem ook adviezen geven, net zoals een navigatiesysteem aanwijzingen geeft. Het informatiesysteem kan bijvoorbeeld aangeven in welke volgorde activiteiten uitgevoerd dienen te worden om de doorlooptijd te minimaliseren.

Over de auteur

Wil van der Aalst (www.vdaalst.com) is hoogleraar informatiesystemen aan de Technische Universiteit Eindhoven (TU/e). Daarnaast heeft hij een deeltijdaanstelling bij Queensland University of Technology (QUT) in Australië en is trekker van de IEEE Task Force on Process Mining.

Process mining boek (kader)

In april is het eerste boek over process mining verschenen bij Springer: "Process Mining: Discovery, Conformance and Enhancement of Business Processes" geschreven door Wil van der Aalst (ISBN 978-3-642-19344-6). Dit boek introduceert de belangrijkste technieken en laat zien hoe deze toegepast kunnen worden. Het boek wordt ondersteund door een verzameling presentaties, voorbeelden van event logs en software (ProM), zie <http://www.processmining.org/book/>. Hierdoor kan de lezer direct aan de slag met process mining.