



Prof.dr.ir Wil van der Aalst is hoogleraar Information Systems aan de Technische Universiteit Eindhoven (TU/e). Hiernaast is hij ook Directeur van het Data Science Center Eindhoven (DSC/e) aan dezelfde universiteit, een van de vele vermeldingen op zijn indrukwekkende CV.

Eind jaren 90 begon Prof.dr.ir. Wil van der Aalst met het meten van handelingen binnen systemen en de event logs van toen waren nog gebaseerd op handmatig ingevulde tijden op post-its op documenten. Vijftien jaar later heeft hij deze analyse op logging in informatiesystemen volledig geautomatiseerd en geprofessionaliseerd. Wil is hiermee de 'founding father' van Process Mining en heeft de hoogste citatie index onder informatici binnen Europa.

Op 14 September is Wil één van de toonaangevende sprekers op het VUroRE Seminar 'IT en Audit in het komende Decennium, Opinie van de experts'. (Process Mining en BPM helpen!). Ter voorbereiding op dit Seminar gingen VUroRE bestuursleden Arnold Westgeest en Stef Schinagl op bezoek bij de TUE/e in Eindhoven.

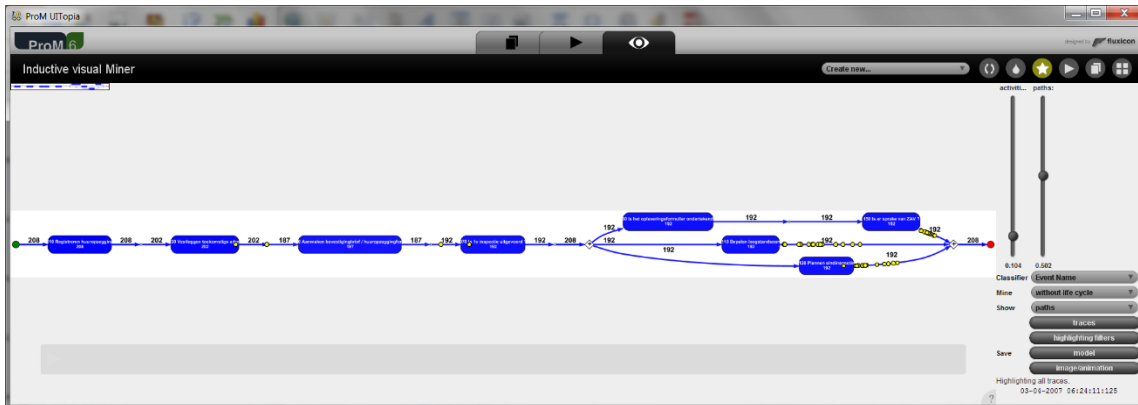
Wat is Process Mining?

Process Mining betreft het aan de hand van logs uit informatiesystemen reconstrueren en visualiseren van de procesgang.

Deze techniek is voortgekomen uit een combinatie van Data Mining (DM) en Business Process Management (BPM). DM gaat in op het opslaan en analyseren van gegevens en BPM houdt zich bezig met het modelleren en beschrijven van procesmodellen. Een combinatie van deze



vakgebieden levert procesmodellering en procesbeschrijving aan de hand van echte data op. Process Mining is hiermee een aanvulling binnen het spectrum van Business Intelligence (BI).



De vakgebieden BPM en DM hebben zich afhankelijk van elkaar ontwikkeld. Process Mining is een van de weinig technieken welke een brug legt tussen deze werelden. Zoals Wil in één van zijn YouTube video's omschrijft is het doel van Process Mining juist niet om meer data te verzamelen. Bij Process Mining is het doel om processen te verbeteren, kosten te verlagen, betrouwbaarheid van systemen te verhogen en om echt te begrijpen wat gebruikers willen.

<https://www.youtube.com/watch?v=azrwGm7BC3s>

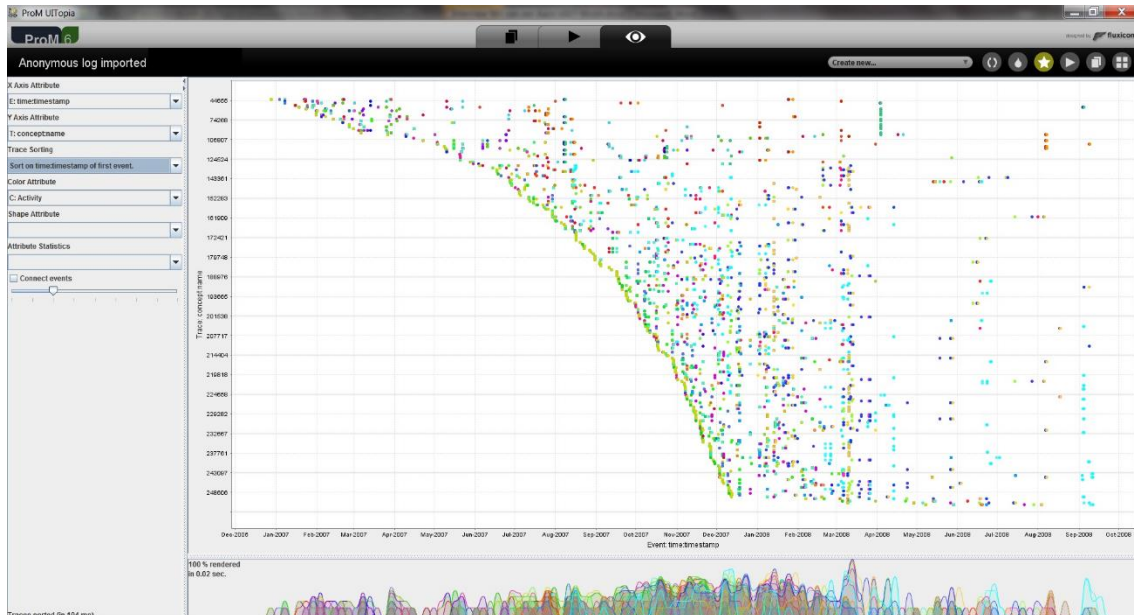
Process Mining is goed te vergelijken met het maken en gebruiken van digitale landkaarten. TomTom kan bijvoorbeeld beschikken over geavanceerde mappen met interactieve en actuele informatie. Procesmodellen kunnen worden gezien als de kaarten van het wegennetwerk binnen de TomTom, welke de operationele processen van organisaties beschrijven. Process Mining technieken maken het mogelijk om automatisch business process maps te genereren. Dit zorgt voor kwalitatief hoogwaardige procesmodellen die laten zien wat er nu *echt* gebeurt binnen een organisatie.

Door het beschikbaar hebben van deze actuele procesmodellen is het mogelijk om eventuele afwijkingen, onnodig meerwerk, en bottlenecks te signaleren net als de TomTom in staat is een file te signaleren. De combinatie van nauwkeurige kaarten, historische informatie, en informatie over de huidige procesgang, maakt het automatisch afleiden van modellen (process discovery), voorspellingen (prediction), het detecteren/voorkomen van ongewenste events (conformance checking) en prestatietesting (performance measurement) mogelijk.



Wat is het verschil tussen Process Mining en Data Mining?

Data Mining analyseert statische data: de data zoals deze op het moment van analyse zijn opgeslagen in de databases. Process Mining daarentegen, kijkt naar de totstandkoming van deze data, de events waarmee een end-to-end beeld ontstaat van de processen. Het gaat om het inzichtelijk maken van processen, het signaleren van afwijkingen, het vinden van onnodig werk en de bottlenecks, waardoor het proces steeds beter kan worden uitgevoerd. Data Mining beperkt zich hierbij alleen tot analyse van de resultaten van het proces.



Als voorbeeld een webshop: met behulp van Data Mining kan het verband tussen aanschaf van verschillende producten inzichtelijk worden gemaakt: Als een klant product X koopt, is het waarschijnlijk ($x\%$ kans) dat deze zelfde klant ook product Y koopt en vice versa. Echter, de volgorde van aanschaf van product X en Y wordt hierbij niet overwogen. Met behulp van Process Mining kan ook de volgorde van aankopen worden geanalyseerd en een nauwkeurigere inschatting worden gemaakt van de kans dat aankopen van producten met elkaar samenhangen, afhankelijk van de volgorde waarin klanten aankopen doen. Dit levert een nog betere inschatting op van de kans dat een klant geïnteresseerd zal zijn in een bepaald product, waarmee klant-specifieke reclame nog effectiever en efficiënter kan worden ingezet.

Waarom is Process Mining voor de auditor van belang?

Bij auditing is het van belang om te kijken of het gerapporteerde overeenkomt met de (gewenste) werkelijkheid. Confrontatie van gedrag, normatief versus werkelijkheid, is hierbij van belang.

Process Mining stelt auditors in staat om integraal aan de hand van logs te observeren hoe een applicatie zich gedraagt en hoe de gebruikers binnen deze applicatie handelen. Dit in tegenstelling tot het inspecteren van configuraties op basis van steekproeven en op geaggregeerd niveau, waarbij de auditor aan moet nemen dat de applicatie zich volgens deze configuraties zal gedragen.



Uit de vele in de praktijk uitgevoerde Process Mining projecten is ook gebleken dat afwijkingen op veronderstelde normen voor processen eerder regel dan uitzondering zijn. Echter, afwijkingen blijken bij nader onderzoek niet altijd problematisch te zijn. Met andere woorden, een afwijking op het normatieve model betekent niet dat er sprake is van een absolute fout. Het is een signaal voor de discussie: Is het model fout of moeten de processen binnen de organisatie

worden verbeterd? Soms zijn afwijkingen juist goed, omdat deze blootleggen dat er iets mis is met het systeem of de procedure. Afwijkingen kunnen juist verbeteringen in de systemen en processen impliceren.

Een vergelijkbaar fenomeen zijn de niet officiële paden die door gebruikers in de loop der tijd worden gecreëerd, de zogenaamde 'Desire Lines' of 'Olifantenpaden'. Eindhoven kende bijvoorbeeld een dergelijk pad vlakbij het station (het "libopad"). De vele voetgangers en fietsers staken diagonaal een grasveld over om sneller het station te bereiken, waardoor een onnatuurlijk pad werd gecreëerd. Het duurde jaren voordat het pad echt werd geasfalteerd, maar de behoefte was duidelijk. Afwijkingen kunnen dus heel positief zijn.

Welke Skills zijn nodig voor het uitvoeren van Process Mining?

Het antwoord op deze vraag hangt nauw samen met een snel opkomend vakgebied: Data Science. Een Data Scientist is er enerzijds in gespecialiseerd om data uit systemen om te zetten naar analyseerbare data en anderzijds om zinvolle analyses op deze data uit te voeren.

Als je data eenmaal analyseert, dan kom je tot de conclusie dat de werkelijkheid complex en weerbarstig is. Daarom is het van belang om met de juiste attitude naar de data te kijken: Wat willen wij weten en hoe moet de beschikbare data worden gefilterd, zodat kan worden ingezoomd op zaken die relevant zijn? Ook dit is de belangrijkste vaardigheid van een Data Scientist.

Je moet dus domeinkennis hebben om juiste vragen te stellen, maar ook de skills om de vraag snel en juist te kunnen beantwoorden. Op basis van een vraag moet een Data Scientist direct een query kunnen draaien: 'Ik wil graag die set gegevens, over die populatie, rond dat tijdstip'.

Het gaat niet alleen om het eenmalig uitvoeren van de analyse, maar ook om de iteraties van de analyse tijdens het proces. Deze iteraties moeten uiteindelijk leiden tot een gewenst abstractieniveau voor presentatie aan de gebruikers.

Bent u niet opzoek naar het schaap met vijf poten?

Wellicht is het zoeken naar het schaap met vijf poten, maar de huidige professionals die vaak al vele jaren in het veld werken zijn opgeleid in de tijd waar de data niet in de huidige verschijningsvorm en omvang aanwezig was. Je kunt van deze groep professionals dus ook niet verwachten dat zij uitputtende kennis hebben op dit gebied. Echter, de studenten van vandaag (en dus de professional van morgen) gaan heel anders om met data. Het is bijna gewoon geworden en zij zijn er mee groot geworden.

Dit is bijvoorbeeld terug te zien binnen het Data Science Center Eindhoven (DSC/e) met ruim 400 onderzoekers verdeeld over 28 onderzoeksgroepen. Deze groepen werken op verschillende gebieden: statistiek, Data Mining, Process Mining, algoritmen, optimalisatie, operations management, intelligent design, wearables, smart grids, marketing, ethiek, etc. Ook binnen de Sociale Wetenschappen zien we een revolutie: Waar men eerst het gedrag observeerde in laboratorium-



achtige omgevingen, wordt gedrag nu geanalyseerd via mobile devices en andere tools. Met andere woorden, de studenten leren onderzoeken aan de hand van echte velddata in plaats van kunstmatige experimenten of vragenlijsten.

Daarnaast is er altijd sprake van een team, waarbij de expertises over verschillende personen zijn verspreid. Kennis van elkaars domein is wenselijk, want bijvoorbeeld de vraag programmeren zonder te begrijpen wat precies de vraag is, is zinloos. Echter, niet alle kennis hoeft bij één persoon te zijn belegd.

Welke opleidingen zijn er op het gebied van Data Science?

Op de Technische Universiteit Eindhoven (TU/e) zijn er al twee Master opleidingen op het gebied van Data Science (een Europese en een Eindhovense variant). Hier komt binnenkort ook een volwaardige Bachelor Data Science bij. Deze Bachelor opleiding is een initiatief samen met de universiteit van Tilburg. Het animo hiervoor lijkt groot en verwacht wordt dat de Bachelor opleiding op het gebied van Data Science op termijn groter wordt dan de Wiskunde en Informatica opleidingen. In Amerika is dit al een feit.

Op de TU/e is Wil Directeur van het Data Science Centre Eindhoven (DSC/e). Dit is één van de grootste initiatieven binnen het vakgebied. Hierbinnen zijn meer dan 25 onderzoeksgroepen en 400 mensen actief. Vanuit het DSC/e wordt ook aansluiting gezocht met het bedrijfsleven. In samenwerking met de Tilburg University is een initiatief gestart om in Den Bosch een nieuwe "data science universiteit" op te zetten. De nadruk ligt hier op ondernemerschap en het aanjagen van nieuwe economische activiteiten rondom data science.

De auditors die zich willen bijscholen op het onderwerp moeten nog even geduld hebben. Door de snelle ontwikkelingen ligt de prioriteit nog niet op avondopleidingen en trainingen. Een alternatief is het volgen van Massive Open Online Courses (MOOC). Bijvoorbeeld de online data science cursussen aangebonden via het Coursera platform. Wil biedt via Coursera ook de cursus "Process Mining: Data Science in Action" aan welke al door meer dan 68.000 mensen is gevolgd. Op 7 oktober starts deze cursus opnieuw (aanmelden via <https://www.coursera.org/course/procmin>).

Welke tools zijn er beschikbaar op de markt voor Process Mining?

In de academische gemeenschap is ProM de-facto standaard en het gereedschap is al meer dan 100.000 maal gedownload. (ProM is open-source en gratis te downloaden). ProM kent meer dan 1000 soorten analyses en is daarmee het meest volledige raamwerk voor Process Mining. Echter, door de vele analysemogelijkheden vergt ProM relatief veel technische kennis en is daarmee complex en lastig voor mensen met weinig ervaring.

Naast ProM zijn er diverse eenvoudigere tools beschikbaar, zoals Disco, Perceptive en Celonis met elk hun eigen voor- en nadelen. Deze tools moeten worden gekocht en zijn dus niet gratis. Echter, zij zijn voor niet-ervaren gebruikers veel gemakkelijker toe te passen.

Voor de functionaliteit in de 'commerciële' tools, in het bijzonder voor conformance checking, valt winst te behalen: Deze laten een aantal significante analysemogelijkheden die ProM wel biedt, nog buiten beschouwing. Het is aan de gebruikers, waaronder auditpartijen, om ervoor te zorgen dat deze functionaliteit ook in de 'commerciële' tools beschikbaar komt: Zij moeten hierom vragen.



Ik voorspel dat het nog maar een kwestie van tijd is dat Process Mining haar intrede doet in het auditvak. Auditen gaat om het onderzoeken van de werkelijkheid en
Process Mining is real!

Prof. dr. ir. Wil van der Aalst is voltijd hoogleraar Information Systems aan de Technische Universiteit Eindhoven. Daarnaast, is hij wetenschappelijk directeur van het Data Science Center Eindhoven. Sinds 2003 heeft Wil een deeltijd aanstelling aan Queensland University of Technology. Zijn interesses gaan uit naar workflow management, process mining, Petri nets, business process management, procesmodering, en simulatie. Wil heeft meer dan 600 artikelen gepubliceerd en heeft diverse invloedrijke boeken geschreven. Wil is een van de meest geciteerde computer wetenschappers ter wereld. Zijn ideeën hebben een grote invloed gehad op standaardisatieprocessen en commerciële softwareproducten. Hij is verkozen tot lid van de KNAW, KHMW, Academia Europaea en heeft eredoctoraten ontvangen van diverse universiteiten (Tsinghua University, Universiteit Hasselt en de Higher School of Economics in Moscow).