

“Het beste idee van 2013”

Olifantenpaden in Data

Prof.dr.ir. Wil van der Aalst

In steden en natuurgebieden laten olifantenpaden zien welke wegen mensen echt volgen. Een olifantenpad is een niet officieel fiets- of wandelpad dat mensen creëren als ze een stukje van de officiële route afsnijden. Doordat ze hun sporen in het gras of plantsoen achterlaten wordt het werkelijke gedrag zichtbaar. Stedenbouwkundige vergissingen worden genadeloos zichtbaar gemaakt door deze olifantenpaden. De breedte en diepte van het olifantenpad vertellen iets over de frequentie van gebruik. Ondanks het gebruik van hekken en andere hindernissen, zijn mensen creatief in het vasthouden van de “ideale lijn”. Dwight Eisenhower, de latere president van de VS, gebruikte dit verschijnsel toen hij president van Columbia University was. Hij nam de campus in gebruik zonder trottoirs aan te leggen. Pas toen de olifantenpaden zich openbaarden en duidelijk was hoe mensen liepen werden de trottoirs aangelegd. Parken zoals Central Park in New York zijn verbeterd door de olifantenpaden te formaliseren.

Op dit moment staat “Big Data” volop in de aandacht. Het is voor iedereen duidelijk dat de exponentiële groei van gedigitaliseerde data en rekenkracht grote invloed zal hebben op onze maatschappij. Alleen organisaties die slim gebruik maken van de stortvloed aan gegevens over het gebruik van producten en diensten zullen overleven. De olifantenpaden die gevonden kunnen worden in “Big Data” laten het daadwerkelijke gedrag van mensen, systemen, en organisaties zien. Moderne process mining technieken zijn in staat procesmodellen te genereren op basis van events. Hierdoor kan event data omgezet kunnen worden in nieuwe verrassende inzichten.

Uitgangspunt voor process mining zijn de zogenaamde event logs. Events worden beschreven door de casus waartoe ze behoren (bijvoorbeeld een patiënt), de naam van de activiteit, een tijdstempel, en een willekeurig aantal aanvullende attributen (bijvoorbeeld de naam van de arts of leeftijd van de patiënt). 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).

Process mining technieken kunnen dus digitale olifantenpaden zichtbaar maken die laten zien wat mensen, systemen en organisaties daadwerkelijk doen. Het process mining spectrum is echter veel breder dan het ontdekken van olifantenpaadjes. Het is bijvoorbeeld mogelijk om te ontdekken waarom mensen afwijken van het officiële pad, en waarom delen van het proces langer duren. Olifantenpaden kunnen gebruikt worden om informatiesystemen te verbeteren en mensen op een betere manier aan te sturen.

Biografie

Prof.dr.ir. Wil van der Aalst is als hoogleraar verbonden aan de Technische Universiteit Eindhoven (TU/e). Internationaal wordt hij gezien als de expert op het gebied van BPM. Hij heeft diverse doorbraken op zijn naam staan op gebieden uiteenlopend van procesmodellering en simulatie tot workflow management en process mining. Volgens Google Scholar is hij de meest geciteerde Europese informaticus (op basis van h-index). Ook heeft hij aan de wieg gestaan van belangrijke open-source initiatieven zoals YAWL en ProM.