

Kleinduimpje in Workflowland

Wil van der Aalst en Ton Weijters
Faculteit Technologie Management
Technische Universiteit Eindhoven
Postbus 513
5600 MB Eindhoven

De huidige generatie informatiesystemen wordt in toenemende mate gedreven door software die geconfigureerd wordt op basis van procesmodellen. Workflow management systemen, maar ook Enterprise Resource Planning (ERP) en Customer Relationship Management (CRM) systemen, pompen werk rond uitgaande van een blauwdruk van het beoogde proces. In veel gevallen wijkt de blauwdruk af van het feitelijke proces of is het feitelijke proces verre van optimaal. Om dit te doorbreken is het verstandig zo nu en dan de zaak om te draaien en uit de feitelijke gebeurtenissen af te leiden wat nu het werkelijke proces is. Workflow mining is een techniek om uit automatisch geregistreerde informatie (de zogenaamde event of workflow logs) procesmodellen te distilleren. Deze modellen kunnen worden gebruikt om systemen te configureren, processen te verbeteren en afwijkingen tussen het ontwerp en de werkelijkheid vast te stellen. Een van de gereedschappen voor workflow mining heet “Little Thumb” in analogie naar het bekende sprookje.

Het sprookje

Er waren eens een houthakker en zijn vrouw die zeven zonen hadden. Ze waren heel arm en besloten hun zonen achter te laten in het bos. Kleinduimpje, die alles had gehoord, stopte steentjes in zijn zak. Hij strooide onderweg de steentjes uit zodat hij de weg terug kon vinden naar huis. De volgende dag bracht de houthakker de zonen nog dieper het bos in. Deze keer strooide Kleinduimpje stukjes brood. Maar aan het eind van de avond kon Kleinduimpje de broodstukjes niet meer vinden omdat sommige waren opgegeten door de vogels. Als gevolg hiervan werden de zeven zonen gevangen genomen door een reus. Na wat omzwervingen met de zevenmijlslaarzen van de reus weet Kleinduimpje echter de situatie ten goede te keren.

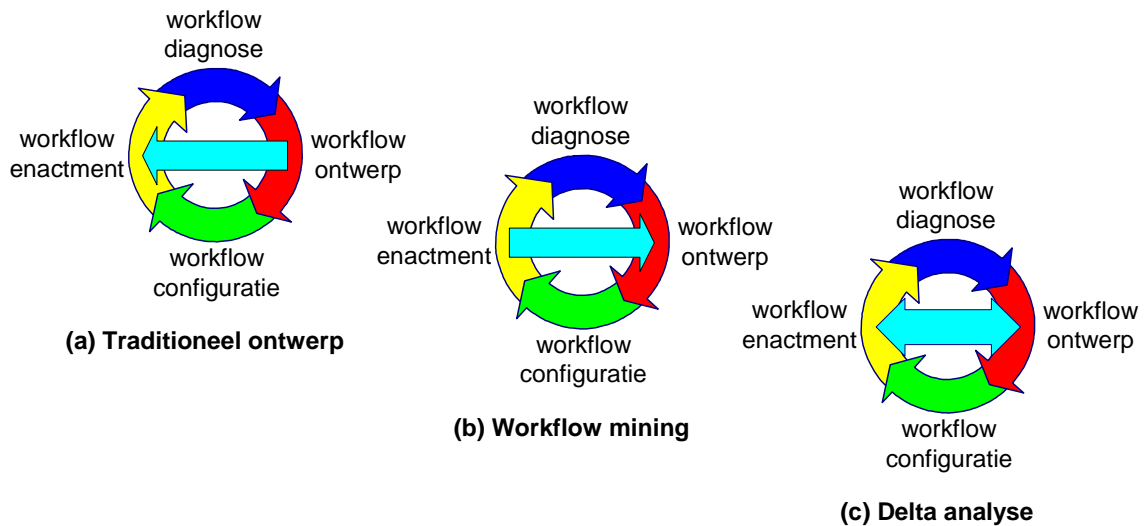
In het sprookje worden sporen in het bos achtergelaten om later de weg terug te vinden. Op eenzelfde wijze laten informatiesystemen sporen na van de bedrijfsprocessen die ze ondersteunen. Vrijwel elk bedrijfsinformatiesysteem registreert events (gebeurtenissen) in een zogenaamd “event-log”. In de event-log is terug te vinden in welke volgorde bepaalde stappen in het proces zijn uitgevoerd. Naast de volgorde wordt vaak ook additionele informatie geregistreerd (bijvoorbeeld het tijdstip, door wie, enzovoorts). De sporen in zo’n event-log kunnen vergeleken worden met de sporen die Kleinduimpje in het bos achterliet. Soms zijn de sporen duidelijk en volledig zoals de steentjes in het verhaal van Kleinduimpje. Soms zijn de sporen onduidelijk en onvolledig zoals de stukjes brood die in het sprookje Kleinduimpje niet naar huis leiden. In het laatste geval zijn er echter heuristieken (“Rules of Thumb”) te bedenken die uit schijnbaar onvolledige informatie toch waardevolle structuren in het onderliggende bedrijfsproces af kunnen leiden. Gezien de alomtegenwoordige beschikbaarheid van event-logs is de praktische relevantie van

technieken voor het destilleren van procesmodellen uit deze logs erg hoog. Daarnaast levert het onderwerp uitdagende wetenschappelijke problemen op. Het is verre van triviaal om op volautomatische wijze zoveel mogelijk informatie uit event-logs te halen/induceren?.

Workflow Mining

In het afgelopen decennium zijn er vele workflow management systemen beschikbaar gekomen. Deze systemen worden op dit moment vooral gebruikt door grote administratieve organisaties zoals banken, verzekeringsmaatschappijen, uitvoeringsinstanties en overheden. Ook al hebben workflow management systemen nog een bescheiden verspreidingsgraad, het is in de afgelopen jaren duidelijk geworden dat workflow technologie een integraal onderdeel zal zijn van de informatiesystemen van morgen. In ERP-systemen zoals SAP, Baan, Peoplesoft en JD Edwards, maar ook software voor bijvoorbeeld CRM, E-commerce en call centers, vinden we workflow componenten terug.

Kenmerkend voor workflow management systemen is dat ze casusgeoriënteerd zijn. Dit wil zeggen dat alle handelingen die uitgevoerd worden in het kader van een specifieke casus (bijvoorbeeld een aangifte, claim, klacht, bestelling, sollicitatie, enzovoorts) uitgevoerd worden. Voor elke casus moeten taken uitgevoerd worden. Deze taken zijn de stappen die gezet moeten worden om een casus van begin tot eind af te handelen. De volgorde van taken kan vast zijn. Het is echter ook mogelijk dat er keuzes gemaakt moeten worden in het proces waardoor bepaalde taken niet of juist wel uitgevoerd worden. Verder is het mogelijk dat taken niet in een vaste volgorde uitgevoerd moeten worden. In het laatste geval spreken we van parallelle routing. Voorbeelden van parallelle routing zijn het versturen van een bestelling en het versturen van de rekening waarbij de onderlinge volgorde er niet toe doet of het afnemen van verschillende onderzoeken (bloed, ECG, enzovoorts) in de polikliniek van een ziekenhuis. Over het algemeen zijn processen met zowel parallelle als conditionele stromen erg lastig in kaart te brengen. Bedrijfsprocessen kunnen honderden taken bevatten. Voor het ontwerpen, analyseren en ondersteunen van sequentiële processen is dit geen probleem. Op het moment dat meer geavanceerde routingsvormen een rol gaan spelen is dit echter verre van triviaal. Vanwege deze complexiteit en het feit dat het werk van mensen moeilijk te vangen is in een eenvoudig procesmodel zijn in het verleden vele workflow projecten mislukt. Een workflow management systeem, maar ook andere informatiesystemen die bedrijfsprocessen ondersteunen, worden immers geconfigureerd op basis van een procesmodel. Als het procesmodel niet klopt ontstaat er een mismatch tussen de wijze waarop het systeem het proces bestuurt en hoe het proces feitelijk werkt. Het resultaat van een dergelijke mismatch is vaak een slecht werkend informatiesysteem dat niet door de organisatie wordt geaccepteerd.



Figuur 1: De positionering van workflow mining en Delta analyse ten opzichte van traditioneel workflow ontwerp.

Figuur 1(a) laat de traditionele ontwerpgerichte aanpak zien. Op basis van een workflow ontwerp wordt het workflow management systeem (of een willekeurige andere applicatie voor procesondersteuning) geconfigureerd. Het aldus geconfigureerde systeem wordt gebruikt voor workflow enactment waarbij de term enactment duidt op de daadwerkelijke uitvoering van het proces onder de supervisie van het workflow systeem. Zoals is aangegeven legt het workflow ontwerp de spelregels op die tijdens de daadwerkelijke uitvoering gevolgd moeten worden. De traditionele ontwerpgerichte aanpak biedt weinig aandacht aan een grondige diagnose van het operationele proces. De workflow diagnose beperkt zich meestal tot het registreren van signalen van externe actoren (“Wat duurt het toch lang!”) of interne actoren (“We krijgen het werk niet gedaan!”). Deze signalen zijn vaak weer aanleiding tot een herontwerp om zo de cirkel van ontwerp, configuratie, enactment, en diagnose te sluiten. In veel gevallen wordt het workflow ontwerp gemaakt door consultants die op basis van informatie van het management de processen in kaart brengen. Het gevaar van deze aanpak is dat de perceptie van het management en de betrokken consultants af kan wijken van de werkelijkheid. Het gevolg is een geïdealiseerd of sterk subjectief workflow ontwerp. Om te voorkomen dat er een geïdealiseerd of subjectief workflow ontwerp gebruikt wordt voor de configuratie van het workflow systeem, worden vaak gebruikers betrokken bij de validatie van het ontwerp. Voor veel eindgebruikers is het echter lastig in te schatten in hoeverre het model de werkelijkheid beschrijft. Ook bij participatie van eindgebruikers tijdens het ontwerpproces is het lastig te komen tot een procesmodel dat daadwerkelijk de juiste gang van zaken specificiert.

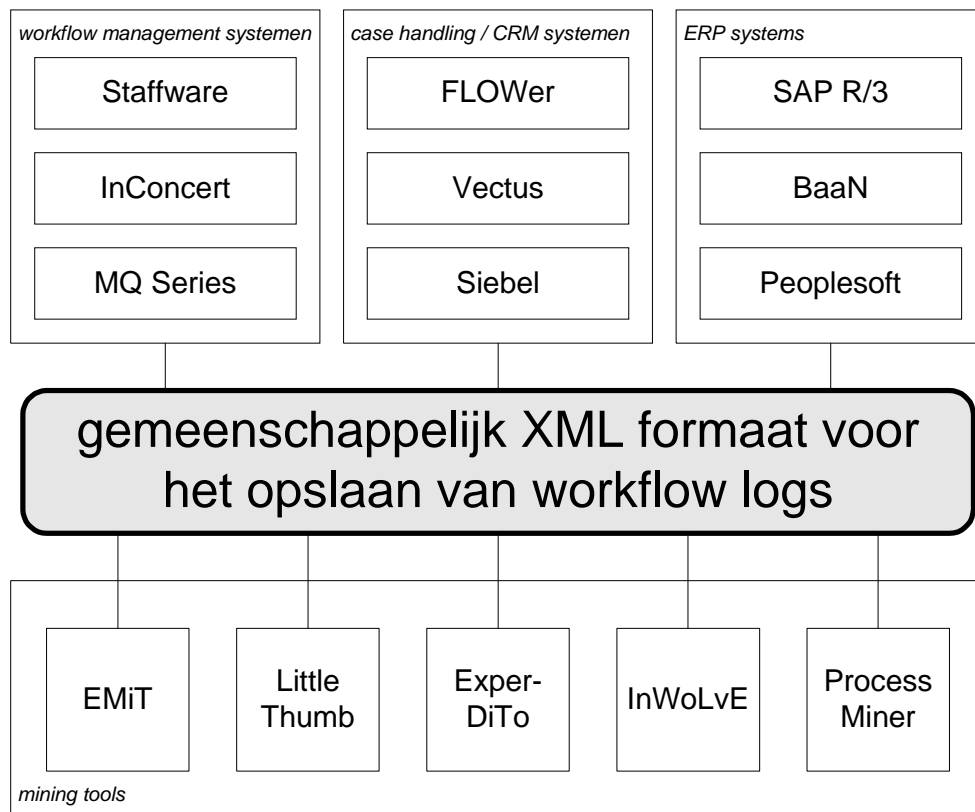
Gegeven deze problemen is het interessant de pijl in Figuur 1(a) om te draaien en uit de feitelijk uitvoering van proces een overeenkomstig workflow ontwerp te destilleren. Door het omkeren van deze pijl wordt de workflow ontwerp fase gevoed door de workflow enactment fase zoals aangegeven in Figuur 1(b). We spreken in dit geval over *workflow mining*. Workflow mining gaat uit van een zogenaamde *workflow log*. Een workflow log bevat een overzicht van alle gebeurtenissen die in een bepaald tijdvenster hebben plaatsgevonden. Deze gebeurtenissen noemen we *events* en elke event komt overeen met de uitvoering van een taak voor een specifieke

casus (daarom noemen we workflow logs ook wel event logs). Behalve de taak en de casus kan een event ook een tijdstempel hebben en andere attributen zoals de persoon die de taak heeft uitgevoerd. De events die betrekking hebben op een specifieke casus beschrijven als het ware de levenscyclus van een case. Op het moment dat een workflow log de levenscycli van een grote hoeveelheid cases bevat, is er voldoende informatie om semi-automatisch een workflow model te construeren. Workflow mining is interessant voor het analyseren van bestaande processen. Voor de introductie van nieuwe technologie of nieuwe processen kan bekeken worden wat er feitelijk gebeurt. Op deze manier kan concrete invulling gegeven worden aan de diagnose fase. Hierdoor kan een workflow ontwerp gemaakt worden dat objectief is en niet onnodig geïdealiseerd.

Workflow mining gaat uit van de situatie dat uit het geregistreerde feitelijke gedrag een procesmodel wordt afgeleid. Op het moment dat het feitelijke gedrag zoals geregistreerd in de workflow log vergeleken wordt met het procesontwerp spreken we over *Delta analyse*. Voor Delta analyse kan gebruik gemaakt worden van dezelfde technieken als voor workflow mining mits er een mechanisme is om procesmodellen met elkaar te vergelijken. Op het moment dat er grote afwijkingen geregistreerd worden is het interessant om te kijken of het workflow ontwerp moet worden bijgesteld of dat de aansturing van het proces moet worden verbeterd. Met name in systemen die veel flexibiliteit bieden (denk hierbij bijvoorbeeld aan case handling tools als FLOWer) is het interessant de afwijkingen tussen het feitelijke en normatieve gedrag in kaart te brengen. Dit is zeer zinvolle informatie voor een eventueel herontwerp.

XML formaat

Op dit moment zijn wereldwijd een aantal groepen bezig met het onderwerp workflow mining. De toepassing van workflow mining is niet beperkt tot omgevingen waar workflow management systemen gebruikt worden. In principe kan elk informatiesysteem dat gebeurtenissen registreert als basis dienen voor workflow mining. Het probleem is natuurlijk dat elk systeem events weer op een verschillende manier opslaat. Door de verschillende formaten en verschillende technieken voor workflow mining ontstaat een complexe situatie die praktische toepassing in de weg staat. Om de krachten te bundelen hebben een aantal partijen een standaard XML formaat afgesproken. Door het formaat te definiëren op basis van XML technologie is het voor alle partijen eenvoudig workflow logs te creëren en te lezen. Figuur 2 laat zien hoe het formaat dient als een koppelpunt tussen (standaard) informatiesystemen die gebeurtenissen registreren en gereedschappen voor workflow mining.



Figuur 2: Het systeemafhankelijke XML formaat ontkoppelt operationele informatiesystemen en mining tools waardoor onnodige conversieslagen worden voorkomen.

Het voert te ver om in te gaan op de details van het XML formaat. Voor een DTD (Document Type Definition) van het formaat verwijzen we naar: <http://www.tm.tue.nl/it/staff/wvdaalst/workflow/mining/>. Zoals aangeven in Figuur 2 zijn er een aantal mining gereedschappen beschikbaar: EMiT, Little Thumb, ExperDiTo, InWoLvE en Process Miner. Van deze gereedschappen zijn de eerste drie ontwikkeld aan de Technische Universiteit Eindhoven. EMiT is gebaseerd op Petri net theorie en probeert, uitgaande van betrouwbare informatie, uit een workflow log een workflow model te destilleren. In tegenstelling tot de andere tools neemt EMiT ook tijdsinformatie mee in de analyse en spoort op deze wijze bijvoorbeeld knelpunten in het workflow proces op. Little Thumb en ExperDiTo zijn ook gebaseerd op Petri netten maar richten zich op de situatie waar de workflow logs incompleet zijn en/of fouten bevatten. We spreken in dit verband ook wel over workflow logs met “ruis” (noise). Little Thumb richt zich op heuristieken die kunnen omgaan met deze ruis. ExperDiTo is vooral gericht op het meten van de kwaliteit van gevonden workflow modellen. Het gereedschap InWoLvE is ontwikkeld binnen DaimlerChrysler en richt zich op workflows waarin taken meerdere malen voorkomen. De Process Miner van OFFIS beperkt zich tot workflow modellen met een zogenaamde blok-structuur. Het gaat te ver om elk van deze gereedschappen in detail te beschrijven. Wel is het van belang in te zien dat deze gereedschappen elkaar aanvullen en dat daarom een standaard formaat van groot belang is.

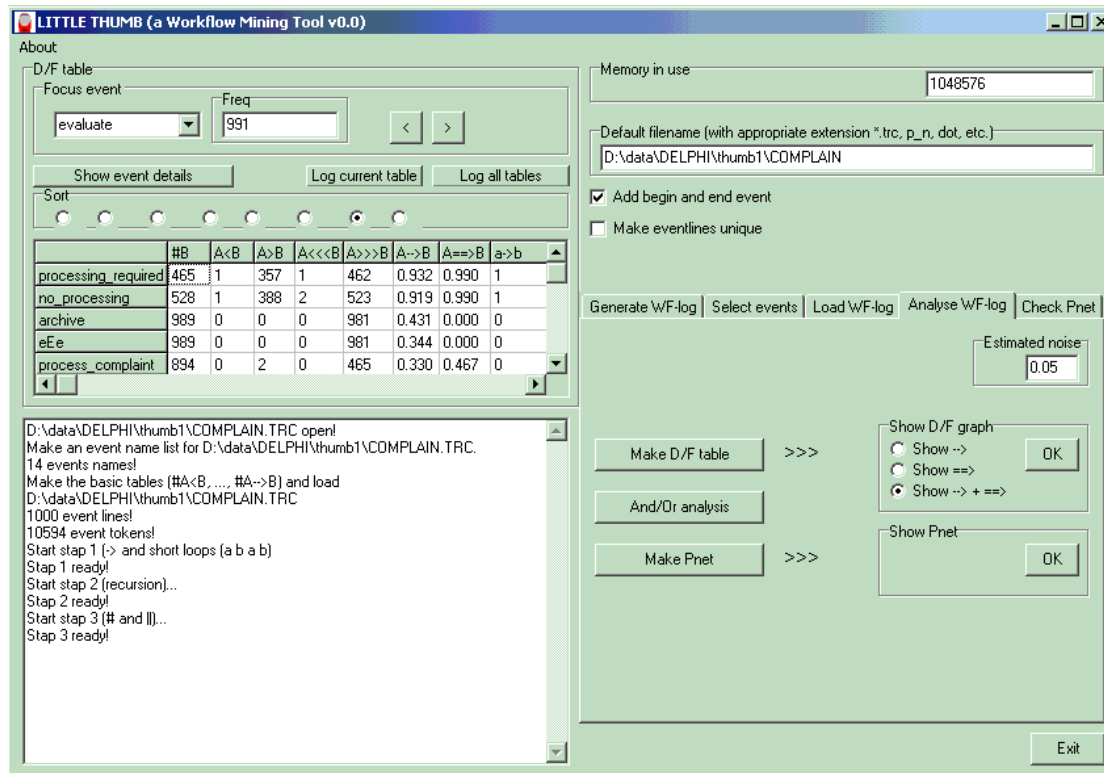
Conclusie

In het sprookje was Kleinduimje niet in staat om de weg naar huis terug te vinden. In Workflow land lijkt het echter heel wel mogelijk te zijn om met behulp van

geavanceerde mining-technieken bedrijfsprocessen te ontdekken in informatiebestanden. Door gebruik te maken van een standaard XML formaat en de beschikbaarheid van gereedschappen als EmiT en Little Thumb liggen de toepassingen voor het oprapen. Dit neemt niet weg dat er nog een aantal wetenschappelijke uitdagingen zijn. Met name vragen als “Hoe om te gaan met ruis?” en “Waar vinden we de juiste gegevens in ons informatiesysteem?” zijn verre van triviaal en blijven onderwerp van verder onderzoek.

Little Thumb

Little Thumb een gereedschap om uitgaande van een workflow log een workflow model te destilleren. Zo'n workflow log mag fouten (noise) bevatten en kan onvolledig zijn. In Figuur 3 zien we een schermafdruck van Little Thumb.



Figuur 3: Een schermafdruck van Little Thumb.

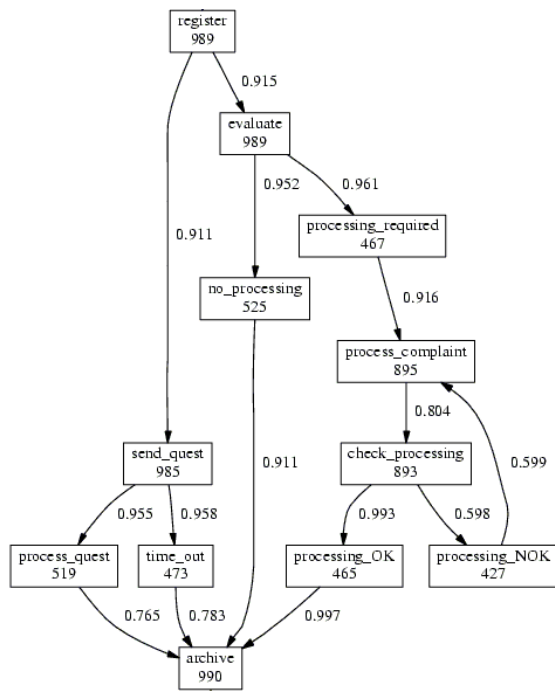
In dit voorbeeld onderneemt Little Thumb een poging een workflow log te analyseren met daarin informatie over de uitgevoerde taken bij de afhandeling van klachten (complaints). De workflow log bevat voor 1000 klachten de registratie van de uitgevoerde taken. Echter 5% (50 cases) van de registraties bevatten fouten. Er ontbreekt een stuk van de registratie van de betreffende case of de volgorde waarin de taken zijn uitgevoerd is foutief geregistreerd. We volgen Little Thumb bij zijn poging om uit deze gegevens een workflow model te destilleren. Dit verloopt grofweg in drie stappen. Stap (i): het destilleren van een zogenaamde dependency/frequency (D/F) graaf, stap (ii): het destilleren van de (and/or) informatie voor de verschillende vertakkingen (splits) en samenvoelingen (joins) en het opstellen van het bijbehorende Petri-net en stap (iii): het uitvoeren van enkele controles waarmee mogelijke tekortkomingen van het model in vergelijking met de voorbeelden in de log kunnen worden opgespoord.

Zoals aangegeven wordt in stap (i) allereerst een D/F-graaf opgesteld. Hierbij wordt gebruik gemaakt van enkele specifiek voor dit doel ontwikkelde maten. In de schermafdruck van Figuur 1 zien we de D/F-table voor taak *evaluate*. De informatie in deze D/F-table wordt door de heuristische regels (rules of thumb) gebruikt om te bepalen of er wel of niet sprake is van een afhankelijkheidsrelatie (dependency relation) met andere taken. In ons voorbeeld resulteert dit in D/F-graaf van Figuur 4. De getallen bij de pijlen geven een indicatie van de sterkte van de

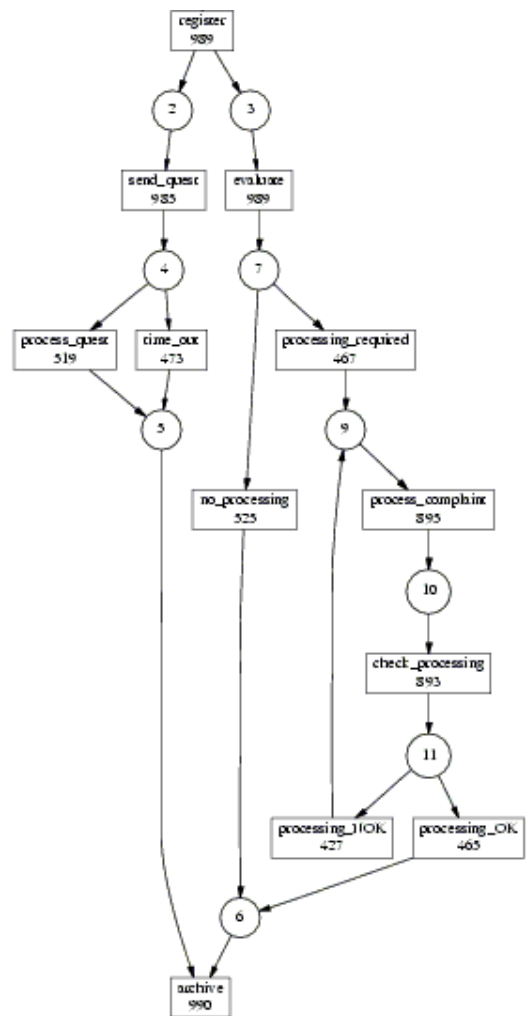
afhankelijkheidsrelatie volgens Little Thumb. Het getal bij elke taak geeft aan hoe vaak de betreffende taak is uitgevoerd. De D/F-graaf kan gebruikt worden voor een eerste validatie van het gedistilleerde workflow model. In overleg met domeindeskundige kunnen eventuele aanpassingen eenvoudig worden aangebracht.

Uitgangspunt voor stap(ii) is de D/F-graaf van stap (i). Opnieuw worden heuristische regels gebruikt om de vertakkingen en samenvoegingen te analyseren. Hierna wordt automatisch een volledig workflow model (Figuur 5) opgesteld. Hiervoor wordt het Petri-net formalisme gebruikt. In het Petri net zien we dat de vertakking van *register* naar *send_quest* en *evaluate* een zogenaamde “AND-split” is. Ook nu is het weer mogelijk om in overleg met een domeindeskundige interactief aanpassingen door te voeren.

In de laatste stap, stap (iii), worden enkele controles uitgevoerd waarmee mogelijke tekortkomingen van het model in vergelijking met de voorbeelden in de log kunnen worden opgespoord. In een eerste controle wordt iedere trace in de workflow log aangeboden aan het workflow model. Er zijn nu twee mogelijkheden: de trace kan daadwerkelijk verwerkt worden door het workflow model of de verwerking loopt ergens vast. Voor iedere trace wordt aangegeven waar deze vastloopt. Wanneer een workflow log ruis bevat zullen sommige traces vastlopen in het model. Wanneer echter veel traces rondom dezelfde taak vastlopen is dit een indicatie dat het model daar niet correct is. Bij een tweede controle wordt gebruik gemaakt de frequency-informatie in de D/F-graaf. Er wordt nagegaan of de in de D/F-graaf opgenomen informatie in overeenstemming is met de uiteindelijke structuur van het netwerk. Zo wijst de frequentie van de taken *evaluate* (989), *processing_required* (467) en *no_processing* (525) op een zogenaamde “OR-split”. Kleine afwijkingen kunnen verklaard worden door ruis, duidelijke afwijkingen wijzen op een fout in het gedistilleerde model.



Figuur 4: De door Little Thumb gevonden D/F-graaf.



Figuur 5: Het door Little Thumb gedistilleerde WF-model.

Oproep

Workflow mining heeft een grote praktische waarde en biedt diverse wetenschappelijke uitdagingen. Op dit moment zijn gereedschappen als Little Thumb en EMiT volwassen genoeg op getest te worden in de praktijk. De gereedschappen zijn toegepast op patiëntstromen in ziekenhuizen en op een administratief proces binnen een overheidsinstelling. Voor het verder uitontwikkelen van de gereedschappen is het echter noodzakelijk te beschikken over meer workflow logs in verschillende domeinen. Daarom roepen we bedrijven op om deel te nemen in verder onderzoek naar workflow mining. Van bedrijven wordt verwacht dat ze bestanden met event data aanleveren en een ruw procesmodel. Uiteraard worden alle gegevens strikt vertrouwelijk behandeld. De toegevoegde waarde voor deelnemende bedrijven is een gedetailleerd inzicht in de werkelijke gang van zaken en eventueel suggesties voor herontwerp. Geïnteresseerden dienen zich te wenden tot Ineke Withagen, I&T, TUE, 040-2472290, it@tm.tue.nl.