

Süreç Madenciliği Manifestosu

(son versiyon)¹

Wil van der Aalst, Arya Adriansyah, Ana Karla Alves de Medeiros, Franco Arcieri, Thomas Baier, Tobias Blickle, Jagadeesh Chandra Bose, Peter van den Brand, Ronald Brandtjen, Joos Buijs, Andrea Burattin, Josep Carmona, Malu Castellanos, Jan Claes, Jonathan Cook, Nicola Costantini, Francisco Curbera, Ernesto Damiani, Massimiliano de Leoni, Pavlos Delias, Boudewijn van Dongen, Marlon Dumas, Schahram Dustdar, Dirk Fahland, Diogo R. Ferreira, Walid Gaaloul, Frank van Geffen, Sukriti Goel, Christian Günther, Antonella Guzzo, Paul Harmon, Arthur ter Hofstede, John Hoogland, Jon Espen Ingvaldsen, Koki Kato, Rudolf Kuhn, Akhil Kumar, Marcello La Rosa, Fabrizio Maggi, Donato Malerba, Ronny Mans, Alberto Manuel, Martin McCreesh, Paola Mello, Jan Mendling, Marco Montali, Hamid Motahari Nezhad, Michael zur Muehlen, Jorge Munoz-Gama, Luigi Pontieri, Joel Ribeiro, Anne Rozinat, Hugo Seguel Pérez, Ricardo Seguel Pérez, Marcos Sepúlveda, Jim Sinur, Pnina Soffer, Minseok Song, Alessandro Sperduti, Giovanni Stilo, Casper Stoel, Keith Swenson, Maurizio Talamo, Wei Tan, Chris Turner, Jan Vanthienen, George Varvaressos, Eric Verbeek, Marc Verdonk, Roberto Vigo, Jianmin Wang, Barbara Weber, Matthias Weidlich, Ton Weijters, Lijie Wen, Michael Westergaard, and Moe Wynn

IEEE Task Force on Process Mining²

<http://www.win.tue.nl/ieeetfpm/>

Özet. Günümüzde bilgi, sistemlerdeki kayıtlar üzerindeki loglardan erişebilir durumdadır. Loglarda gizli olan bilgi ortaya çıkar. Süreç madenciliği teknikleri ile; süreçleri ortaya çıkartabilir, geliştirebilir veya yönetebiliriz. Süreç madenciliğine olan ilginin artmasının iki temel sebebi vardır. Süreç madenciliğine olan ilginin ilk sebebi; veri yığınları arasından süreç madenciliği teknikleri ile detaylı bilgi sağlanmasıdır. Diğer bir sebep ise, sürekli değişen rekabetçi iş ortamında süreçlerin gelişmesi için güçlü destek vermesidir. Bu manifesto, IEEE Task Force tarafından süreç madenciliği kavramını tanıtmak için oluşturulmuştur. Buna ek olarak; temel prensipler tanımlanmış ve önemli sorunlar listelenmiştir. Manifesto; yazılım geliştirme uzmanları, bilim adamları, danışmanlar, yöneticiler ve son kullanıcılar için rehber olmayı amaçlamaktadır. Manifestonun amacı; tasarım/yeniden tasarım, kontrol, destek gibi operasyonel işlerde kullanılan yeni bir araç olan süreç madenciliğinin, gelişmesini sağlamaktır.

1. Süreç Madenciliği üzerinde IEEE Görev Gücü

Manifesto bir grup insan tarafından “prensip ve niyetleri açıklayan genel bir bildirgedir”. Süreç madenciliği manifestosu IEEE Task Force üyeleri ve destekçileri tarafından yazılmıştır. Task Force’un görevi; süreç madenciliğinin araştırılmasını, geliştirilmesini, eğitimlerinin verilmesini, uygulanmasını, yayılmasını sağlamaktır.

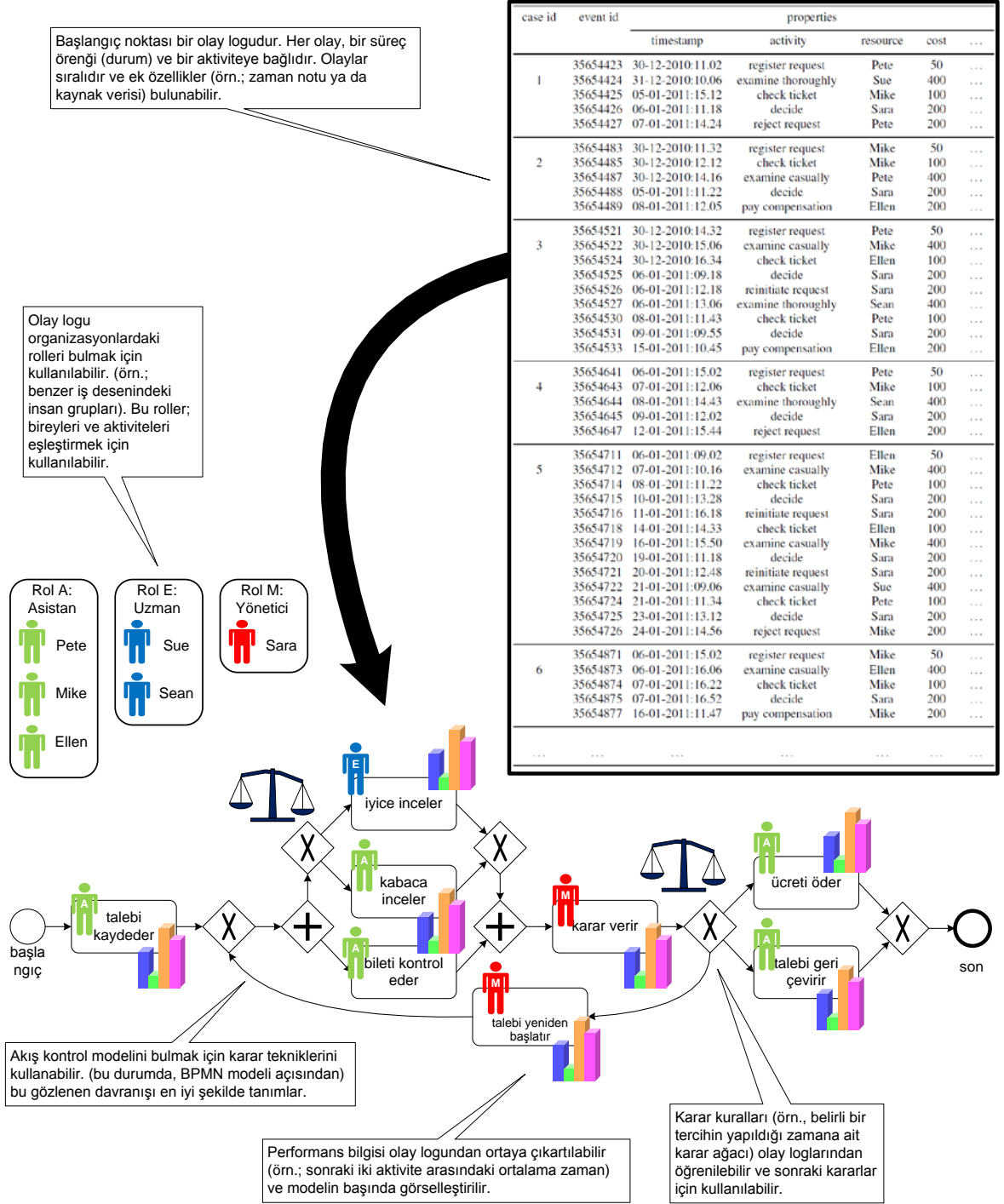
Süreç madenciliği; bir yandan süreç modellemesini ve süreç analizini, diğer yandan iş zekasını veri madenciliğini kapsayan, yeni sayılabilecek bir araştırma alanı ve disiplindir. Süreç madenciliğinin amacı; günümüz bilgi sistemlerinde kolayca erişilebilir bulunan loglardan (yani; kabul edilen süreçlerden değil) gerçek süreçleri ortaya çıkartmak, keşfetmek, kontrol etmek ve geliştirmektir. Süreç

¹ Türkçeye çeviren: Aylin Şahin (aylin.sahin@ymail.com)

² Orijinal Versiyon BPM 2011 Workshops proceedings, Lecture Notes in Business Information Processing, Springer Yayınevi, 2011 .

Süreç Madenciliği Manifestosu

madenciliğinin içerdiği konular: otomatikleşmiş süreç keşfi -automated process discovery- (diğer bir deyişle; olay logundan süreçlerin ortaya çıkarılması), uygunluk kontrolü -conformance checking- (diğer bir deyişle; model ve logun karşılaştırılması ile sapmaların kontrolü), sosyal ağ/organizasyonel madencilik -social network/organizational mining-, simülasyon modellerinin otomatikleşmek, model geliştirmek -model extension-, model tamiri yapmak-model repair-, durum tahmini yapmak -case prediction-, geçmiş bazlı öneriler yapmak.



Şekil 1. Süreç Madenciliği teknikleri olay loglarındaki bilgileri şu amaçlar için ortaya çıkarır: süreçleri keşfetmek, izlemek ve geliştirmek

Süreç Madenciliği Manifestosu

Süreç madenciliği; veri madenciliği, süreç modelleme ve analizi arasında önemli bir köprü kurar. İş zekası (business intelligence) başlığı altında bir çok raporlama ve tablo araç kavramları tanıtılır. İş Aktivite Kontrolü (Business Activity Monitoring) iş süreçlerine eş zamanlı kontrolü sağlayan teknolojiden bahseder. Karmaşık Olay İnceleme (Complex Event Processing) ise büyük miktardaki olayları kontrol edecek, yönetecek ve gerçek zamanda optimize edecek teknolojiden bahseder. Kurumsal Performans Yönetimi (Corporate Performance Management) bir süreç ya da organizasyonun performansını ölçmeyi sağlar. Sürekli Süreç Geliştirme(Continuous Process Improvement), Süreç Geliştirme (Business Process Improvement), Toplam Kalite Yönetimi (Total Quality Management) ve 6 Sigma gibi ilgili diğer yönetimsel yaklaşımlar da vardır. Bu yaklaşımların ortak özelliği süreçleri “mikroskobun altına koymak” ve herhangi bir geliştirmenin mümkün olup olmadığına bakmaktır. Süreç madenciliği yukarıda bahsettiğimiz CPM, BPI, TQM, 6 Sigma gibi konular için teknoloji sağlar.

6 sigma, TQM gibi yönetimsel yaklaşımlar ve iş zekası araçları (BI Tools); iş akışı hatalarını ve sürelerini azaltır, organizasyonel yönetime ve risk uyumuna vurgu yaparak operasyonel performansı artırır. SOX (Sarbanes –Oxley Act) ve Basel 2 Accord’ un temel konuları da uyum sorunlarıdır. Süreç madenciliği; uyum ölçümlerini titizlikle yapar ve organizasyonların ana süreçleri hakkında gerçekçi ve geçerli bilgiler sunacağını ifade eder.

Verilere erişim son on yılda oldukça kolaylaştı ve bu süre içerisinde süreç madenciliği teknikleri gelişti. Buna ek olarak, biraz önce de bahsettiğimiz yönetimsel trendler (yani; 6 Sigma, TQM, CPI, ve CPM) ve uyum ile ilgili olan SOX, BAM vb. süreç madenciliğinden faydalanabilir. Süreç madenciliği algoritmaları şimdiye kadar birçok akademik ve ticari sistemde kullanıldı. Süreç madenciliği üzerinde çalışan aktif bir araştırma grubu vardır. İş Süreç Yönetiminin (BPM) “popüler araştırma konularından” birisi de süreç madenciliğidir. Bunun yanı sıra; endüstri süreç madenciliğine büyük ilgi gösteriyor. Yazılım satıcıları her geçen gün daha fazla süreç madenciliği özelliğini, araçlarının içerisine ekliyor. Süreç madenciliği özelliği olan yazılım ürünlerine örnek olarak şunları verebiliriz: ARIS Process Performance Manager (Software AG), Comprehend (Open Connect), Discovery Analyst (StereoLOGIC), Flow (Foursparke), Futura Reect (Futura Process Intelligence), Interstage Automated Process Discovery (Fujitsu), OKT Process Mining suite (Exeura), Process Discovery Focus (Iontas/Verint), ProcessAnalyzer (QPR), ProM (TU/e),Rbminer/Dbminer (UPC), ve Reectjone (Pallas Athena). Task Force’un kurulmasını motive eden, log temelli süreç madenciliğine olan büyük ilginin artışındaki hızdır.

Task Froce 2009’da IEEE’ deki CIS ve DMTC konseptleri içerisinde kuruldu. Task Force’un üyeleri yazılım satıcılarını (örneğin; Pallas Athena, Software AG, Futura Process Intelligence, HP, IBM, Infosys, Fluxicon, Businesscape,Iontas/Verint, Fujitsu, Fujitsu Laboratories, Business Process Mining,sterelogic),danışmanlık firmaları/son kullanıcıları (örneğin, ProcessGold, Business Process Trends, Gartner, Deloitte, Process Sphere, Siav SpA, BPM Chili, BWI Systeme GmbH, Excellentia BPM, Rabobank) ve araştırma enstitülerini (örneğin; TU/e, University of Padua, Universitat Politecnica de Catalunya, New Mexico State University, IST - Technical University of Lisbon, University of Calabria, Penn State University, University of Bari, Humboldt-Universitaet zu Berlin, Queensland University of Technology, Vienna University of Economics and Business, Stevens Institute of Technology, University of Haifa, University of Bologna, Ulsan National Institute of Science and Technology, Craneld University, K.U. Leuven, Tsinghua University, University of Innsbruck, University of Tartu) temsil eder.

Süreç Madenciliği Manifestosu

Task Force'un somut hedefleri:

- Son kullanıcıların, yazılım geliştirme uzmanlarının, danışmanların, yöneticilerin ve araştırmacıların süreç madenciliğindeki en son teknolojiden haberdar olmasını sağlamak,
- Süreç madenciliği tekniklerini ve araçlarını desteklemek, yeni uygulamalara teşvik etmek,
- Olay verilerini loglanması için olan çalışmalarda görev almak,
- Eğitimler, özel sezonlar, atölye çalışmaları, paneller organize etmek ve
- Makaleler, kitaplar, videolar ve dergilerde özel sayılar yayınlamak.

Task Force; 2009'daki kuruluşunun ardından yukarıdaki hedefler doğrultusunda birçok aktivite gerçekleştiriyor. Birçok atölye çalışması ve özel bölümler, Task Force tarafından veya Task Force yardımı ile gerçekleştirilmiştir. Mesela, iş süreç zekâsı hakkındaki atölye çalışmaları (BPI'09, BPI'10, BPI'11) yapılmıştır ve ana IEEE konferanslarında özel bölümler (örneğin; CIDM'11) sunulmuştur. Süreç Madenciliği hakkındaki temel bilgi, yapılan eğitimler ile yayılmıştır (örn.; WCCT'10 ve PMPM'09). Ayrıca yaz okulları açılmıştır (ESSCaSS'09, ACPN'10, CICH'10, vb.) ve eğitim videoları yayınlanmıştır (www.processmining.org). Springer tarafından yakın zamanda basılan ilk süreç madenciliği kitabını da içeren, birçok eser yayınlanmıştır. Task Force aynı zamanda ilk Business Process Intelligence Challenge'i (BPOC'11) düzenlemiştir (düzenlemeye yardımcı olmuştur): bu yarışmada katılımcılar uzun ve karmaşık bir olay log verisinden anlamlı bir bilgiyi çıkarmak zorundaydı. 2010'da Task Force aynı zamanda XES standardizasyonunu (www.xes-standart.org) gerçekleştirmiştir. XES genişletilebilir standart bir loglama formatıdır; OpenXES kütüphanesi (www.openxes.org) ve ProM, XESame, Nitro gibi araçlarla desteklenir.

Okuyucular; Task Force aktiviteleri hakkında bilgi almak için <http://www.win.tue.nl/ieeefpm/> sayfasını ziyaret edebilir.

2. Süreç Madenciliği: En Son Teknoloji

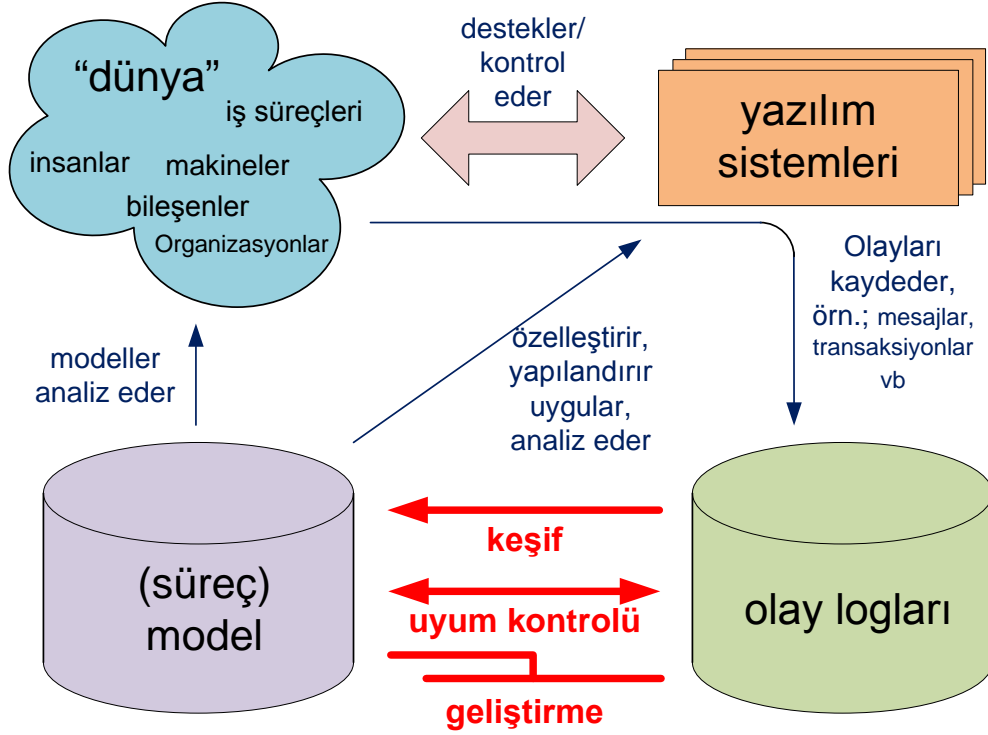
Moore kanunu bilgisayar sistemlerine bağlı olan, bilgi sistemleri ve diğer sistemlerin kapasitesini oldukça iyi tanımlar. Intelin ortak kurucusu, Gordon Moore, 1965 de devrelere eklenecek olan bileşenlerin sayısının her sene çifte katlanacağını tahmin etmiştir. Son 50 yıl boyunca büyüme biraz yavaş olsa da artarak devam etmiştir. Bu gelişmelerle "dijital dünya" çarpıcı boyutlara erişmiştir (yani, bütün veriler elektronik ortamda saklanmıştır ve/veya değiştirilmiştir). Bunun yanı sıra, dijital ve gerçek dünya her geçen zaman birbirlerine daha da yaklaşmaya devam ediyor.

Dijital ve gerçek dünyanın birbirine çok yaklaşması ile organizasyonel süreçleri de dijital ortamda kaydetmek ve analiz etmek mümkündür. Organizasyonel süreçler çok çeşitli olabilir: ATM'den para çekmek, X-Ray makinasının ayarlanması, sürücü belgesi için başvurmak, vergi beyannamesinin teslimi ya da elektronik bilet almak gibi. Zor olan, elektronik ortamda bulunan verilerden anlamlı süreçleri ortaya çıkarmaktır. Çeşitli hedefler doğrultusunda süreçler ortaya çıkartılır. Bunlardan bazıları şunlardır: süreçlere genel bir bakış akıcı sağlamak, süreçler içerisindeki kısır döngüleri bulmak, süreçlerdeki kural ihlallerini bulmak ve bu ihallere karşı önlemler sunmak ve aerodinamik süreçler keşfetmek. Süreç madenciliğinin hedefleri de tam olarak bu bilgileri ortaya çıkarmaktır.

Süreç madenciliğinin çıkış noktası olay logudur. Her olay bir aktiviteyi temsil eder, bu nedenle tüm süreç madenciliği tekniklerinin kabul ettiği, olayların sıralı bir biçimde kaydedilmesinin mümkün olduğudur (aktivite, süreçlerde iyi tanımlanmış olan adımdır). Her aktivite de, sadece belirli bir olaya yani süreç örneğine aittir. Loglarda olaylara ait ek bilgiler de kaydedilebilir. Aslında, süreç madenciliği

Süreç Madenciliği Manifestosu

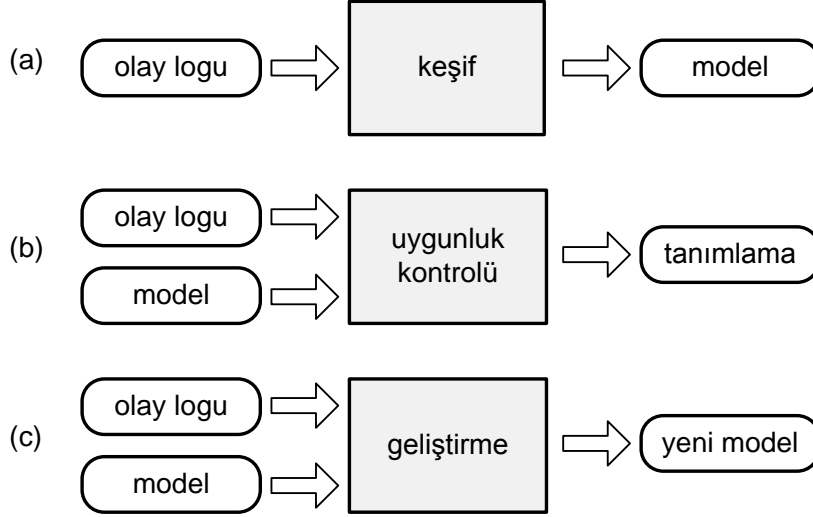
teknikleri uygulanırken, mümkün olan her yerde uygun ek bilgiler kullanılır. Örneğin; aktiviteyi başlatan ya da yürüten kaynak (yani; bir kişi ya da bir cihaz), olayların oluş ve bitiş zamanı (aktivite bazlı olabilir) ya da olayla beraber kaydedilen veri elementi (bir siparişin büyüklüğü gibi).



Şekil 2. Süreç Madenciliğinin üç ama çeşidi konumlandırılır: (a) keşif, (b) uyum kontrolü, (c) geliştirme

Şekil 2’de gösterildiği gibi, süreç madenciliği 3 çeşittir. Loglar bu üç çeşit süreç madenciliğini uygulamak için kullanılır. Süreç madenciliği çeşitlerinden ilki, keşiftir. Keşif tekniğinde öncü bilgiler olmadan, log alınır ve model oluşturulur. Süreç keşfi en önemli süreç madenciliği tekniğidir. Birçok organizasyon için, loglar üzerinden gerçek süreçlerin ortaya çıkması şarttır. Süreç madenciliğinin ikinci tipi ise uyum kontrolüdür. Uyum kontrolünde var olan süreçler ile loglardan keşfedilmiş süreçler incelenir, gerçeğin loglarda kaydedildiği gibi gerçekleşip gerçekleşmediği kontrol edilir. Uyum kontrolü farklı modeller için de gerçekleştirilebilir, örneğin; iş kuralları/poliçeleri, kanunlar, bildirilen süreç modellerinin kontrolü, organizasyonel modeller, prosedür modelleri gibi. Uygunluk kontrolü; gerçek süreçlerle model süreçler arasındaki farkı ölçer. Burada asıl amaç, gerçek süreçten gelen bilgileri kullanarak model süreci geliştirmektir. Süreç madenciliğinin üçüncü türü olan geliştirmenin amacı ise, önceden var olan modeli değiştirmek ya da geliştirmektir. Örneğin, loglardaki zaman notlarını kullanarak darboğazlar, servis seviyeleri, çıktı zamanları uygun şekilde değiştirilir.

Süreç Madenciliği Manifestosu



Şekil 3. Süreç Madenciliğinin temel üç çeşidi girdiler ve çıktılar açısından açıklanmıştır: (a)keşif, (b)uygunluk kontrolü, (c)geliştirme

Şekil 3 girdiler ve çıktılar açısından süreç madenciliği çeşidini tanımlar. Log kaydının alınıp model üretilmesi, süreç madenciliğinin keşif yöntemidir. Keşfedilen Petri ağı, BPMN, EPC ya da UML diyagramı gösteriminde olan süreç modelidir; bu model aynı zamanda sosyal ağlar gibi farklı perspektifler de içerir. Uygunluk kontrolünde girdi olarak bir log ve model gereklidir. Çıktı ise; model ve log arasındaki farklılıkları ve ortak noktaları gösteren bilgilerdir. Model geliştirmek için olan teknikler de girdi olarak log ve modele ihtiyaç duyar; çıktıları geliştirilmiş veya genişletilmiş modeldir.

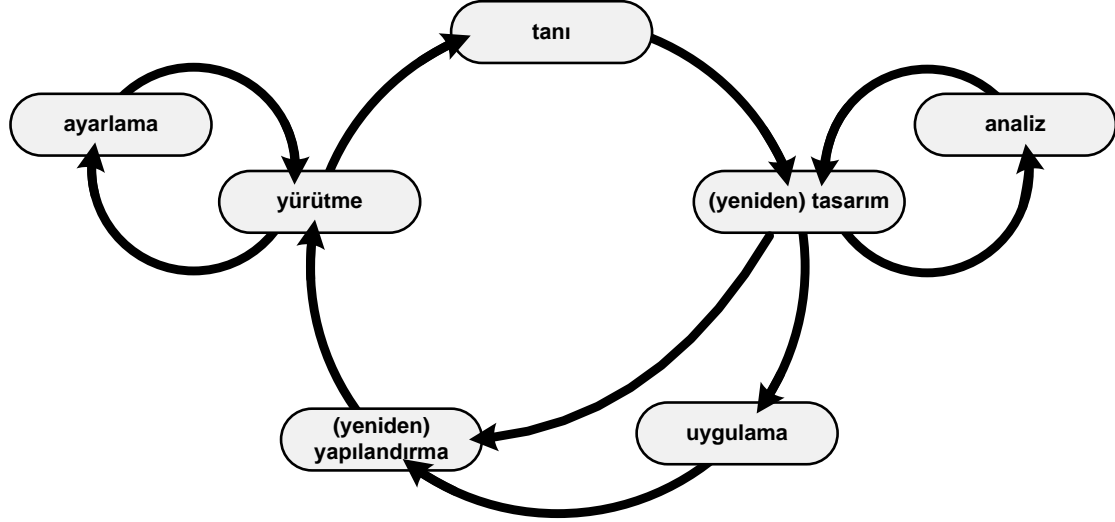
Süreç Madenciliğinde akış kontrolü, olay ve zaman gibi perspektifler de vardır. Akış kontrolü perspektifi, aktivitelerin sıralanması üzerinde odaklanmıştır. Amacı mümkün olan tüm yollar için en iyi tanımlamaları bulmaktır. Sonuçlar genellikle petri ağları olmak üzere, EPC, BPMN ya da UML diyagramları ile ifade edilir. Organizasyonel perspektif, logları gerçekleştiren aktörler (insanlar, sistemler, roller, departmanlar) ve bu aktörlerin birbirleri ile olan ilişkilerini ortaya çıkartır. Organizasyonel perspektifin amacı sosyal ağ göstermek, diğer bir ifade ile modeli organizasyon ve roller açısından sınıflandırmaktır. Olay perspektifi olayların özelliklerine odaklanır. Bir olay; süreçleri ve bu süreçleri gerçekleştiren aktörleri ile tanımlanır. Buna ek olarak belirtmeliyiz ki, veri elementlerinin değerleri ile de olaylar tanımlanabilir. Örneğin sipariş yenilemesini içeren bir olayda tedarikçiyi ya da sipariş edilen ürünün sayısını bilmek önemlidir. Zaman perspektifinin odak noktası ise olayların gerçekleşme zamanları ve sıklıklarıdır. Olaylarda zaman bilgisi varsa darboğazlar keşfedilebilir, servis seviyeleri ölçülebilir, kaynakların kullanımını izlemek mümkün olur ve devam eden olaylar için kalan zaman tahmin edilebilir.

- Süreç madenciliği iş akış kontrolü (work flow control) ile sınırlı değildir. Loglardan süreçleri keşfetmek hem akademik dünyanın hem de kullanıcıların ilgisini çekiyor. Bu nedenle iş akış kontrolü (work flow control) keşfi süreç madenciliğinin en ilgi çeken bölümüdür. Ancak süreç madenciliği bununla sınırlı değildir. İş akış kontrolü (work flow control) süreç madenciliğinin üç temel özelliğinden birisidir ancak en önemli alanı diyemeyiz. Çünkü organizasyonel, olay ve zaman perspektifleri de süreç madenciliği için önemlidir.
- Süreç madenciliği, veri madenciliğinin farklı bir türü değildir. Süreç madenciliği; veri madenciliği ve geleneksel BPM arasındaki "kayıp linktir". Veri madenciliği teknikleri süreç merkezli değildir. Veri madenciliğinin karar ağaçları ve ilişki kuralları gibi temel özelliklerine karşın; süreç

Süreç Madenciliği Manifestosu

madenciliği eş zamanlı sunduğu veri analiz teknikleri ile eşsizdir. Süreç madenciliğinde tamamen yeni temsiller ve algoritmalar gereklidir.

- Süreç madenciliği çevrimdışı analizler ile sınırlı değildir. Süreç madenciliği teknikleri geçmiş olay verilerinden bilgiyi ortaya çıkartır. Kullanılan “Post Mortem” verilere rağmen sonuçlar devam eden bir olayda kullanılabilir. Örneğin; devam eden müşteri siparişinin tamamlanması için gerekli olan süre, daha önceden keşfedilmiş olan süreçlerden yola çıkarak tahmin edilebilir.



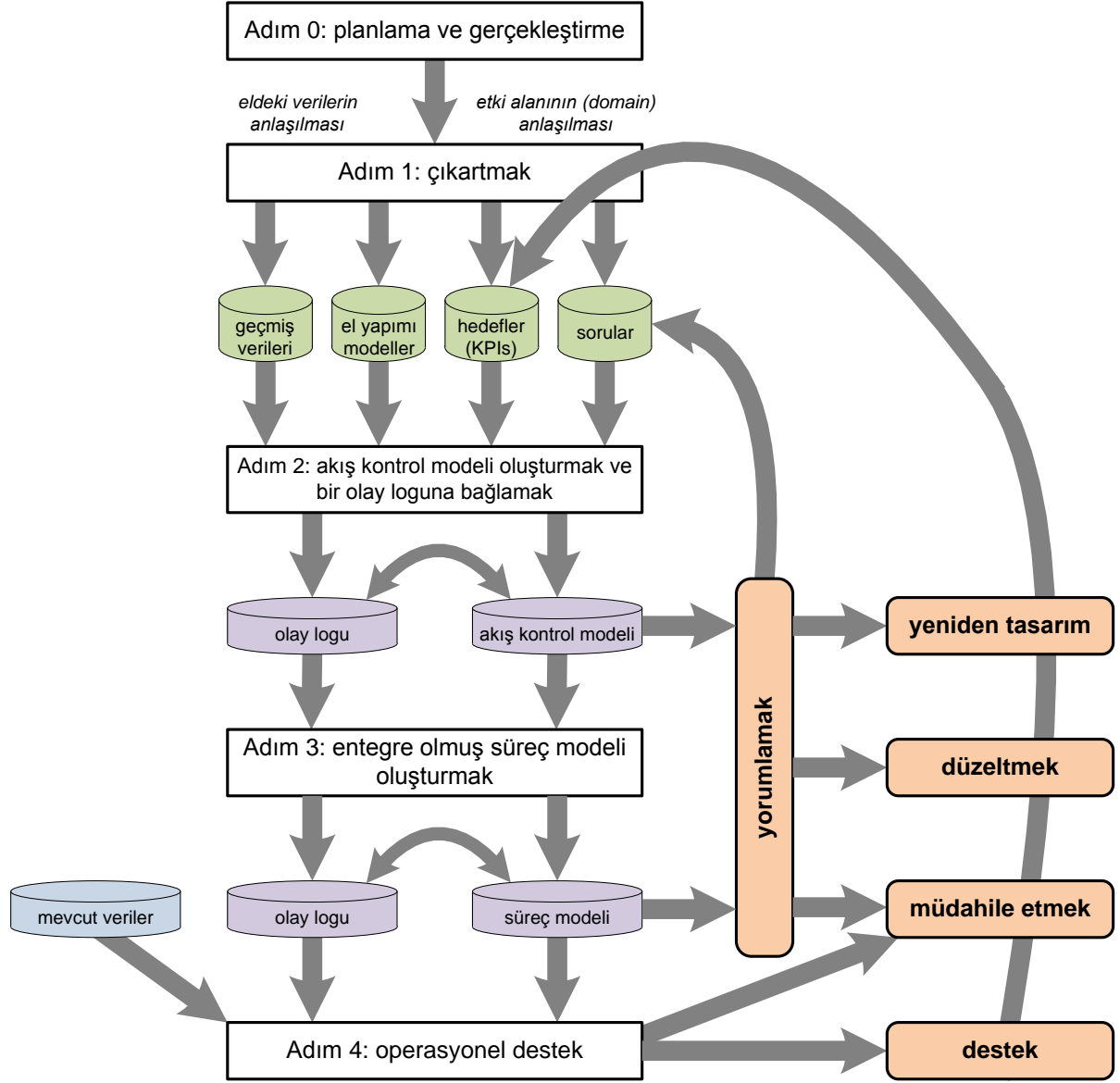
Şekil 4. BPM hayat döngüsü iş süreçlerinin farklı fazlarını ve bu fazların ilgili oldukları bilgi sistemlerini belirler; süreç madenciliği (potansiyel olarak) tüm fazlarda rol alır (uygulama fazı hariç).

Süreç madenciliğinin yerinin bulunması için şekil 4 deki İş Süreç Yönetimi (BPM) yaşam döngüsü kullanılır. BPM yaşam döngüsü hem iş sürecinin yedi aşamasını, hem de ilişkide oldukları bilgi sistemlerini gösterir. *Tasarım/yeniden tasarım* fazında bir süreç modeli yaratılır ya da var olan süreç modelinde değişiklikler yapılır. *Analiz* fazında ise planlanan model ve alternatifleri analiz edilir. *Tasarım/yeniden tasarım* fazından sonra *uygulama* fazı bulunur. Bu fazda planlanan model uygulanır. *Yeniden yapılandırma* fazında, var olan bir model yapılandırılır/yeniden yapılandırılır. *Yürütme* fazında ise tasarlanan model uygulanır; süreç incelenir. Buna ek olarak *ayarlar* fazında, yeniden yapılandırma fazına geçmeden önce var olan model üzerinde küçük ayarlamalar yapılır. Yürütülen süreç, *tanı* fazında analiz edilir ve çıktıları *yeniden tasarım* fazında başlayacak olan sürecin girdisini etkiler. Süreç madenciliği, şekil 4’de gösterilen fazların birçoğu için değerli bir araçtır. Özellikle *tanı* fazının süreç madenciliğinden yararlanacağı oldukça açıktır. Daha önce de bahsedildiği gibi süreç madenciliği *tanı* fazı ile sınırlı değildir. *Yürütme* fazında süreç madenciliği operasyonel destek olarak kullanılabilir. Devam eden olayların analizi, eski verilerden oluşturulmuş modellere bağlı olarak tahmin edilir ve öneriler yapılır. Karar destek sistemlerinin benzer formları süreçleri ayarlamak ve yeniden yapılandırmak için rehber olarak kullanılır.

Şekil 4 genel BPM hayat döngüsünü gösterirken, şekil 5 somut süreç madenciliği aktiviteleri ve eserleri üzerinde odaklanır. Şekil 5 süreç madenciliği projelerindeki olası adımları tanımlar. Herhangi bir süreç madenciliği, planlama ve bu planlamanın gerçekleşmesi ile başlar (adım 0). Proje başladıktan sonra; olay verileri, modeller, hedefler ve sorular; sistemlerden, ana uzmanlardan ve yönetimden öğrenilir (adım 1). Erişilebilen veriler anlaşılabilir olmalıdır (“Analiz için neler kullanılabilir?”) ve araştırma alanı netleştirilmelidir (“önemli sorular nelerdir?”). Bu araştırmanın sonucunda şekil 5’deki

Süreç Madenciliği Manifestosu

sonuçlar (yani; geçmiş verileri, manuel yapılmış modeller, hedefler ve sorular) ortaya çıkar. İkinci adımda, akış kontrol modeli oluşturulur ve olay logu ile bağlantısı yapılır. Bu adımda otomatikleştirilmiş süreç keşif teknikleri kullanılabilir.



Şekil 5. L*'ın hayat döngüsü; beş adımdan oluşan bir süreç madenciliği projesini tanımlar: Planlama ve gerçekleştirme(Adım 0) , çıkarmak (Adım 1), akış kontrol modeli oluşturmak ve bunları bir olay loguna bağlamak (Adım 2), entegre olmuş süreç modeli oluşturmak (Adım 3), ve operasyonel destek sağlamak (Adım4).

Model bu ilk hali ile, aranan bazı cevapları verir ve ayarlama aksiyonlarını tetikler. Buna ilaveten, olay logu, model olarak kullanılarak filtrelenebilir ya da uyarlanabilir (nadir olan aktiviteler ya da sıra dışı olaylar kaldırılır ya da eksik olaylar eklenir). Bazen, aynı süreç örneğine ait olayların ilişkilendirilmesi çok fazla çaba gerektirir. İlişkilendirilemeyen olaylar süreç modelindeki varlıklar ile ilişkilendirilir. Süreç yapılandırıldığında, üçüncü adım boyunca akış modeli veri, zaman ve kaynaklar gibi farklı perspektifler ile geliştirilir. İkinci adımda olay logu ve model arasındaki ilişki geliştirilir. (örneğin; ilgili olay zaman bildirelerinin kullanılması ile aktivite bekleme sürelerinin tahmin edilmesi gibi). İlişkiler; cevapları aranan soruların keşfi için kullanılır ve ek aksiyonları da tetikler. Süreç

Süreç Madenciliği Manifestosu

madenciliğinin en sonunda, üçüncü adımda yapılandırılan modeller operasyonel destek için kullanılır (adım4). Devam eden olaylara müdahale etmek ve tavsiyelerde bulunmak için; geçmiş olay log dastasından elde edilen bilgi, devam eden bu olayın verileri ile birleştirilir. Üçüncü ve dördüncü adımlarda süreç kararlı ve yapılandırılmış bir olgunluğa erişir.

Şekil 5 de gösterilen adımları destekleyecek teknikler ve araçlar günümüzde vardır. Ancak süreç madenciliğinin yeni bir paradigma olmasından dolayı, oldukça yeni olan araçların gelişmesi henüz tamamlanmamıştır. Buna ek olarak, muhtemel kullanıcılar süreç madenciliğinin potansiyel faydalarından ve limitlerinden haberdar değildir. Manifestonun üçüncü bölümü rehber prensiplere, dördüncü bölümü de gözlenen zorluklara ayrılmıştır. Manifesto; araştırmacılar, yazılım geliştirme uzmanları ve süreç madenciliğini kullananlar için bir rehber niteliğindedir.

3. Rehber Prensipler

Süreç madenciliğinin gerçek uygulamalarında da diğer yeni teknolojilerde olduğu gibi yapılan bazı hatalar vardır. Süreç madenciliği kullanıcılarını ve analistleri bu hatalardan korumak için altı rehber prensip belirlenmiştir.

3.1 RP1: Olay Verilerine Birinci Sınıf Vatandaş gibi Hizmet Edilmelidir

Veri madenciliği yöntemlerinin başlangıç noktası kaydedilen olaylardır. Olayların toplanmış haline log adı verilir. Ancak log kayıtlarının olması, verilerin daha önceden belirlenmiş bir log dosyasında saklanması gerektiği anlamına gelmez. Olaylar; veri tabanı tablolarında, mesaj loglarında, posta arşivlerinde, transaksyon loglarında ve diğer veri kaynaklarında da kaydedilebilir. Önemli olan, kayıtların hangi formda kayıt edildiği değil, bu kayıtların kalitesidir. Süreç madenciliği sonuçlarının kalitesi, girdi verilerinin kalitesine bağlıdır. Bu nedenle analiz edilecek bilgi sistemlerindeki loglar birinci sınıf vatandaşlar gibi hizmet görmelidir. Ne yazık ki olay logları çoğu kez “ürüne göre” hata gidermek ve hataların gösterilmesi için kullanılır. Örneğin; Philips Healthcare yazılımında, uzmanlar kodun içerisinde olayları basitçe kaydedecek özellik eklemişlerdir. Böylece Philips Healthcare olayları basitçe kaydeder. Olay loglarının kalitesini arttırmak için sistematik yaklaşımlar gereklidir. Olay logları ikinci sınıf vatandaşlar olarak değil, birinci sınıf vatandaşlar olarak muamele görmelidir.

Verilerin kalitesini ölçmek için birçok kriter vardır. Öncelikle veriler güvenilir olmalıdır; olayların gerçekleştiğinden ve doğru özelliklerin kaydedildiğinden emin olmak gereklidir. Olay logları tamamlanmış olmalıdır. Oluşan taslak logda, herhangi bir veri kaybı olmamalıdır. Logların semantikleri iyi tanımlanmış olmalıdır. Buna ek olarak veriler gizlilik ve güvenlik açılarından da yeterli olmalıdır. Örneğin; aktörler kaydedilen olayların tiplerinden ve kullanımlarından haberdar olmalıdır.

Tabela 1 logların kalite skalasını zayıf kaliteden (*) mükemmel kaliteye(*****) kadar göstermektedir. Örneğin; Philips Healthcare’ın olay logları *** seviyesinde bulunur çünkü olaylar otomatik olarak kaydedilmesine ve gerçekte uyumlu olmasına rağmen, verilerin kaydı ve çeşitli süreçler sırasında herhangi bir sistematik yoktur. ***,**** ve ***** kalite seviyelerindeki loglarda süreç madenciliği teknikleri uygulanır. Aslında * ve ** kalitesindeki loglarda da süreç madenciliği uygulanabilir ancak bu analizlerin sonucu genel olarak problemlidir ve güvenli olduğundan bahsedemeyiz. * seviyesindeki loglar için süreç madenciliğini uygulamak fazla anlamlı değildir diyebiliriz.

Süreç madenciliği yöntemlerinden en iyi şekilde faydalanmak için organizasyonlar logların kalitesini en yüksek seviyede tutmayı hedeflemelidir.

Süreç Madenciliği Manifestosu

3.2 RP2: Logların Sorular ile Ortaya Çıkarılması Gerekir

Şekil 5 de görüldüğü gibi süreç madenciliğinde aktivitelerin sorular ile yönetilmesi gerekir. Yalnızca somut sorular ile anlamlı olay verileri ortaya çıkartılır. Örneğin; SAP sistemlerindeki binlerce ERP tablolarını düşünün. Somut sorular olmadan ilgili tabloları seçmek imkânsızdır.

Şekil 1 belirli bir modeldeki olayların hayat döngüsünü süreçlerle ifade eder (yani; süreç örnekleridir). Bundan dolayı; süreç madenciliği teknikleri uygulanmadan önce analiz edilecek olayların tipi seçilmelidir. Cevabı aranan sorulara göre tercih yapılması oldukça önemlidir. Örneğin müşteri siparişlerini düşünün. Her müşterinin siparişi birden fazla sipariş hattını içerebilir ve/veya bir sipariş içerisinde müşteri birden fazla sipariş verebilir. Ayrıca, bir müşteri siparişi birden fazla teslimatı içerebilir. Bundan dolayı, sipariş ve tedarik arasında çoktan çoğa (n den m'e) bir ilişki; sipariş ve sipariş hatları arasında ise birden çoğa (1'den n'e) ilişki vardır. Olay verilerini kaydeden bir veri tabanında siparişler, sipariş hatları ve teslimat ilişkilidir. Bu veri tabanından farklı süreçler keşfedilebilir. Sadece tekil siparişlerin hayat döngülerini tanımlamak için veriler kullanılabilir.

Tabela 1 Olay loglarının olgunluk seviyeleri

Seviye	Özellik
*****	En yüksek seviye: olay logu mükemmel kalitededir (diğer bir deyişle; log kayıtları güvenilir ve tamamlanmıştır) ve olaylar iyi tanımlanmıştır. Olaylar; otomatik, sistematik, güvenilir ve güvenli olarak kaydedilir. Gizlilik ve güvenlik konularına yeterince dikkat edilmektedir. Buna ek olarak; kaydedilen olayların ve tüm özelliklerin temiz bir semantik yapısı vardır. Bu özellik, bir ya da daha fazla ontolojinin varlığını gösterir. Olaylar ve özellikleri bu ontolojiye işaret ederler. Örneğin; BPM sistemlerinin semantik açıklamalı logları.
****	Olaylar otomatik, sistematik ve güvenli bir şekilde kaydedilir; yani loglar güvenilir ve tamamlanmıştır. Operasyon seviyesi *** olan loglara göre, süreç örneği (durum) ve aktivite gibi kavramlar açık bir şekilde desteklenmektedir. Örneğin; geleneksel BPM/workflow sistemlerindeki geleneksel log olayları.
***	Olaylar otomatik olarak kaydedilir ancak, olayları kaydetmek için herhangi bir sistematik yaklaşım yoktur. Ancak; ** seviyesindeki loglardan farklı olarak, kaydedilen olayların gerçeği belirli bir ölçüde karşıladığından bahsedebiliriz (diğer bir ifade ile; olay logları güvenilir ancak loglar mutlaka tamamlanmış değildir). Bir ERP sisteminde kaydedilmiş olan olayları ele alalım. Olaylar çok farklı tablolardan çıkartılır ve bilgi doğru olarak varsayılır (örneğin; ERP sistemi tarafından kaydedilmiş bir ödemenin gerçekleştiğini kabul etmek güvenilirdir ya da bu olayın tam tersini de kabul edebiliriz). Örnekler: ERP sistemlerindeki tablolar, CRM sistemlerinin kayıtları, mesajlaşma sistemlerinin işlem logları, yüksek teknoloji sistemlerin olay logları vs.
**	Olaylar otomatik olarak kaydedilir. Diğer bir ifade ile, bazı bilgi sistemlerinde yan ürün olarak kaydetme işlemi yapılır. Kapsama alanı çok değişkendir, yani hangi olayların kaydedileceğine dair herhangi sistematik bir yaklaşım yoktur. Buna ek olarak; bilgi sistemlerini bypass etmek mümkündür; bu nedenle, olaylar eksik ya da düzgün kaydedilmemiş olabilir. Örnekler: bir doküman ya da ürün yönetim sistemindeki olay logları, gömülü sistemlerdeki hata kayıtları, servis mühendislerinin notları, vs.
*	En düşük seviye: olay logları zayıf kalitededir. Kaydedilen olaylar gerçeği karşılamayabilir ve olaylar eksik olabilir. Olayları el ile kaydedilen loglar genellikle bu özelliklere sahiptir. Örnekler: kağıda yazılmış tıbbi notlar, organizasyon içerisinde gezen notlar içerisinde kalmış izler ("sarı notlar")

Verilerden tekil sipariş hatlarının ya da tekil teslimlerin hayat döngüsünü keşfetmek mümkündür.

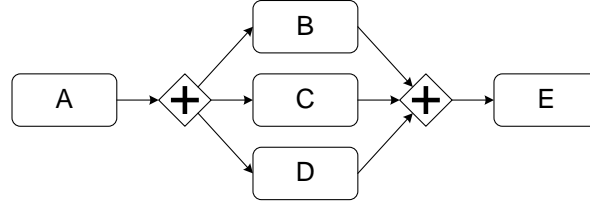
3.3 RP3: Tutarlılık, Seçim ve Diğer Temel Akış Yapıları Desteklenmelidir

Birçok süreç modelleme dili vardır (BPMN, EPC, Petri Ağları, BPEL ve UML aktivite diyagramı gibi). Bu dillerden bazıları birçok modelleme elementi içerirken (örneğin; BPMN 50 den fazla farklı grafik elementi sunar) bazıları da oldukça temeldir (örneğin; petri ağları üç temel elementten oluşur; yerler, geçişler ve arklar). Kontrol akış tanımlanması süreç modelleri için temel dayanma noktasıdır. Tüm ana diller tarafından; temel iş akış yapısı (aynı zamanda desen-pattern- olarak da bilinir), sıra¶lel yönlendirme (AND ayrımları ve birleşmeleri), seçim (XOR ayrımları ve birleşmeleri) ve döngüler oluşturulur. Bu desenlerin (pattern) süreç madenciliği tarafından desteklendiği açıktır. Buna rağmen; bazı teknikler eş zamanlılıkla başa çıkmazlar ancak Markov zincirleri/ geçiş sistemlerini desteklerler.

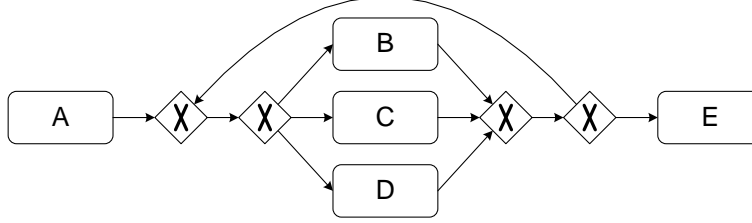
Şekil 6 süreç madenciliğinde tutarlılık keşfinin (AND ayrımları ve birleşmeleri yoktur) olmadığını gösterir. Olay logu L yi ele alalım; $L = \{ (A,B,C,D,E), (A,B,D,C,E), (A,C,B,D,E), (A,C,D,B,E), (A,D,B,C,E), (A,D,C,B,E) \}$. L; A ile başlayıp E ile biten olayları içerir. B,C ve D aktiviteleri A ile E arasındaki herhangi bir yerde ortaya çıkabilir. Şekil 6(a) daki BPMN modeli iki AND bağlantısı olan olayların altında yatan yoğun süreçleri gösterir. Süreç madenciliğinin AND bağlanmalarını desteklemediğini düşünelim. Bu durumda; şekil 6 daki iki BPMN modeli kesinlikle aday olurdu. Şekil 6b daki BPMN modeli oldukça yoğundur, ancak daha fazla davranışa izin verir (örneğin; (A,B,B,B,E) modele göre gerçekleşebilir bir durumdur ancak olay loguna göre olması pek mümkün değildir). Şekil 6c de olan BPMN modeli L'deki durumların olmasına izin verir ancak tüm dizileri açıkça kodlamasına rağmen, logun yoğun bir temsilcisi değildir. Örnekler gösterir ki, gerçek zamanlı aktivitelerin olası binlerce aktivitesi vardır. Sonuçlanan modeller açıkça ortalamanın altındadır (yani çok fazla davranışa izin vermektedir) ve/veya tutarlılıkla desteklenmediği durumlarda oldukça karmaşıktır.

Şekil 6 da gösterildiği gibi en azından temel iş akışı desenini (work flow pattern) desteklemek önemlidir. Bahsi geçen temel desenlerin yanı sıra OR ayrımları ya da birleşmelerini desteklemesi de beklenir. Böylece kısmi senkronizasyonlar ve kararlar gibi yoğunluklar da desteklenir.

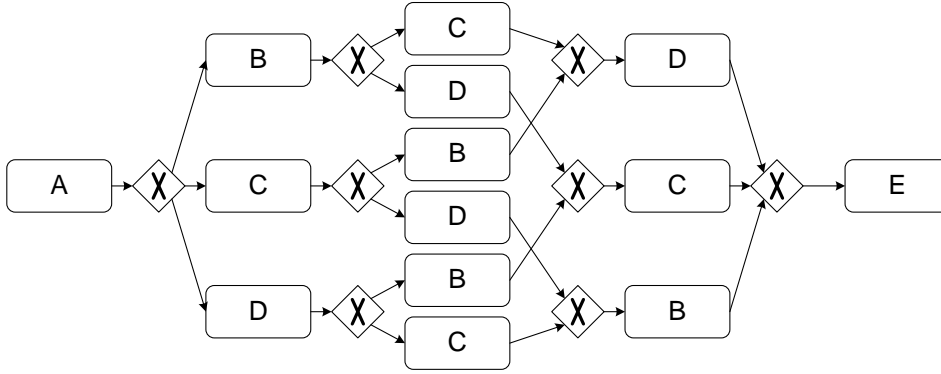
Süreç Madenciliği Manifestosu



(a) B, C, ve D herhangi bir sıra ile gerçekleşebilir



(b) B, C, ve D hem herhangi bir sıra ile hem de birçok kez gerçekleşebilir



(c) B, C, ve D herhangi bir sıra ile gerçekleşebilir fakat modelin tamamen incelenmesi için ikişer kere gerçekleşmelidir.

Şekil 6. süreç madenciliğinde tutarlılık keşfinin (AND ayrımları ve birleşmeleri yoktur) direk olarak açıklayamadığı duruma ait bir örnektir. Sadece B,C ve D aktiviteleri eş zamanlıdır. 10 tane aktivitenin eş zamanlı olarak devam ettiği bir durumdaki sonuç süreç modelini düşünün ($2^{10} = 1024$ durum ve $10! = 3,628,800$ tane işleme sırası mümkün olur).

3.4 RP4: Olaylar Model Elementleri ile İlişkili Olmalıdır

Süreç madenciliği bölüm 2 de de anlatıldığı gibi sadece akış kontrolü keşfi (control flow discovery) ile sınırlı değildir. Şekil 1 de gösterildiği gibi keşfedilmiş süreç modeli birçok bakış açısını kapsar (organizasyonel perspektif, zaman perspektifi, veri perspektifi vs). Buna ek olarak; keşif, şekil 3 de gösterilen süreç madenciliği çeşitlerinden sadece birisidir. Süreç madenciliğinin diğer iki çeşidi olan uyum ve geliştirme; modeldeki elementler ve olay logundaki olaylar arasındaki ilişkilere çok sıkı bağlıdır. Model ve olay logundaki ilişki “yeniden yürüt (replay)” de kullanılır.

Replay (yeniden yürütme)’de olay logu ve model arasındaki uyumsuzluklar ortaya çıkartılır. Logdaki bazı olayların modele göre imkansız olduğu görülebilir. Uygunluk kontrolündeki teknikler bu gibi uyumsuzlukları ortaya çıkartır ve miktarlarını belirler. Loglarda zaman bilgisini içeren notlar bulunur. Yeniden yürütme fazında; var olan zaman bilgileri ile süreçteki geçici davranışlar analiz edilir. Nedensel olarak birbiri ile ilişki aktivitelerin arasındaki zaman farkı belirlenir. Farkın belirlenmesi ile modelde süre gelen diğer aktivitelerin tamamlanması için gerekli olan bekleme süresi tahmin edilir. Bu örnekler; loglardaki olaylar ile modeldeki elementlerin farklı analizler için çıkış noktası olarak kullanıldığını gösterir.

Süreç Madenciliği Manifestosu

Bazı durumlarda, bu gibi ilişkiler belirlemek oldukça önemlidir. Örneğin; bir olay farklı iki aktivite ile ilişkili olabilir ya da hangi aktivite ile ilişkili olduğu tam olarak net değildir. Bu gibi belirsizlikler ortadan kaldırılmalıdır. Çünkü belirsizlikler var olduğu sürece, süreç madenciliği sonuçlarını düzgünce yeniden yürütmek mümkün olmaz. Olaylardan aktivitelere aktarılan problemin yanı sıra, olaylardan süreç örneklerine aktarılan problemde vardır. Bu duruma genel olarak olay korelasyonu denilir.

3.5 RP5: Modeller, Gerçeğin Soyutlamaları gibi ele alınmalıdır

Olay verilerinden oluşmuş modeller gerçek süreçlerin görülmesini sağlar. Logda bir davranışın soyut hali belirli bir nedene hizmet etmek amacıyla yakalanmış olur. Bir olay logundan faydalı olabilecek birçok bilgi vardır. Buna ek olarak, organizasyonlardaki farklı paydaşlar farklı bilgilere ihtiyaç duyar. Aslında olay loglarından ortaya çıkarılan modeller birer “harita” olarak görülmelidir (aynı coğrafi haritalar gibi). Bu rehber prensip aşağıda bahsedilen başlıca iki özellik olmak üzere, önemli diğer ayrıntıların da ortaya çıkarılmasını sağlar.

Öncelikle belirtilmelidir ki, belirli bir coğrafi bölge için “bir harita” yoktur. Farklı kullanım amaçlarına uygun olarak geliştirilen haritalar vardır: yol haritaları, yürüyüş haritası, bisiklet yolu haritası vs. Bu haritaların her biri gerçeğin farklı açılardan görüntüsüdür ve herhangi birini “mükemmel harita” olarak belirlemek uygun olmayacaktır. Süreç modelleri için de benzer bakış açısı geçerlidir: her bir model belirli bir kullanıcı grubu için uygun olan bilgileri vurgulamalıdır. Keşfedilen modeller kontrol akışı, veri akışı, zaman, kaynak, maliyet ve benzeri bakış açılarına odaklanabilir ve farklı bakış açılarına farklı seviyelerde bağlı olur. Yani; bir yönetici maliyetlere kabaca odaklanıp resmi olmayan süreç modelini isterken, bir süreç analisti normal akışlardan sapmaları gösteren ayrıntılı bir süreç akış modelini isteyebilir. Buna ek olarak; paydaşlar da süreçleri farklı seviyelerden izlemek ister: stratejik seviye (bu seviyedeki kararlar uzun vadede etkiler ve olay verilerinin uzun vadede toplanması temellidir), taktiksel seviye (bu seviyedeki kararlar orta seviyede dönemde etki ederler ve çoğunlukla güncel verilere dayanır) ve operasyonel seviye (bu seviyedeki kararlar ani etkilerdir ve durumlarda gerçekleşen olay verilerine dayanır).

İkinci olarak; özellikle konu anlaşılabilir haritalar yaratmaya geldiğinde, kartografiden bilgiler almak faydalıdır. Örneğin; sadece önemli yol ve şehirleri içeren yol haritalarında daha az önemli olan şeyler ya çıkartılmıştır ya da bir şekil içerisinde uygun olarak toplanmıştır (örneğin; sokaklar ve banliyöler şehirler içerisinde toplanmıştır). Haritalar daha az ilgili detaylardan arındırılmıştır. Bunun yanı sıra haritalar önemli özelliklere dikkat çekmek için renkler kullanır. Bunlara ek olarak; grafiksel elementlerin özelliklerini belirtmek için belirli büyüklüklerde bulunurlar(örneğin; çizgilerin ve noktaların sayılarında değişiklikler olur). Coğrafi haritalar aynı zamanda x ve y düzlemlerinde düzgün bir yansıma sağlar. Örneğin; harita üzerindeki elementlerin koordinatlarının bir anlamı olduğu için, elementler rasgele yerleştirilmemiştir. Tüm bu özellikler süreç modelleri için oldukça zittir. Süreç modellerin daha anlaşılır olması için; renk, büyüklük veya yer özellikleri kullanılmaz. Buna rağmen, haritalardan alınan özellikler kolayca, keşfedilmiş haritaların üzerine eklenebilir. Örneğin; bir aktivitenin büyüklüğü onun sıklığını yansıtmak için kullanılabilir ya da önemini anlatmak için başka bir özellik kullanılabilir (örneğin; kaynak ya da maliyet kullanımı). Karşılıklı nedensel bağlılığın önemini ifade etmek için özelliklerden herhangi biri kullanılabilir. Renklendirmeler kısırdöngüleri öne çıkarmak için olabilir.

Yukarda anlatılanlar; doğru bir sunum şeklinin ve hedef kitleye göre özelleştirilmiş sunumun önemini göstermektedir. Bu seçim; sonuçların görsel hale getirilip son kullanıcılara uygun görseller

Süreç Madenciliği Manifestosu

oluşturmak için ve uygun modeller için yol gösteren algoritmaların keşfi için önemlidir (zorluk 5'e bakınız).

3.6 RP6: Süreç Madenciliği Devamlı Bir Süreç Olmalıdır

Süreç Madenciliği, olay verileri ile direk bağlantılı olan "haritaları" sağlamakta yardımcı olur. Bu modelde hem geçmiş zaman verileri hem de güncel veriler kullanılır. Buna ek olarak, süreçler analiz edilirken değişebilir. Süreçlerin dinamik olan doğasından dolayı, süreç madenciliğine tek zamanlı bir aktivite olarak bakmak önerilmez. Hedef belirli bir model yaratmak olmamalıdır. Ancak süreç modelleri güncel durumlara uyum sağlamalıdır ki, kullanıcılar ve analistler süreçlere günlük olarak bakmaya teşvik edilsin.

Coğrafi etiketlemede kullanılan mashups ile karşılaştırılabilir. Google haritalarında kullanılan binlerce mashups (örneğin; belirli bir harita üzerindeki trafik durumu, gayrimenkul, fast food lokantalarını ya da sinema saatlerini getiren uygulamalar) vardır. Sorunsuz bir şekilde bu tip haritalar üzerinde odaklanıp, haritalar ile iletişime geçilir (örneğin; trafik sıkışıklıkları harita üzerinde belirlidir ve kullanıcı detayları görmek için haritaya daha fazla yakınlaşabilir). Gerçek zamanlı olayların verileri üzerinden süreç madenciliği yapmak da mümkün olmalıdır. "Harita Benzetmesini" kullanarak, olayların GPS koordinasyonlarının olduğunu var sayabiliriz ve bu koordinasyonlar haritalar üzerinde gerçek zamanlı olarak gösterilebilir. Navigasyon sistemlerindeki analoglar ve süreç madenciliği araçları ile; (a) son kullanıcıların süreçlerine navigasyon yapar(b) ayrıca süreç haritaları üzerine dinamik bilgiler sunar (iş süreçlerinde "trafik sıkışıklığını göstermek"), bunlara ek olarak (c) devam eden durum için tahminler yapar (örneğin; gecikmesi olan bir süreç için "varış zamanı" tahmin edilir). Bu örnekler de gösteriyor ki, süreç madenciliğini kullanmak gerçekten faydalıdır. Bu nedenle, süreç madenciliğine farklı zaman ölçülerinde (dakikalar, saatler, günler, haftalar ve aylar) bilgi sunan, devamlı bir süreç olarak bakılmalıdır.

4. Zorluklar

Süreç madenciliği özellikle önemli operasyonel süreçleri olan organizasyonlar için önemli bir araçtır. Bir taraftan bilgi sistemlerinde kaydedilen verilerdeki artış oldukça fazladır. Diğer yandan da; süreçler ve içerdikleri bilgiler; uyum, verimlilik ve müşteri memnuniyeti gibi sebepler nedeniyle mükemmelleştirilmelidir. Süreç madenciliği oldukça kolay uygulanabilir. Ancak belirtilmelidir ki; süreç madenciliği gelişmekte olan bir disiplin olduğundan, çeşitli zorlukları vardır. Aşağıda süreç madenciliğinin çeşitli zorluklarından bahseden liste vardır. Bu liste zorlukları belirtmek için oluşturulmuştur. Ayrıca zaman içerisinde yeni zorluklar ortaya çıkabilir ya da var olan zorluklar süreç madenciliğinin diğer faydaları ile ortadan kalkabilir.

4.1 Z1: Olay Verilerinin Bulunması, Ayarlanması ve Temizlenmesi

Süreç madenciliğine uygun verileri oluşturmak için önemli derecede çaba harcamak gerekmektedir. Genel olarak birçok engelin aşılması gerektiğinden bahsedebiliriz.

- Veri birçok kaynak tarafından *dağıtılmış* olarak bulunabilir. Bu verilerin birleştirilmeye ihtiyacı vardır. Bu zorlu durum; farklı kullanıcıların farklı veri kaynaklarını kullandıkları durumlarda ortaya çıkar. Örneğin; bir sistem bir kişiyi tanımlarken kullanıcı adı ve doğum tarihini kullanırken, diğer bir sistem kullanıcının sosyal güvenlik numarasını kullanabilir.
- Olay verileri çoğu kez "süreç odaklı" değil "nesne odaklıdır". Örneğin; bireysel ürünlerin, paletlerin ve konteynerlerin RFID etiketleri vardır ve kayıtlı olaylar bu etiketleri belirtir. Bu

Süreç Madenciliği Manifestosu

nedenle, belirli bir müşterinin siparişini takip ederken bu nesne merkezli olayların birleştirilmesi ve sürecinin yeniden işletilmesi gerekir.

- Olay verileri *tamamlanmamış* olabilir. Olayların açıkça bir süreç örneğini hedeflememesi genel bir problemdir. Hangi olayın hangi süreç örneğine bağlı olduğu oldukça uğraş gerektiren araştırma sonucunda ortaya çıkartılabilir. Diğer bir sorunda; bazı olaylar için zaman bilgisi kayıp olabilmektedir. Zamanlama bilgisinin kullanılması için; bu bilginin eklemesi gerekir.
- Bir olay logunda *sıradışılar (outliner)*; yani gürültü olarak kabul edilen istisnai durumlar olabilir. Sıradışılar nasıl tanımlanmalıdır? Sıradışılar nasıl belirlenmelidir? Olay verilerini temizlemek için bu soruların cevaplandırılması gerekir.
- Loglar farklı seviyelerde detay(tanecik) yapıya sahip olaylar içerebilir. Bir hastane bilgi sistemindeki olay logunda; basit kan testinden, karmaşık bir ameliyat prosedürüne kadar olaylar yer alır. Yani; milyon saniye hassaslığından (28-09-2011:h11m28s32ms342) kaba tarih bilgisine (28-09-2011) kadar uzanan farklı detay (tanecik) yapısındaki zaman özellikleri bulunur.
- Olaylar belirli bir *durumda* ortaya çıkar (hava, iş yükü, haftanın bir günü vs). Bu durum kesin bir fenomeni açıklayabilir. Örneğin; bir süreç işlemdeyse tamamlanma süresi beklenen kadar olacaktır ya da tatil zamanı ise bekleme süresi tahmin edilen kadar olacaktır. Bu durum, olay verileri ile durumsal verilerin birleştirilmesini anlamına gelir. Fazla değişkenin analize katılması ile, analiz çok etkileşimli hale gelir. Bu duruma analizin “çok yönlülüğün zorluğu” ortaya çıkar.

Yukarıdaki problemleri çözmeyi hedefleyen daha iyi araçlar ve metodolojiler oluşturulmalıdır. Buna ek olarak, daha öncede belirtildiği gibi, organizasyonlar olay loglarına ürünlerden oluşturulmuş bilgiler olarak değil, birinci sınıf vatandaşmış gibi muamele etmelidir. Hedef tabele 1 deki gibi ***** sınıfındaki olay loglarını elde etmektir. Yüksek kalitede olay logları elde etmek için veri ambarındaki öğrenilmiş dersler faydalıdır. Örneğin; veri girişindeki basit kontroller olay verilerindeki hatalı oranını gözle görünür biçimde azaltır.

4.2 Z2: Farklı Karakterleri Olan Karmaşık Olay Logları ile Başa Çıkmak

Olay loglarının farklı karakterleri olabilir. Bazı olay logları kendileri ile başa çıkmayı zorlaştıracak kadar büyük olabilirken, bazıları gerçekçi sonuçlar ortaya çıkarmayı engelleyecek kadar kısa olabilir.

Bazı alanlarda, akıllara durgunluk veren sayıda olaylar kaydedilir. Bu nedenle, performansı ve ölçeklendirilebilirliği arttırmak için ekstra çabaya gerek vardır. Örneğin; ASML, yonga plakasının tarayıcılarının hepsini sürekli olarak kontrol eder. Bu yonga plakası tarayıcıları birçok organizasyon tarafından cip (ortalama olarak ciplerin üretiminin %70 inde ASML yonga plakasının tarayıcısı kullanır) üretmek için kullanılır (örneğin; Samsung ve Texas Instruments). Bu gibi alanlardaki peta bitlik veriler ile başa çıkmakta var olan araçlar zorlanırlar. Birçok olayın kaydının yanı sıra; durum başına ortalama olay sayısı, durumlar arasındaki benzerlikler, tekil olayların sayısı, tekil yolların sayısı gibi farklı özellikler de vardır. L1 olay logunu şu özellikleri taşıyan: 1000 durum, durum başına ortalama 10 olay ve az sayıda varyasyon (yani, birçok durum benzer ya da aynı yolu kullanıyor). Log L2 de ise 100 durum olduğunu ve durum başına ortalama 100 olay olduğunu ve her olayın da tek bir yolu takip ettiğini düşünelim. İki olayın benzer büyüklüğünün olmasına rağmen (ortalama olarak 10 000 olay) ; L2'yi analiz etmek L1'i analiz etmekten açıkça çok daha zordur.

Olay logları sadece örnek davranışları içerirken, bu davranışlar tamamlanmış olarak kabul edilmemelidir. Süreç madenciliğinde “açık dünya varsayımı” vardır. Bu nedenle tamamlanmamış durumlar ile başa çıkılmalıdır. Bir olay örneğinin, log örneği üzerinde gerçekleşmemiş olması, hiçbir

Süreç Madenciliği Manifestosu

zaman gerçekleşmeyecek olduğu anlamına gelmez. Bu durum; birçok varyasyonu olan küçük olay logları ile başa çıkmayı daha da riskli hale getirir.

Daha önce bahsedildiği gibi, bazı loglar çok düşük seviyede soyutlanmış olaylara sahiptir. Loglar üzerindeki kayıtlar oldukça büyük olabilir. Büyük veri yığınları içerisindeki düşük tekil seviyedeki olaylara karşı, paydaşlar daha az ilgili olur. Bu nedenle düşük seviyedeki bu olaylar yüksek seviyedeki olaylarda toplanır. Örneğin; belirli bir grup hasta için yapılan tanı ve tedavi sürecini analiz ederken, bir paydaş hastane laboratuvarının bilgi sistemindeki sonuçlar ile tek tek ilgilenmez.

Bu gibi zamanlarda, organizasyonun bir deneme yanılma yaklaşımı kullanmasına ihtiyaç vardır. Böylece logun süreç madenciliği için uygun olup olmadığını anlar. Bu nedenle, bir veri setinin uygunluğunu ölçmek için kullanılacak araçların kullanımı hızlı olmalıdır. Ölçüm için kullanılacak test; potansiyel performans problemlerini gösterebilmelidir. Tamamlanmıştan çok uzak ya da çok karmaşık olan logları bu test uyarmalıdır.

4.3 Z3: Temsil Noktalarını Yaratmak

Süreç madenciliği hala gelişmekte olan bir teknolojidir. Bu nedenle de iyi temsil noktaları hala eksiktir. Onlarca süreç madenciliği tekniği ve satıcılar tarafından sunulan çok farklı ürünler olmasına rağmen bu ürünlerin kalitesi konusunda hiçbir fikir birliği yoktur. Süreç madenciliği ürünlerinin fonksiyonelliği ve performansında büyük farklılıklar vardır. Ancak bu farklı teknik ve araçları karşılaştırmak mümkün değildir. Bu nedenle; örnek veri setlerinden temsili kalite kriterleri geliştirmek gereklidir.

Klasik veri madenciliği teknikleri için oldukça yeterli temsil noktaları vardır. Bu temsil noktaları; araştırmacıların ve araç sunucuların kendi tekniklerini geliştirmeleri için uyarır. Süreç madenciliği durumunda daha risklidir. Örneğin; ilişkisel modeli Codd tarafından 1969 da basit ve oldukça destekleyici şekilde tanıtıldı. Sonuç olarak; veriyi bir veri tabanından diğere aktarmak oldukça kolay ve herhangi bir yorum problemi olmadan gerçekleşiyor. Süreçler için böyle bir model eksiktir. Süreç modellemesi için önerilen standartlar, ilişkisel modellere göre çok daha fazla karmaşıktır. Sadece az sayıda satıcı sadece tam olarak aynı süreç madenciliği kavramlarını desteklemektedir. Süreçler, tablo verilerinden oldukça daha fazla karmaşıktır.

Buna rağmen, süreç madenciliği için temsil noktası yaratmak önemlidir. Bazı temel çalışmalar yapılmıştır. Örneğin; süreç madenciliği sonuçlarının kalitesini karşılaştırmak için birçok metrik mevcuttur (uygunluk, basitlik, doğruluk ve genelleme). Buna ek olarak, birkaç olay logu erişilebilir durumdadır (cf. www.processmining.org). BPIC'11 de Task Force tarafından organize edilmiş ve kullanılmış örnek olay loguna bakınız (cf. doi:10.4121/uuid:d9769f3d-0ab0-4fb8-803b-0d1120ffc54).

Gerçek zamanlı veri setleri temelli noktalar geliştirilmesi gerekmektedir. Diğer yandan, belirli özellikleri yakalayacak suni veri setlerinin yaratılmasına ihtiyaç vardır. Bu gibi suni veri setleri sayesinde birçok özel durum için süreç madenciliği tekniklerini geliştirilebilir. Bu özel durumlar arasında; tamamlanmamış olay logları, gürültülü olay logları veya belirli süreç bütünleri sayılabilir.

Temsili noktaların yaratılmasının yanında, süreç madenciliğinin sonuçlarının kalitesini karşılaştırmak için kriterlerde fikir birliğine ihtiyaç vardır (zorluk 6 ya bakınız). Buna ek olarak; veri madenciliğindeki karşılıklı doğrulama teknikleri sonuçları değerlendirmek için kullanılabilir. Örnek K-durumu kontrolünü düşünün. K olay logu parçası uygulanabilir. K-1 parçası bir süreç modelini öğrenmek için

Süreç Madenciliği Manifestosu

kullanılabilir ve uygunluk kontrolü kalan parçaya göre değerlendirilebilir. Bu uygulama k kere tekrar edebilir, böylece modelin kalitesi hakkında bazı bilgiler edinilir.

4.4 Z4: Kavramsal Sapmalar ile Başa Çıkmak

Sürecin analizi sırasında sürecin değişmesi durumuna kavramsal sapma denilir. Bir log başındaki belirli iki aktivite eş zamanlı iken, logun diğer aşamasında söz konusu iki aktivitenin paralel hale gelmiş olması buna örnek olarak gösterilebilir. Süreçler; periyodik olarak, farklı koşullara göre (yani; “Pazar günleri piyasa daha rekabetçi hale geliyor”) ve hatta mevsimsel olarak (örneğin, “aralıkta daha fazla talep vardır” ya da “ Cuma akşamları daha az çalışan erişilebilir”) değişebilir. Bu gibi durumlar süreçleri etkiler. Bunların belirlenmesi ve analiz edilmesi gerekir. Log küçük parçalara ayrılır ve bu küçük logların analiz edilmesi ile süreçte kavramsal sapma olduğu ortaya çıkarılır. Logun yeniden incelenmesine “ikinci dizilim” analizi denilir. Bu analiz için çok daha fazla veriye ihtiyaç duyulur. Ne yazık ki, pek az süreç kavramsal sapmalardan arınmıştır ve bu durumda kavramsal sapma süreçlerin yönetimi için birincil öncelikli konu haline gelmektedir. Anlam kaymasını yeterince anlamak için ek araştırmalar ve araçların kullanımı bu nedenle gerekli ve önemlidir.

4.5 Z5: Süreç Keşfi için Kullanılan Temsili Yargıları Güçlendirmek

Süreç madenciliği tekniği BPMN ya da Petri Ağları gibi belirli bir dil kullanarak modeller üretir. Ancak, elde edilen sonuç modelin gösterimi ile, fiili keşif sürecinde kullanılan temsilleri birbirinden ayırmak önemlidir. Hedef dil seçimi, genel olarak birçok varsayım içerir. Seçilen dil araştırma uzayını kısıtlar; seçilen hedef dil tarafından gösterilemeyen süreçler, doğal olarak keşfedilemez. Dil seçimi, “temsilden kaynaklı ön yargı ” olarak adlandırılan bu durumdan dolayı sadece grafiksel gösterim tercihi baz alınarak yapılmamalı, dikkatli ve bilinçli bir şekilde yapılmalıdır.

4.6 Z6: Uygunluk, Basitlik, Doğruluk ve Genelleme gibi Kalite Kriterlerini Ayarlamak

Olay loglarında genellikle örnek davranışlar verilmiş olup, tamamlanmış değildir. Süreç modelleri üzerinde sonsuz sayıda, hızla büyüyen izler yani döngü durumları mümkündür. Buna ek olarak; bazı durumların gerçekleşme olasılığı diğerlerinden çok daha azdır. Bu nedenle mümkün olabilecek tüm izlerin olay logunda olmasını beklemek gerçekçi değildir. Sadece 10 adet paralel aktivitesi olan bir sürece ilişkin log yaklaşık olarak 10000 durumu içerir. Bu modeldeki 10 somut aktivitenin bir araya gelmesi ile oluşabilecek ilişkilerin sayısı ise $10! = 3628800$ dür. Yani daha az durumu olan (10000) log da tüm potansiyel izinlerin olması (3628800) mümkün değildir. Log içerisinde milyonlarca durum olsa da, olması muhtemel tüm varyasyonların yer alması pek olası değildir. Ayrıca bazı alternatiflerin gerçekleşme olasılığının diğerlerinden az olması da karmaşık bir durum ortaya çıkartır. Bu tip durumlar belki “gürültü” olarak da adlandırılabilir. Bu gibi gürültü davranışları için nedensel bir model yaratmak mümkün değildir. Keşfedilen model gürültüsel durumlardan soyutlanmalıdır. Özellikle uyum kontrolünde olma sıklığı az olan davranışlar araştırılmalıdır.

Loglardaki gürültü ve tamamlanmamışlık süreç keşfini daha da zorlu bir problem haline getirmektedir. Birbiri ile rekabet halinde olan 4 durum vardır: (a)uygunluk, (b)basitlik, (c)hassasiyet, (d)genelleme. Bir model uygunluk özelliğini taşıyorsa, log tüm izleri ile yeniden yürütülebilir olmalıdır. En basit model ise logda görülen davranışı en basit şekilde açıklayan model en iyi olandır. Bu prensip Occam’ın Razor’u olarak bilinir. Keşfedilmiş süreç modelinin kalitesini belirlemek için sadece uygunluk ve basitlik yeterli değildir. Örneğin; basit bir petri netz (“flower model”) kurmak oldukça kolaydır. Petri netz ile olay logundaki tüm olayların yeniden oynatılabilmesinin yanında aynı aktivite setine sahip olan bir olay logu da oynatılabilir. Aynı şekilde, sadece olay logunda olan olayın bire bir

Süreç Madenciliği Manifestosu

gösteriminin yapılması istenmez. Daha önce de bahsedildiği gibi, log sadece örnek davranışları içerir ve birçok izin o log da olmaması olasılığı da oldukça yüksektir. Bir model “çok fazla” davranışa izin vermiyorsa *kusursuzdur*. Petri netzin hassasiyet açısından yeterli olmadığını açıkça belirtebiliriz. Kusursuz olmayan bir model doğal olarak uygunluk ölçeklerinin de altındadır. Eğer log üzerindeki davranışların genelmesi çok fazla yapılıyorsa uygunluk seviyesinin altında olan bir model olduğundan bahsederiz(yani; model logda görülenin dışında birçok oldukça farklı davranışa da izin verir). Bir model genel olmasının yanı sıra, sadece logda görülen örnek davranışlar ile sınırlandırılmamış olmalıdır. Genelleme yapmayan bir model uygunluk sınırlarının üstündedir. Aynı şekilde uygunluk sınırlarının üstünde olmakta bir problemdir; yani logda olan davranışın örnek olduğu bilinirken bunun genellemesinin yapılmasıdır (model belirli örnek bir logu açıklar, fakat gelecek örnek log aynı süreç için tamamen farklı bir süreç modeli üretebilir).

Uyum, basitlik ve hassasiyet seviyelerini ayarlamak, ayrıca genelleme yapmak zorlayıcıdır. Bu nedenle en güçlü süreç madenciliği teknikleri birçok parametre sunar. Geliştirilmiş olan algoritmalar ile bu dört rekabetçi kalite boyutunun uyumlu olması sağlanır. Buna ek olarak, herhangi bir algoritmanın son kullanıcı tarafından anlaşılır olması bir zorunluluktur.

4.7 Z7: Karşılıklı Organizasyonel Madencilik

Süreç madenciliği geleneksel olarak tek bir organizasyon içerisinde kullanılır. Ancak günümüzde süreç teknolojisinde, tedarik zincir entegrasyonu ve bulut bilişim yaygın şekilde kullanılır hale geldiğinden, birçok organizasyonun loglarının analizinin mümkün hale geldiği durumlar mevcuttur. Prensip, karşılıklı organizasyonel süreç madenciliği için iki düzenleme vardır.

İlk olarak, farklı organizasyonların süreç örneklerini oluşturmak için birlikte çalıştığı işbirlikçi düzenlemeleri düşünmeliyiz. Bu karşılıklı organizasyonel süreçte, genel bir süreç parçalara bölünmüştür ve süreçler organizasyonlar üzerinde dağıtılmıştır. Bu gibi organizasyonlarda başarı ile sonuçlanmış işler yapmak için iş birliğine ihtiyaç vardır. Bu sebeplerden dolayı; karşılıklı organizasyonel süreç “jigsaw yapbozu’na” benzetilir. Baştan sona bir süreci keşfetmek için organizasyonlardan herhangi birini analiz etmek yeterli değildir. Farklı organizasyonların logları birleştirilmelidir. Logların birleştirilmesinde dikkat edilmeli, karşılıklı organizasyonların sınırları içerisinde ilişkilendirilmenin yapılması gerekmektedir.

İkincil olarak, farklı organizasyonların aynı süreçleri uygularken ortak bir yapıyı paylaştıklarını; birbirlerinin bilgi ve tecrübelerinden faydalandıklarını göz önünde bulundurmalıyız. Salesforce.com’u örnek olarak düşünebiliriz. Salesforce.com da birçok organizasyonun satış süreçleri birleştirilir ve sistem tarafından desteklenir. Bunun yanı sıra; sistem aynı sürecin farklı organizasyonlardaki farklı uygulamalarını da kabul eder böylece organizasyonlar katı süreç modellerini desteklemek zorunda kalmaz. Diğer bir örnek olarak; herhangi bir belediyede uygulanan bina yapım izinlerinin verilmesi gibi temel işleri düşünebiliriz. Bir ülkedeki tüm belediyelerin aynı temel süreç setini desteklemesi beklenirken, farklılıklar da olabilir. Farklı organizasyonlar arasında olan bu farklılıkları analiz etmek oldukça öğreticidir. Birbirleri ile etkileşim halinde olan organizasyonlar birebirlerinden öğrendikleri doğrultusunda, sunucularının servislerini değiştirebilir. Bu organizasyonlara uygulanacak, karşılıklı organizasyonel süreç madenciliğinin sonucunda, değer katılmış servisler sunabilir.

Karşılıklı organizasyonel süreç madenciliğinin iki tipi için de yeni analiz tekniklerinin geliştirilmesi gerekmektedir. Bu teknikler geliştirilirken, gizlilik ve güvenlik konularına da dikkat etmelidir.

Süreç Madenciliği Manifestosu

Organizasyonlar rekabet ya da yeterli güvenin olmaması nedeniyle bilgilerini paylaşmak istemeyebilir. Bu nedenle, güvenlik koruması olan süreç madenciliği teknikleri geliştirmek önemlidir.

4.8 Z8: Operasyonel Destek Sağlamak

Süreç madenciliğinin öncelikli odak noktası geçmişe dair verilerin analiziydi. Oysa günümüzde birçok veri kaynağı neredeyse gerçek zamanlı (oldukça yakın) bilgiler ile yüklendi. Böylece olaylar vuku bulurken onları analiz edebilecek yeterli bilgisayar gücü mevcuttur. Bu nedenle süreç madenciliği çevrimdışı analizler ile sınırlandırılmamalı ve çevrimiçi operasyonel destek için de kullanılmalıdır. Belirlemek, tahmin etmek ve önermek başlıkları altında üç farklı operasyonel destek aktivitesi tanımlanır. Bir durum daha önceden tanımlanmış olan süreçlerinden daha farklı olarak ilerlerse, sistem tarafından belirlenir ve sistem alarm verir. Sürece hala müdahale etme imkanı varken, alarm verilmesi değişime olanak sağlar ve bu uygulama çevrimdışı haldeyken yapılamaz. Modelleri oluşturmak için geçmişe dair veriler kullanılır. Oluşan modeller devam eden süreçler için rehber olur. Örneğin süren bir durumun tamamlanması için kalan süreyi tahmin etmek mümkün olur. Bu gibi tahminlerden yola çıkarak, harcanılan zamanı ve maliyetleri azaltacak ayrıca belirli etkiler de içerecek sistemler yaratılır. Süreç madenciliğini online uygulamalara eklemek, bilgi işlem ve veri kalitesi gibi alanlarda ek zorluklar yaratabileceği de bir gerçektir.

4.9 Z9: Süreç Madenciliğini Diğer Analiz Çeşitleri ile Birleştirmek

Birçok yönetim bilim dalı başta olmak üzere, operasyonel yönetimler ve araştırmalar modellemeye güvenir. Linear programlama ve proje planlamadaki modelleri sıralamadan, Markov zincirine ve simülasyona kadar birçok matematiksel model yelpazesi kullanılır. Veri madenciliği genellikle büyük olan veri setlerinden kolayca bulunamayan ilişkileri bulur, kullanıcılara bulunan bu yeni ilişkileri anlatır. Buna ek olarak, yeni ilişkilerin kullanıcılar tarafından faydalanacakları şekilde oluşturmak da veri madenciliğinin işidir. Analizleri yapmak için veri madenciliği başlığı altında çeşitli teknikler geliştirilmiştir: sınıflandırma (yani; karar ağacı öğrenmesi), regresyon, kümeleme (yani, k-ortalama analizi) ve desen keşfi (yani; ilişki kural öğrenmesi).

Hem operasyonel yönetim hem de veri madenciliği değerli analiz teknikleri sunar. Zor olan bu iki tekniği süreç madenciliği altında bir araya getirmektir. Simülasyonu düşünelim. Geçmiş verilerine bağlı olarak yapılan süreç madenciliği teknikleri bir simülasyon modelindeki öğrenmede kullanılabilir. Daha sonra simülasyon modeli operasyonel destek sağlamak için kullanılabilir. Model replay (yeniden yürütme) için kullanılır ve model/log simülasyonu başlatır ve “hızlı ileri tuşu” ile canlı veri üzerinde gelir.

Süreç madenciliğinin, görsel analitik ile birleştirilmesi istenmektedir. Otomatik görsel analiz de; interaktif görseller birleştirilerek geniş ve karmaşık veri gruplarının daha iyi anlaşılması sağlanır. Görsel analitik, insanların yapılandırılmış veriler içindeki desenleri görme yeteneğinden faydalanır. Otomatik süreç madenciliği tekniklerinin interaktif görsel analitik ile birleştirilmesi ile olay verilerinden daha fazla anlam çıkarmak mümkündür.

4.10 Z10: Uzman Olmayanlar İçin Kullanılabilirliği Geliştirmek

Süreç madenciliğinin bir hedefi de arşiv de sonlanan bir model yerine, günlük olarak kullanılan bir süreç modeli geliştirmektir. Bu tip süreç modellerine “yaşayan süreç modeli” adı verilir. Yeni olay verileri gelişmekte olan davranışların ortaya çıkarılması için kullanılır. Olay verileri ile süreç modeli arasında güçlü bir bağ vardır. Bu durum, güncel duruma projektör ile bakmak gibidir. Son aktivitelerin de güncel süreç modelinde olmasından dolayı, son kullanıcılar günden güne süreç madenciliğinin

Süreç Madenciliği Manifestosu

sonuçları ile etkileşime geçer. Bu gibi etkileşimler oldukça değerlidir ancak aynı zamanda sezgisel bir kullanıcı ara yüzünde ihtiyaç duyulur. Ara yüzler ardına sofistike algoritmaları yerleştirilerek, bu algoritmaların otomatik parametreler kurmasını sağlamak ve uygun analiz çeşidini önermesi sürecin en zor kısmıdır.

4.11 Z11: Uzman Olmayanlar için Anlaşılabilirliği Arttırmak

Süreç madenciliği ile araştırma sonuçlarına oldukça kolayca erişilebilir. Ancak belirli bir sonuca erişmek, faydalı veriler elde edildiği anlamına gelmez. Son kullanıcıların elde edilen bu verileri anlaması ve faydalı bilgiler edinebilmesi gereklidir. Dolayısı ile sonuçlar uygun bir şekilde sunulmalıdır (aynı zamanda RP6'ya bakınız). Buna ek olarak, sonuçların güvenilirliği her zaman açıkça belirtilmelidir. Belirli sonuçları savunmak için çok az veri olabilir. Süreç madenciliği teknikleri düşük ya da yüksek uygunluk seviyelerinde uygulanmaz. Süreç madenciliği; sonucu kanıtlamak için var olan veri setinin oldukça az olduğu durumlar da bile her zaman bir model ortaya çıkartır.

5. Son Söz

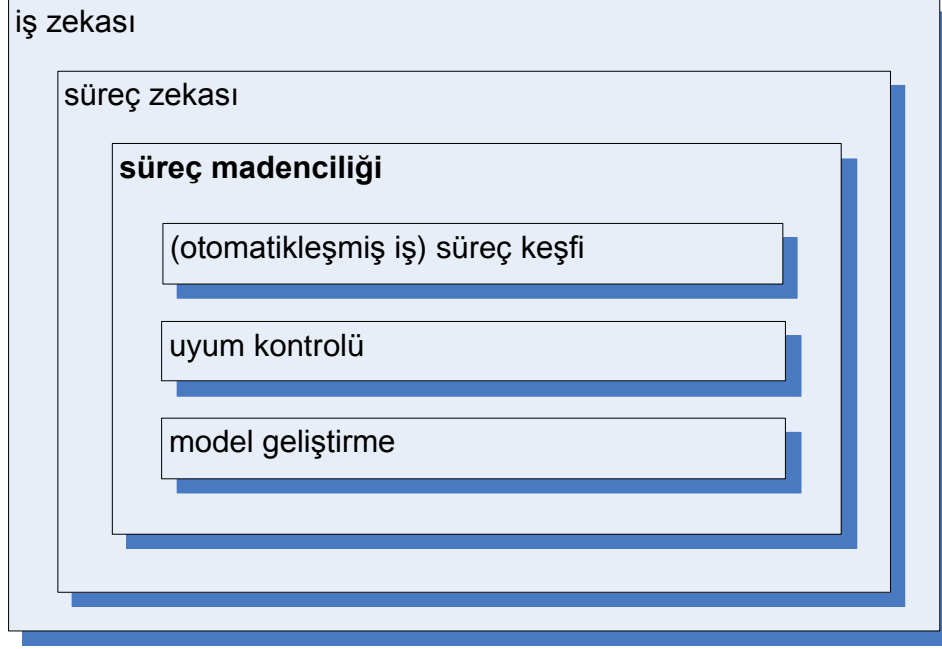
IEEE Task Force; süreç madenciliği uygulamalarını destekler(a). Yazılım geliştirme uzmanları, danışmanlar, iş yöneticileri ve son kullanıcıların son teknoloji ürünü olan süreç madenciliği tekniklerini kullanırken onlara rehberlik eder(b). Buna ek olarak IEEE Task Force süreç madenciliği üzerine olan araştırmalar hakkında önerilerde bulunur(c). Süreç madenciliği manifestosu Task Force'un amacını ve ana prensiplerini açıklar. Manifestoda öncelikle süreç madenciliği tanımı açıklanmış daha sonra bölüm 3'de rehber prensipler ve bölüm 4'de süreç madenciliğindeki zorluklar sınıflandırılmıştır. Rehber prensiplerin yardımı ile uygulama sırasındaki net hatalardan kaçınılır. Listelenen zorluklar ile süreç madenciliğindeki araştırma geliştirme konularının belirlenmesi hedeflenmiştir. Her iki bölümde temel olarak süreç madenciliği uygulamaların olgunlaşmasını hedeflemiştir.

Manifestoyu sonlandırmadan önce terminolojiden birkaç tanım açıklanmalıdır. Süreç madenciliği alanlarında kullanılan tanımları şöyle sıralayabiliriz: iş akışı madenciliği, (iş) süreç madenciliği, otomatikleştirilmiş (iş) süreç keşfi ve (iş) süreç zekası. Gözlemlerimize göre farklı organizasyonlar örtüşen konseptler için farklı tanımlar kullanıyor. Gartner'in "otomatikleşmiş iş süreci" tanımını ABPD desteklerken, SoftwareAG "Süreç Zekası" tanımını platformu kontrol etmek için kullanıyor. Süreç madenciliğindeki birçok uygulamadan sadece birisi olduğu için, iş akışı yaratımında "iş akışı madenciliği" daha az uygun gibi gözükür. Tanımlara eklenen "iş" sözcüğü ile, süreç madenciliğindeki uygulama alanlarından bazıları kesin olarak kısıtlanmış oluyor. Süreç madenciliğinde sınırsız uygulama alanı olduğundan dolayı (örneğin; ileri teknoloji sistemlerinin analiz edilmesi ya da web sitelerinin analiz edilmesi gibi), tanımlara iş başlığının eklenmesi uygun değildir. Süreç keşfi, süreç madenciliği yelpazesindeki uygulamalardan sadece bir tanesidir. Süreç madenciliği ile gelen diğer uygulama alanları arasında; uyum kontrolü, tahmin, organizasyonel madencilik de yer almaktadır.

Bahsedilen bu terimler ile şekil 7 ilişkilidir. Tüm teknolojiler ve metotlar; iş zekâsı şemsiyesi altında bulunan karar verme sitelerini destekleyici, aksiyon alınabilecek bilgiler sunmayı amaçlar. (İş) süreç zekâsı; BI ve BPM'in birleşimi olarak görülür; yani süreçleri analiz etmek, geliştirmek ve yönetmek için BI yöntemleri kullanılır. Süreç madenciliği olay loglarını çıkış noktası olarak alan süreç zekâsı olarak tanımlanır. Üç temel süreç madenciliği çeşidinden birisi de, otomatikleşmiş iş süreç madenciliğidir. Manifesto da anlatılan birçok süreç madenciliği fonksiyonunu BI desteklemez. Aslında

Süreç Madenciliği Manifestosu

şekil 7 eksik yönlendirilmiştir diyebiliriz. BI için yapılan tanımlar sadece belirli araç ve metoda uygun kullanımları içermektedir oysaki bu tanımlar BI yelpazesinin sadece küçük bir kısmını karşılamaktadır.



Şekil 7. Farklı kavramların ilişkisi

Süreç madenciliği için kullanılan farklı tanımlamaların temelinde farklı ticari nedenler vardır. Farklı satıcılar, farklı bakış açılarını vurgulamak ister (keşif yerine zeka'yı kullanmak gibi). Ancak manifesto da belirtilmiş olan disiplini kapsamak için "süreç madenciliği" tanımını kullanmak gereklidir.

Sözlük

Activity(Aktivite)

Süreçlerde iyi tanımlanmış adımlardır. Olaylar; özel bir süreç örneği için başlama, tamamlama, iptal edilme vs gibi aktiviteleri içerir. Süreç keşfine bakınız.

Automated Business Process Discovery

(Otomatik İş Süreç Keşfi):

Business Intelligence (BI):

Karar verme sistemi desteklemek için kullanılan metod ve araçların toplamıdır

Süreç Madenciliği Manifestosu

Business Process Intelligence(İş Süreç Zekası): Business Process Management (BPM) (İş Süreç Yönetimi):	İş Süreç Zekası: Süreç Zekasına bakınız Hem bilgi teknolojisi hem de yönetim bilimleri disiplinindeki bilgileri toplayarak her ikisini de operasyonel iş sürecinde uygular Süreç örneğine bakınız
Case (Durum) Concept Drift (Konsept Kayması):	Süreçlerin zamanla değişmesiyle ilgilidir. İzlenen süreç kademeli olarak (ya da aniden) zamansal ya da artan rekabet durumuna göre değişir, bu nedenle analiz karmaşıklaşır.
Conformance Checking(Uygunluk Kontrolü):	Log kaydının ve modelin gerçekle uyumunun analizidir. Uygunluk kontrolünün hedefi değişiklikleri tespit etmek ve bunların derecesini ölçmektir. Uygunluk kontrolü süreç madenciliğindeki temel üç çeşitten birisidir
Cross-Organizational Process Mining(Karşılıklı Organizasyonel Süreç Madenciliği):	Farklı organizasyonlar tarafından gerçekleşen olay loglarında süreç madenciliği tekniklerini uygulamaktır.
Data Mining (Veri Madenciliği):	Genellikle geniş veri grupları olmak üzere; veri gruplarındaki beklenmeyen ilişkileri için analiz etmek, ortaya çıkarmak ve bu yeni görüşü kanıtlayacak şekilde veri gruplarını özetlemektir.
Event(Olay):	Log da kaydedilmiş bir harekettir yani belirli bir süreç örneği için başlangıç, tamamlanma ya da bir iptal durumu olabilir.
Event Log (Olay Logu):	Süreç madenciliği için girdi olarak kabul edilen olayların toplamıdır. Olayların farklı loglarda saklanmasına gerek yoktur. (yani, olaylar farklı veritabanı tabloları üzerine yayılmış olabilir.)
Fitness (uygunluk):	Olay logunda gözlenen olayın model üzerinde uygulanırken ne kadar yeterli olduğunu belirleyen bir ölçümdür. Eğer başlangıçtan sona kadar gerçek loglar model üzerinde uygulanabilirse model en mükemmel seviyede yeterli demektir.
Generalization (Genelleme):	Modelin elimizde olan davranışa göre uygunluğunu ölçmektir. Uygunluğun üzerindeki bir model yeterince genelleştirilemez.
Model Enhancement (Model Geliştirme):	Süreç madenciliğinin üç temel özelliğinden biridir. Log dan ortaya çıkarılmış olan bilgi ile model geliştirilir ya da genişletilir. Kısır döngüleri örnek verebiliriz. Zaman notlarına dikkat edilerek yeniden oynatılan bir olay logunda kısır döngüler belirlenir.
MXML:	Olay loglarını değiştirmek için XML temelli bir formattır. MXML'in yerini yeni araçlardan bağımsız süreç madenciliği formatı olan XES almaktadır.
Operational Support (Operasyonel Destek):	Devam eden süreç örneğini izlemek ve etkilemek için olay loglarının online analizidir.3 farklı operasyonel destek aktivitesi tanımlanabilir: belirlemek(normal davranışlardan sapmalar

Süreç Madenciliği Manifestosu

Precision (Hassasiyet):	olduğunda bir alarm sistemi oluşturmak), tahmin etmek(geçmiş davranışlar temelli olarak gelecek davranışları hakkında ön görüde bulunmak, örneğin kalan zamanı tahmin etmek gibi) ve önermek(örneğin maliyetleri düşürmek gibi belirli bir hedef için uygun eylemleri önermek) Olay loglarındaki davranış ile modeldeki davranışın hassasiyetleri arasındaki farkın ölçülmesidir. Düşük hassasiyetlerdeki model ortalama uyumun altındadır.
Process Discovery (Süreç Keşfi):	Süreç madenciliğinin üç çeşidinden birisidir. Olay logundan süreç modelinin öğrendiğini kabul eder. Örneğin; olayların toplanmasından sonra alfa algoritması ile analiz edilmesinin sonucunda, olaylar içerisindeki desenler tespit edilir ve süreçleri gösteren bir Petri ağı oluşturur.
Process Instance (Süreç Örneği):	Analiz edilen süreçte ele alınan örnektir. Olaylar süreç örneklerinden oluşur. Süreç örnekleri olarak; müşteri siparişleri, sigorta talepleri ve kredi uygulamaları verilebilir.
Process Intelligence (Süreç Zekası):	İş süreç madenciliği üzerinde odaklanmış iş zekası dalıdır.
Process Mining (Süreç Madenciliği):	Günümüz veri tabanlarında erişilebilir olan olay loglarından bilgiyi ortaya çıkarma yöntemleri, araçları, keşif metotlarıdır. Buna ek olarak gerçek süreçleri izler ve geliştirir.
Representational Bias (Temsili Değerler):	Süreç madenciliği sonuçlarını yapılandırmak ve sunmak için seçilen hedef dildir.
Simplicity (Basitlik):	Bu prensip Occam'ın Razor'u olarak bilinen ölçümdür. Yani; en basit model ise logda görülen davranışı en basit şekilde açıklayandır ve bu model en iyi modeldir.
XES:	Olay logları için bir XML standartıdır. IEEE Task Force tarafından süreç madenciliği için geliştirilmiştir. Olay loglarını belirli bir formata dönüştürmek için kullanılır.(www.xesstandard.org)