

Yazılım Geliştirme ve Test Döngüsü Üzerinde Süreç Madenciliği Yaklaşımı

Rabia Saylam^{1*} ve Ozgur Koray Sahingoz²

¹ Bilgisayar Mühendisliği Bölümü, Hava Harp Okulu, İstanbul, Türkiye
rsaylam@hvkk.tsk.tr

² Bilgisayar Mühendisliği Bölümü, Hava Harp Okulu, İstanbul, Türkiye
sahingoz@hho.edu.tr

Özet. Veri madenciliği ile başlayan büyük verinin işlenmesi ve analiz edilmesi yaklaşımı, son yıllarda süreç odaklı bir araştırma alanını ortaya çıkardı. Süreç madenciliği, bugünün bilgi sistemlerinde kolayca erişilebilen olay günlükleri veri bilgisiyle gerçek süreçleri ortaya çıkartmayı, keşfetmeyi, kontrol etmeyi ve geliştirmeyi amaçlayan bir iş süreçleri yönetimi teknolojisidir. İş Süreç Modelleme (Business Process Modelling-BPM) analizlerini geliştirmeyi hedefleyen, nispeten yeni ve gelişmekte olan bir araştırma alanıdır. Süreç madenciliğinin en zor kısmı, uygun verinin uygun tablo ve alanlardan çıkarılmasıdır. Bu çalışmada, SAP kullanan bir yazılım kurumundaki yazılım geliştirme ve test döngüsü olay günlükleri alınarak bir analiz yapılmaktadır. Binlerce verinin bulunduğu bu sistemde, gerekli veriler ilgili tablolardan çıkarılmaktadır. Veri, süreç madenciliği araçları için geçerli format olan MXML formatına dönüştürülmektedir. Süreç madenciliği yani veri işleme aşamasına gelindiğinde ProM ve DISCO araçları açıklanmaktadır. Yazılım geliştirme ve test döngüsü sürecinin bir yıllık verisi ile organizasyondaki olgu, aktivite, kişi ve zaman istatistikleri ortaya konmaktadır.

Anahtar Kelimeler: Süreç Madenciliği, ProM, DISCO, Süreç Madenciliği Analizleri.

1 Giriş

Bilgi sistemleri kabiliyetlerinin son 50 yılda büyük oranda genişlediği görülmektedir. Bu ilerlemeler sonucu saklanan ve sistemler arası taşınan veri miktarında olağanüstü artışlar meydana gelmektedir. Kurumlar değişen ve gelişen çevre ile birlikte iş süreçlerini bu verileri kullanarak mevcut duruma adapte etmek zorundadırlar. Ancak tüm değişiklikler sonucunda başlangıç modelinden farklı ya da uygulanamayan modeller ortaya çıkmaktadır. Bu modelleri geliştirmek için insanların ve prosedürlerin nasıl çalıştığının süreç olay loglarından ortaya çıkması faydalı olabilir.

Süreç, girdilerin çıktılara dönüşümünü sağlayan birbiriyle ilişkili veya etkileşen faaliyetler kümesidir(ISO 9000). İş süreçleri madenciliği ya da süreç madenciliği, kısa bir biçimde, saklanan verinin analizi için yenilikçi bir analiz yolu getiren ve olay günlükleri ile karar vericilere büyük resmi sunarak iş süreçlerini iyileştirmeyi amaçlayan gelişmekte olan bir araştırma alanıdır. Bu özelliği nedeniyle, bazı araştırmacılar, süreç madenciliğini bir tür makine öğrenme görevi olarak tanımlamaktadır [1]. Ana fikir, analiz bilgi sistemi tarafından kaydedilen günlüklerden elde edilen bazı kritik bilgileri öğrenmek ya da çıkarmaktır. Temelde, olay günlükleri analiz sisteminde bir faaliyet ya da duruma bakarak, olaylar hakkında gerekli bilgileri içerir. SAP gibi *Kurumsal Kaynak Planlama sistemleri*, insanlar ve işlemleri kapsayan tüm işlemleri kaydettiğinden süreç madenciliği için iyi bir örnek olarak kabul edilebilir. Öte yandan, Müşteri İlişkileri Yönetimi sistemleri, müşterileri ile olan ana etkileşimleri loglarlar. Bu tip süreçler yeterince analiz edilemezse, darboğazları yakalamak veya atlanan alt süreçleri yakalamak çok zor olacaktır. Bu nedenle, süreç madenciliği teknikleri kullanılarak hata yapma olasılığı azalır ve fırsatları yakalama şansı artar.

2 Literatür Araştırması

Süreç Madenciliği, Şekil 1 [2]'de gösterildiği gibi iş zekâsı şemsiyesi altında karar verme sistemlerini destekleyici bilgiler sunmayı hedefler. Bu noktada (İş) süreç zekâsı ve modelleme notasyonları anlaşılmalıdır.



Şekil 1 : İş Zekâsı Şemsiyesi

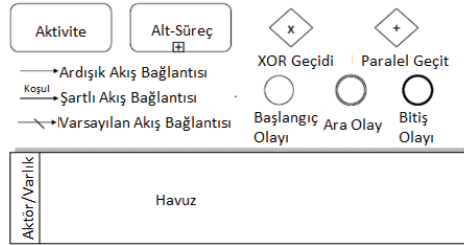
2.1 İş Süreçleri Zekâsı

Süreç günlüklerinin bir ambar içerisinde temizlenmesi, toplanması ve iş süreçleri zekâsı (Business Process Intelligence-BPI) teknolojileri ile analiz edilmesi sonucu geçmişte meydana gelen durumlar hakkında bilgi sahibi olunabilir. Bunun yanında

çalışan süreçlerin potansiyel problemlerini tahmin etmek ve neden ortaya çıktığını önceden açıklamak için yine bu bilgiler kullanılabilir 3.

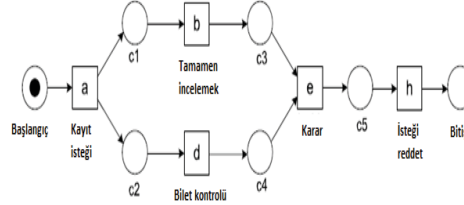
2.2 Süreç Modelleme Notasyonları

İş süreçlerine dayalı yazılım ürünleri geliştirirken süreç adımları belirlenir ve bazı gösterim sistemleri bunları formüle eder. Süreç madenciliği tekniği BPMN ya da Petri Ağları gibi belirli bir dil kullanarak modeller üretir 2. İş Süreçleri Modelleme Notasyonu (BPMN), tüm katılımcılar arasında iletişim, el sıkışma ve dokümantasyon sağlayan bir grafik gösterimidir.



Şekil 2 : İş Süreçleri Modelleme Notasyonu

Şekil 2’de [4] BPMN’de kullanılan nesnelere yer almaktadır. Hafızada yer alan modelin BPMN’de süreçlere dönüşümü her bir aktör için boş bir havuz oluşturulmasıyla başlar. Süreç modelinde bir başlangıç olayı ve bir veya daha fazla bitiş olayı vardır. Eylem türüne bağlı olarak faaliyet ya da bir alt süreç oluşturulur. Bağlantılar, eylemlerin sırasını uygun olarak yaratılır. Özel Durumlar süreç modelinde ağ geçitlerine dönüşür [4].



Şekil 3 : Petri Ağı

Petri ağları [5] eş zamanlı modellemeye izin veren en eski ve en iyi araştırılmış süreç modelleme dilidir. Grafikselsel gösterimi sezgisel ve basit olmasına rağmen, yürütülebilir ve analiz edilebilirdir. Petri ağları yerler ve geçişleri içeren iki parçalı bir grafikdir. Ağ yapısı statiktir ama atış (ilgili aktivitenin etkinleşmesi ve meydana gelmesi) üstünlüğü ile yönetilen jeton ağ üzerinde dolaşır. Her yer bir daire her geçiş bir kare ile gösterilir. Şekil 3’te yer, geçiş ve jeton bulunan bir Petri Ağı örneği bulunmaktadır.

2.3 İlgili Çalışmalar

Rozinat ve arkadaşlarının yaptığı çalışmada [6]; Gofret tarayıcı üreticilerinden ASML şirketi test süresini azaltmayı amaçlamıştır. Gofret tarayıcı için test çok önemli ve zaman alıcı bir süreçtir. Bu amaçla test süreçlerine süreç madenciliği teknikleri uygulanmıştır. Hata tespiti yerine test sürecinin bizzat kendisine odaklanılmıştır. Test süreci olay günlükleri kullanılabilirliği ile kolayca kaydedilmektedir. Süreç madenciliği aracı olarak ProM kullanılmakta olup ana soru: “Aslında testler nasıl uygulanır?”dır. Test faaliyetlerinin sipariş ve sıklığını gösteren süreç modelleri olay günlükleri temel alınarak oluşturulmakta ve daha sonra ideal model ile karşılaştırılmaktadır. Sonunda, test sürecinde en çok zaman harcanan süreç ortaya çıkmaktadır.

Bir başka çalışma da Mans ve arkadaşları tarafından yapılmıştır [7]. Bu çalışmada sağlık alanında süreç madenciliği yaklaşımına odaklanılmıştır. Aynı tanı konulmuş bir hasta grubu için prosedür (bakım yolları) tanımlamak ve bölümler arası işbirliği sağlamak için süreç madenciliği kullanılabileceği gösterilmiştir. Çalışmada süreç madenciliği; kontrol akış perspektifi, organizasyon perspektifi ve performans perspektifi olmak üzere üç perspektife dayanmaktadır.

Başka bir çalışmada [8] süreç madenciliği teknikleri ilk kez gerçek bir ticaret alanına uygulama olarak alınmaktadır. Kullanılan teknikle, denetim alanında mevcut değerlendirme uygulamalarıyla ortaya çıkması beklenmeyen olağandışı işlemler ortaya çıkarılmaktadır. Amaç iç denetçilerin mevcut denetim tekniklerini kullanarak tespit edemedikleri olağan dışı işlem ve ilişkileri algılamalarına izin vermektir. Süreç madenciliği teknikleri kullanılarak; onaylanmamış ödemeler, şirket prosedürleri ihlalleri veya görev çek farkları gibi ilginç durumlar tespit edilmektedir.

Aalst ve Gunther'in çalışmasının [9] iki farklı görüşü vardır. İlk olarak klasik sentez yaklaşımların süreç madenciliğine göre “aşırı uygunluğa” eğilimli olmaları sebebiyle yararlı olduğu kanıtlanmaktadır. Süreç madenciliği “yetersiz uygunluk” ve “aşırı uygunluk” arasında denge kurar. İkinci olarak gerçek yaşam süreçleri çok kullanılan süreç keşif teknikleri ile "spagetti model gibi" modellenmiştir. Bu gerçek çoğunlukla yaşam süreçlerinin karmaşık ve yapılandırılmamış süreçler olmasından kaynaklanmaktadır. Bu nedenle model basitleştirme teknikleri bu çalışmada sunulmaktadır. Soyutlama ve kümeleme daha iyi anlaşılır modeller oluşturmak için kullanılmaktadır.

3 Süreç Madenciliği ve Geliştirilen Uygulama

3.1 Süreç Madenciliği

IEEE çalışma grubuna göre üç çeşit süreç madenciliği tipi vardır. Bunlar sırasıyla; öncü bilgiler olmadan olay loglarından model oluşumu ile ilgilenen **Keşif**; var olan süreçler ile loglardan keşfedilmiş süreçlerin karşılaştırılması ve uyumluluk ve/veya sapmaları tespit edilmesini amaçlayan **Uyum Kontrolü Tekniği**; ve son olarak önce-

den var olan modeli olay loglarından elde edilen bilgilerle genişletme ya da geliştirmeyi amaçlayan *Geliştirme*'dir [10].

ProM [5] en yaygın ve popüler süreç madenciliği aracıdır. Süreç Madenciliği Platformlarına ayrı bir araç olarak eklenti olabilen birçok algoritma vardır ve ProM tüm madencilik özelliklerini tek bir sistemde kapsayan en büyük platform olarak kabul edilir. Farklı algoritmalar süreçleri keşfetme, sosyal ağ analizi ya da iş kurallarını doğrulama gibi farklı amaçları hedefler [2]. Bu araç açık kaynak ve genişletilebilirdir. Başka bir deyişle yeni eklentiler yaratılarak geliştirilebilir. Bu güne kadar 280'in üzerinde eklenti bu araca dahil edilmiştir. En uygun ve önemli olan eklentiler madencilik üzerine olanlardır. Alfa Algoritması, Sezgisel Madencilik Algoritması ve Sosyal Ağ Analizi kullanılan süreç madenciliği algoritmalarıdır.

DISCO, Fluxicon [11] tarafından tamamen süreç madenciliği için hazırlanmış bir yazılımdır. DISCO; otomatik süreç keşfi, süreç haritası animasyonu, detaylı istatistikler, filtreleme, veri girdi-çıkışı ve proje yönetimi gibi amaçlar için kullanılmaktadır.

3.2 Geliştirilen Uygulama

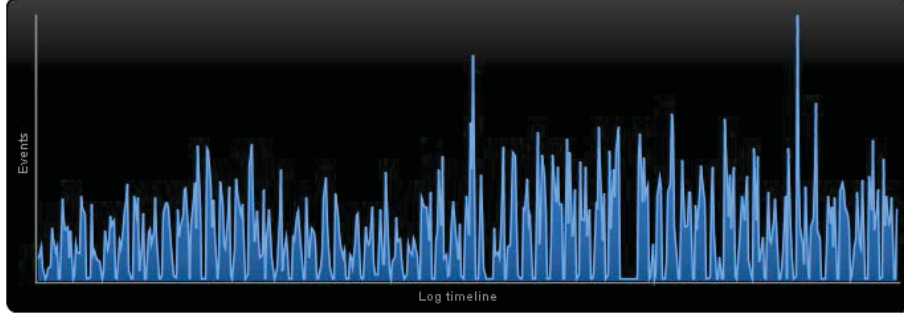
Uygulama SAP (Veri İşlemede Sistemler, Uygulamalar ve Ürünler) kullanan bir kuruluştaki Yazılım Geliştirme ve Test döngüsü süreci olay günlüklerinin ayıklanarak yapılmıştır [12]. 30.000'in üzerinde personeli olan, Mali Muhasebe (FI), Malzeme Yönetimi (MM), Kalite Yönetimi (QM), Proje Sistemleri (PS) ve İnsan Kaynakları (İK) modüllerinde yazılım geliştirme ve bakım sorumluluğu olan yazılım kurumundaki 3000'in üzerinde yazılım geliştirme süreci durumu keşfedilmektedir.

Yazılım değişiklik isteği durumunda ihtiyaç tanımlamasından sonra ilk olarak değişiklik yapılacak işlem koduyla ilişkili SOLMAN mesajı yaratılır. Kişiler arasındaki mesaj iletimi "*Message Processor*" alanına hedef kişinin ID'sinin eklenmesi ile gerçekleşir. Mesajın durumu "*Status*" alanından güncellenir. Bu mesaj tekildir. Her mesaj bir duruma karşılık gelir ve "*kişi*", "*durum*" ve "*zaman*" değişkenlerine sahiptir. Tasarlanan bir programla "mesaj numarası, kullanıcı, zaman, atanan kişi ve atanan statü" verileri elde edilmektedir. Bu veriler XLS formatında alınarak önce CSV sonra MXML formatlarına dönüştürülmüştür.

Dönüşüm sonucunda çıkan istatistikler [13]'de yer alan kılavuz ışığında değerlendirilmiştir.

DISCO aracı ile alınan analizler aşağıda açıklanmıştır:

Şekil 4 **Zaman İçerisinde Aktiviteler** grafiğinin yatay eksen (log-zaman) logun kapsadığı toplam zaman çerçevesini temsil eder. Dikey eksen, üzerinde işlem yapılan faaliyet sayısı çizilerek aktivite düzeyini gösterir. Organizasyonun bir yıl içerisindeki istatistiği incelendiğinde yapılan işlem sayısının yılın ikinci yarısında ilk yarıyıldan yüksek olduğu, özellikle Mayıs-Haziran aylarında ise ortalamadan düşük olduğu gözlemlenmektedir.



Şekil 4 : Zaman İçerisinde Aktiviteler

Şekil 5 **Zaman İçerisinde Aktif Olgular** grafiği aynı zamanda devam etmekte olan olgu sayısını göstermektedir. Yeni durumlar eklendiğinde dikey ekseninde artış olur. Durumlar tamamlandığında aktif durum seviyesinde düşme olur. Grafikte aktif olgu sayısının Temmuz ayı sonunda en üst noktaya çıktığı gösterilmektedir..



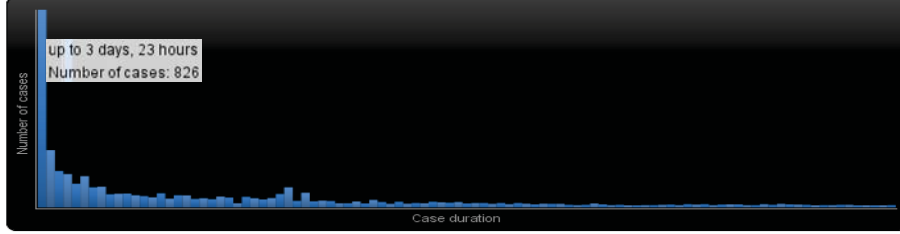
Şekil 5 : Zaman İçerisinde Aktif Olgular

Şekil 6 **Olgu Başına Aktiviteler** grafiği süreç boyunca meydana gelen aktivite dağılımını gösterir. Yatay ekseninde soldan sağa doğru artan şekilde aktivite sayısı, dikey ekseninde olgu başına aktivite sayısı görüntülenir. Bu grafikte uygulama verilerinden çıkan sonuca göre olguların büyük kısmının 3-6 aktiviteden oluştuğu gösterilmektedir.



Şekil 6 : Olgu Başına Aktiviteler

Olgu süreleri Şekil 7’de gösterilmektedir. Olguların 1/3’nün ilk üç günde tamamlandığı gösterilmektedir. Olguların tamamının ise genel olarak ilk üç ayda tamamlanmaktadır.



Şekil 7 :Olgu süreleri

Aktivite istatistikleri Şekil 8’de verilmektedir. Uygulamada 17 farklı aktivite bulunmaktadır. Çizelge 1’de aktiviteler ile meydana gelme sıklıkları verilmektedir. Kullanımı seyrek aktivitelerin gereksinimi tartışılmaktadır.



Şekil 8 : Aktivite İstatistikleri

Çizelge 1 : Aktiviteler ve meydana gelme sıklıkları

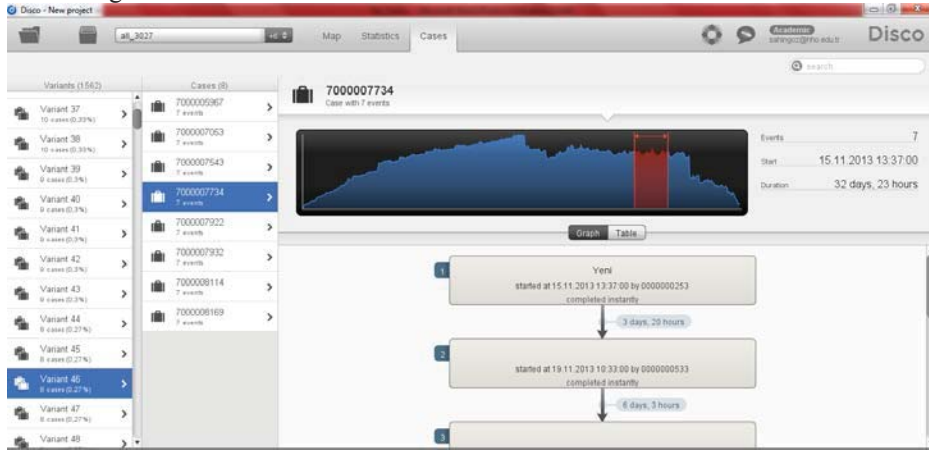
Aktivite	Meydana gelme sıklığı
Yeni	2739
Analiz	508
Uyarlama	67
Geliştirmesiz Tamamlandı	346
Tasarım Bekler	316
Tasarım	252
Tasarım Düzenle	70
Tasarım Onay	556
Geliştirme Bekler	1125
Geliştirme	737
Teknik Karar	42
Geliştirme Testi	1420
Geliştirme Hata	71
Fonksiyonel Onay	669
Test – Tekrar Test	1236
Test Hata	59
Test Başarılı	2020

Kaynak yani **personel istatistikleri** incelendiğinde sistemde 133 kaynak bulunduğu görülmektedir. Şekil 9’de özet grafik ve alt tarafta tablo formatında detaylı kaynak durumları gösterilmektedir. Sık adı geçen kaynaklar incelendiğinde; en dikkati çeken 217 (959 defa) numaralı kaynaktır. Bu kaynak yani aktörün “*Geliştirme Lideri*” olduğu görülmektedir. 181, 373 ve 169 gibi en çok sıklıkla (300-500 defa) çalışan kişiler ise *Modül Yöneticileri*’dir. Modül yöneticileri geliştirmeye hazır hale gelen tasarımı geliştirme liderine göndermektedir. Geliştirme liderinde biriken işler bir süre sonra darboğaz oluşturmaya başlamaktadır. Sistem loglarıyla su yüzüne çıkan bu sonuçlar ile organizasyonda değişikliğe gidilmiş, geliştirme lideri aktörü kaldırılmış ve bu görev modül yöneticilerine aktarılmıştır.



Şekil 9 : Kaynak İstatistikleri

Süreçte aynı diziyi izleyen çoklu aktiviteler olabilir. DISCO’da bu aktivite dizilerine varyant denir. **Varyant** sekmesi ile her bir varyantın kaç farklı aktivite izlediği, ortalama süreleri izlenebilir. Şekil 10’da 46 numaralı varyantın detayları gösterilmektedir. 8 olgunun takip ettiği varyantta; 7 farklı event bulunmakta olup ortalama tamamlanma süresi 39 gündür.



Şekil 10 : Varyant Detayları

4 Sonuç

Süreç madenciliği ile kurumlar; süreçlerini anlama, sürekli olarak iyileştirme ve denetleme kabiliyeti kazanmaktadır. Tasarımın içerisinde tahmin edilemeyen riskler

ve fırsatlar gerçek süreç akışlarının ortaya konulması ile belirlenmekte; düzeltici işlemler ciddi zaman ve maliyet tasarrufları sağlayabilmektedir. Gerçek verilerden yola çıkan bu yaklaşım iş süreçlerini yeniden kurmak için gerekli verileri otomatik olarak bulmaktadır. Süreç madenciliği sayesinde karar yolları, kontrol akışı ve performans gibi yürütme bilgileri elde edilebilir. Bu bilgiler kullanılarak simülasyon modelleri üretmek mümkündür.

Çalışma kapsamında yeni bir alan olarak anılan bu yaklaşıma ait araştırma ve geliştirmeye yönelik konular belirlenerek bir çalışma yapılmıştır.

Bu çalışmada Süreç Madenciliği, ProM ve DISCO Platformu ve Süreç Madenciliği açıklanmaktadır. Olgu çalışması ise, 30.000'in üzerinde personeli olan bir kurumun, Mali Muhasebe (FI), Malzeme Yönetimi (MM), Kalite Yönetimi (QM), Proje Sistemleri (PS) ve İnsan Kaynakları (İK) gibi SAP modüllerinde yaptığı yazılım geliştirme ve bakım çalışmaları loglarından yürütülmektedir. Süreç madenciliğinin en zor kısmı olarak anılan veri toplama aşaması için ilgili loglar ilişkili tablolardan çıkarılmakta ve süreç madenciliği platformları için gerekli olan formlara dönüştürülmektedir.

Organizasyondaki olgu, aktivite, zaman ve kişi istatistiklerini ortaya koymak amacıyla DISCO aracı kullanılmaktadır. 3000 üzerinde durumla yapılan analizler sonucunda; özellikle mayıs-haziran aylarında aktivite sayılarında ki düşüşler, olguların aktivite sayılarında farklılıklar, kişilere dağılan aktivite sayılarındaki ciddi farklılıklar, olgu kapanma süreleri ve detayları sunulmaktadır. Çalışmanın devamında ortaya çıkan sonuçlarla organizasyonda bazı değişiklikler olabilecek ve bu değişiklikler sonucu oluşan loglar incelenecektir.

Referanslar

1. Weber, P.; Bordbar, B.; Tino, P., "A Framework for the Analysis of Process Mining Algorithms," IEEE Transactions on Systems, Man, and Cybernetics: Systems, , vol.43, no.2, pp.303,317, March 2013.
2. IEEE Task Force on Process Mining, Process Mining Manifesto (Süreç Madenciliği Manifestosu-Türkçeye çeviren: Aylin Şahin), in BPM Workshops, volume 99 of Lecture Notes in Business Information Processing. SpringerVerlag, Berlin, 2011, 169–194.
3. Casati F, Dayal U, Sayal M, Shan MC, "Business Process Intelligence", Software Technology Laboratory, HP Laboratories Palo Alto, adres: <http://www.hpl.hp.com/techreports/2002/HPL-2002-119.html>
4. Sinha A, Paradkar A, "Use Cases to Process Specifications in Business Process Modeling Notation", 2010 International Conference on Web Services.
5. Wil M.P. van der Aalst, "Process Mining: Discovery, Conformance and Enhancement of Business Processes", 2010 by Springer.
6. A. Rozinat, I.S.M. de Jong, C.W. Günther, and W.M.P. van der Aalst, "Process Mining Applied to the Test Process of Wafer Steppers in ASML", IEEE Transactions On Systems, Man, And Cybernetics—Part C: Applications and Reviews, 2009

7. R. S. Mans, M. H. Schonenberg, M. Song, W. M. P. van der Aalst, P. J. M. Bakker, "Application of Process Mining in Healthcare – A Case Study in a Dutch Hospital", *Biomedical Engineering Systems and Technologies Communications in Computer and Information Science* Volume 25, 2009, pp 425-438, Springer Berlin Heidelberg.
8. Jans M., Alles M. and Vasarhelyi M, "Process Mining of Event Logs in Internal Auditing: A Case Study", 2nd International Symposium on Accounting Information Systems, Rome, 2011.
9. W.M.P. van der Aalst and C.W. Günther, "Finding Structure in Unstructured Processes: The Case for Process Mining", *Proceedings of the Seventh International Conference on Application of Concurrency to System Design*. p. 3–12, IEEE, 2007.
10. Saylam R., Sahingoz O.K., "Process mining in business process management: Concepts and challenges", *International Conference on Electronics, Computer and Computation (ICECCO)*, IEEE, 2013.
11. Process Mining and Process Analysis - Fluxicon Adres:
<http://www.fluxicon.com/> Erişim Tarihi: 11.04.2014
12. Saylam R., Sahingoz O.K., "A Process Mining Approach in Software Development and Testing Process", *World Congress on Engineering, WCE 2014*, Accepted.
13. Process Mining and Automated Process Discovery Software for Professionals - Fluxicon Disco. Adres: <http://fluxicon.com/disco/> Erişim Tarihi: 11.04.2014