



**Jorge Miguel
Mourinha Martins**

Obtenção de Processos de Negócio Informais

Descobrir, modelar e analisar processos
de negócio informais nas organizações

Dissertação/Trabalho de Projeto/Relatório de
Estágio submetida como requisito parcial para
obtenção do grau de Mestrado em Informática
de Gestão

Júri

A designar

Orientador: Nuno Miguel Vicente Pina
Gonçalves

Dezembro de 2014

Sumário

LISTA DE FIGURAS	4
LISTA DE TABELAS	6
AGRADECIMENTOS	7
RESUMO	8
ACRÓNIMOS	9
INTRODUÇÃO	10
INTRODUÇÃO DE CAPÍTULO	10
PROBLEMA	10
OBJETIVO	11
ENQUADRAMENTO	12
METODOLOGIA	12
RESULTADOS ESPERADOS	14
CONCLUSÕES DO CAPÍTULO	15
ESTADO DA ARTE	16
INTRODUÇÃO DE CAPÍTULO	16
RESUMO	16
DATA MINING	17
<i>Evolução Histórica</i>	17
<i>O que é?</i>	17
BUSINESS PROCESS MANAGEMENT	35
<i>Evolução Histórica</i>	35
<i>O que é?</i>	38
PROCESS MINING	47
<i>Evolução Histórica</i>	47
<i>O que é?</i>	48
DIFERENÇAS ENTRE DATA MINING E PROCESS MINING	60
FERRAMENTAS DE SUPORTE	61
CONCLUSÕES DO CAPÍTULO	63
METODOLOGIA	64
INTRODUÇÃO DE CAPÍTULO	64
PLANEAMENTO	65
<i>Introdução</i>	65
<i>Descrição</i>	65
ABORDAGEM	69
<i>Outlook</i>	70
<i>Carregamento da informação</i>	71
<i>Tratamento da Informação</i>	72
<i>Gestão de Eventos</i>	73
<i>Gestão de Recursos</i>	74
<i>Clustering</i>	76
<i>Rules</i>	77
<i>Text Mining</i>	79
<i>Componentes do Processo</i>	81

<i>XPDL</i>	82
<i>Resultados</i>	84
CONCLUSÕES DO CAPÍTULO	85
PROTÓTIPO	86
INTRODUÇÃO DE CAPÍTULO	86
APRESENTAÇÃO DO PROTÓTIPO	87
<i>Arquitetura</i>	87
<i>Pré-Requisitos Aplicacionais</i>	89
<i>Modelo de Utilização</i>	89
CONCLUSÕES DO CAPÍTULO.....	107
VALIDAÇÃO	108
INTRODUÇÃO DE CAPÍTULO	108
VALIDAÇÕES	109
CONCLUSÕES DO CAPÍTULO	112
CONCLUSÕES E TRABALHO FUTURO	113
INTRODUÇÃO DE CAPÍTULO	113
CONCLUSÕES.....	114
TRABALHO FUTURO	115
REFERÊNCIAS	116

Lista de Figuras

Figura 1 - Exemplo de uma árvore de decisão (Aalst, W. M. P. van der., 2011).....	24
Figura 2 - Exemplo de <i>Clustering</i> com duas variáveis em estudo, peso e altura (Aalst, W. M. P. van der., 2011).....	24
Figura 3 - Exemplo de uma abordagem <i>top down</i> do algoritmo <i>agglomerative hierarchical clustering</i> . Adaptado de (Yong Joseph Bakos, 2014)	26
Figura 4 - Exemplo de uma abordagem <i>bottom up</i> do algoritmo <i>agglomerative hierarchical clustering</i> . Adaptado de (Yong Joseph Bakos, 2014)	26
Figura 5 - Fórmula de cálculo da métrica <i>support</i> . (Aalst, W. M. P. van der., 2011).....	28
Figura 6 - Fórmula de cálculo da métrica <i>confidence</i> . (Aalst, W. M. P. van der., 2011).....	28
Figura 7 - Fórmula de cálculo da métrica <i>lift</i> . (Aalst, W. M. P. van der., 2011)	28
Figura 8 - Exemplo de três episódios distintos. (Aalst, W. M. P. van der., 2011).....	30
Figura 9 - Matriz de confusão para duas classes. (Aalst, W. M. P. van der., 2011).....	32
Figura 10 - Métricas para medir a performance de um classificador. (Aalst, W. M. P. van der., 2011) 33	
Figura 11 - Exemplo de funcionamento da validação cruzada. (Aalst, W. M. P. van der., 2011)	34
Figura 12 - Processo de Negócio.....	38
Figura 13 - Análise histórica da evolução dos sistemas de BPM (José António Sena Pereira, 2013). 39	
Figura 14 – Atividades chave do BPM, adaptado de (Aalst, W. M. P. van der., 2012)	42
Figura 15 - Representações gráficas do BPM, adaptado de (Aalst, W. M. P. van der., 2012) (W. M. P. van der Aalst, A. H. M. terHofstede, B. Kiepuszewski & A. P. Barros, 2003).....	43
Figura 16 - Sistemas que têm tecnologia BPM, adaptado de (Aalst, W. M. P. van der., 2012)	44
Figura 17 - Ciclo de vida do BPM, adaptado de (Aalst, W. M. P. van der., 2012) (Aalst, W. M. P. van der., 2004).....	45
Figura 18 - Três principais tipos de <i>process mining</i> : <i>discovery</i> , <i>conformance</i> e <i>enhancement</i> (Aalst, W. M. P. van der., 2011).....	48
Figura 19 - Terminologias sobre a relação entre modelos de processos e a realidade (Aalst, W. M. P. van der., 2011)	51
Figura 20 - Ciclo de Vida do BPM com as diferentes utilizações dos modelos de processos (Aalst, W. M. P. van der., 2011).....	57
Figura 21 - Esquema sobre a abordagem desenvolvida ao longo da investigação	69
Figura 22 - Exemplo de um <i>event log</i>	74
Figura 23 - Exemplo dos recursos guardados com vista à criação do <i>event log</i>	75
Figura 24 - Arquitetura aplicacional	88
Figura 25 - Ecrã Inicial	90
Figura 26 - Seleção da pasta a descarregar para a aplicação	91
Figura 27 - Carregamento da informação selecionada efetuado com sucesso	91
Figura 28 - Estatísticas relativas ao carregamento de itens	92
Figura 29 - Ecrã de visualização de eventos	93
Figura 30 - Ecrã que apresenta as estatísticas relativas aos eventos criados	94
Figura 31 - Ecrã onde é possível observar os recursos previamente carregados tendo em conta os eventos criados	94
Figura 32 - Ecrã onde são apresentadas as estatísticas relativas aos recursos descortinados	95
Figura 33 - Criação dos ficheiros de validação	96
Figura 34 - Apresentação dos Menus de configuração bem como do ecrã utilizado para efetuar a análise dos eventos.....	96
Figura 35 - Ecrã de configuração de eventos a serem analisados.....	97
Figura 36 - Ecrã de gestão de <i>roles</i>	98
Figura 37 - Ecrã inicial relativo à análise de eventos.....	98
Figura 38 - Ecrã relativo à pesquisa desenvolvida aos eventos a serem analisados	99
Figura 39 - Representação gráfica dos eventos	100
Figura 40 - Ecrã relativo aos eventos que irão ser analisados posteriormente.....	100

Figura 41 - Ecrã relativo à informação pormenorizada do evento	101
Figura 42 - Ecrã com o agrupamento dos eventos tendo em conta as suas propriedades	102
Figura 43 - Ecrã com as regras descortinadas relativamente aos recursos envolvidos nos eventos selecionados.....	102
Figura 44 - Ecrã relativo à análise desenvolvida ao corpo e anexo dos eventos selecionados.....	103
Figura 45 - Ecrã com a informação recolhida para a criação de um processo	104
Figura 46 - Ecrã de validação de <i>roles</i>	104
Figura 47 - Ecrã de validação de tarefas associadas a <i>roles</i>	105
Figura 48 - Ecrã com o conteúdo do ficheiro <i>XPDL</i>	105
Figura 49 - Modelação do processo descortinado	106

Lista de Tabelas

Tabela 1 - <i>Data Set</i> acerca dos pedidos num café.....	18
Tabela 2 - <i>Data Set</i> acerca da análise de uma pauta de estudantes de uma determinada escola.. ...	21
Tabela 3 - Diferenças entre o <i>process mining</i> e o <i>data minig</i>	60
Tabela 4 - Representação do planeamento sobre a forma de tabela.....	68

Agradecimentos

O desenvolvimento desta dissertação de mestrado teve a colaboração e apoios de extrema importância, sem os quais não teria sido possível dar por terminada e aos quais estarei para sempre grato.

Ao professor Nuno Pina Gonçalves, pela sua orientação, por toda a paciência demonstrada, pelos ensinamentos, pelo saber providenciado, pelo acompanhamento ao longo dos últimos 5 anos, 3 de licenciatura e 2 de mestrado, por todas as críticas saudáveis que fez, pelo tempo despendido e por todos os conselhos dados, estarei eternamente grato pois grande parte do meu crescimento a nível académico, profissional e pessoal lhe é devida.

Aos meus colegas de mestrado, Nuno Nunes e Nélio Letras, que deixaram de ser colegas para passarem a ser grandes amigos, pois sem todo o conhecimento que me forneceram, a disponibilidade total para ajudarem seja no que for, os momentos complicados que juntos superamos, o companheirismo e acima de tudo o mais importante, a amizade demonstrada, não seria possível dar por concluído este mestrado.

Há minha família, aos meus pais, há minha irmã e aos meus tios, aos meus pais e há minha irmã pois sem eles nada disto seria possível, toda a paciência que tiveram, todos os incentivos, todas críticas construtivas, toda a exigência solicitada, todo o apoio demonstrado, tudo o que me deram e os esforços que fizeram para poder frequentar e conseqüente adquirir um curso superior, aos meus tios por todo apoio demonstrado, por todas a palavras de incentivo e por sempre terem estado presentes ao longo de todo o meu percurso.

Há minha namorada, pela sua paciência, por todo o tempo que cedeu para que eu pudesse dar por terminada esta dissertação de mestrado, por estar presente tanto nos bons como nos maus momentos, pelas constantes palavras de incentivo e pela sua forma de ser que a torna única aos meus olhos.

A todos os professores do instituto politécnico de Setúbal que tive o privilégio de conhecer e que contribuíram para o meu crescimento quer a nível profissional quer a nível académico.

Por fim, mas não o menos importante, bem pelo contrário, quero dedicar esta dissertação de mestrado e conseqüente conclusão deste mestrado aos meus avós que infelizmente já não estão cá para ver, porque lhes devo muito e sem eles eu não seria o que sou hoje.

Resumo

Na era das redes sociais, sejam elas ao nível empresarial ou ao nível interpessoal, existe um elemento comum e fundamental nos dias de hoje, a comunicação.

Tendo em conta a comunicação supracitada podemos dividi-la em comunicação formal ou informal, tendo como base a informação informal que circula nos canais de comunicação das empresas torna-se imperativo apontar a nossa atenção para esta, por forma a descobrir novos processos negócio para a organização ou otimizar processos já existentes.

Toda a informação relativa à comunicação existente no núcleo das organizações descrita em cima é crítica, assim como a sua qualidade, rapidez e veracidade.

Nesta dissertação será desenvolvida uma metodologia de recolha e consequente análise à informação informal existente no seio das organizações, mais precisamente no que diz respeito aos dados que circulam no correio eletrónico (*Microsoft Outlook*).

A abordagem passará por recolher e consequentemente tratamento dos dados que circulam no correio eletrónico das organizações, sendo esta recolhida do *Microsoft Outlook*, como referenciado em cima. Este tratamento assenta primordialmente na transformação dos itens provenientes em eventos capazes de serem analisados, após o tratamento desses dados torna-se imperativo desenvolver metodologias para os analisar.

As metodologias utilizadas para analisar os dados prendem-se principalmente com a utilização de algoritmos de *data mining*, como o *clustering*, por forma a agrupar os eventos com temáticas semelhantes, posteriormente a utilização do *Apriori*, com vista à análise dos recursos que intervêm nos eventos e por fim a abordagem ao corpo e aos anexos dos eventos através do *texto mining*.

Posteriormente à análise de toda a informação, será possível obter os componentes necessários à criação de processos de negócio e consequente modelação, sendo que tanto a criação como a modelação assentam na criação de ficheiros *XPDL*.

Com esta análise será possível descobrir, validar e modelar novos processos de negócio que apresentem relevância para a organização, tendo em conta a realidade existente. Será também possível vislumbrar estatísticas, padrões e regras daí consequentes.

Acrónimos

BPM – *Business Process Management*;
BPEL – *Business Process Execution Language*;
BPMN – *Business Process Modeling Notation*;
BPMS – *Business Process Management Systems*;
SI – Sistema de Informação;
TI – Tecnologia da Informação;
UML – *Unified Modeling Language*;
ATM – *Automated Teller Machine*;
IRS – Imposto sobre os rendimentos;
PAISs – *Process-Aware information Systems*;
CRM – *Customer Relationship management*;
YAWL – *Yet Another Workflow Language*;
BAM – Business activity monitor;
CPM – *Corporate performance management*;
CEP – *Complex event processing*;
KPIs – *Key perform indicators*;
ABC – Activity based costing;
RCA – *Resource Consumption Accounting*; e
XPDL – *XML Process Definition Language*.

Introdução

Introdução de Capítulo

Este capítulo tem por base uma apresentação do tema e dos pressupostos utilizados para o desenvolvimento desta dissertação de mestrado.

Problema

De uma forma genérica e simplista, o problema a resolver surge da crescente necessidade das organizações otimizarem e identificarem novos processos de negócio, quer a nível interno, para uma melhor eficiência dos seus recursos, quer a nível externo, para uma melhor relação com os seus intervenientes.

Esta necessidade advém, em parte, do crescimento das organizações, pois com o aumento da sua complexidade hierárquica e a criação de estruturas departamentalizadas por área de negócio, surgiu a descentralização da informação. Consequentemente a manutenção e controlo total sobre as atividades da organização tornou-se complexa, pois muitas vezes existem diferentes chefias, diferentes localizações físicas, diferentes áreas de atividade, diferentes formas de comunicação entre os diferentes setores da organização.

Para além da dificuldade em termos de monitorização e posterior controlo de atividades, é imperativo desenvolver novos processos e otimizar processos já existentes, de modo a tornar os recursos organizacionais mais eficientes e mais capazes de responder a todas as necessidades.

Apesar de diversas organizações já utilizarem ferramentas capazes de desenvolverem processos de negócio, estes focam-se em processos formais, como por exemplo, o processamento salarial, mas existe uma crescente necessidade de desenvolver processos através de uma iteração informal entre os seus colaboradores, como por exemplo através da utilização do *Skype*, do *Outlook* ou do *Lync*.

Hoje em dia existem aplicações capazes de desenvolver este tipo de abordagem a nível empresarial, contudo estas ferramentas apresentam custos e como tal não existiu a possibilidade de efetuar uma análise completa aos seus pontos fortes e fracos. Em termos de ferramentas capazes de desenvolver este tipo de abordagem temos por exemplo o *TIBCO*, o *IBM BPM*, *BPM Suite*, o *Bonita BPM*, entre outros.

Objetivo

A falta de uniformidade nos processos de negócio de uma organização leva a que processos semelhantes possam ser efetuados de diferentes formas, caso sejam elaborados ou executados por diferentes intervenientes. Assim, apesar de potencialmente terem o mesmo *output*, as atividades realizadas para o atingir tendem a ser diferentes, levando a um controlo deficitário e bastante complexo de um processo que poderia facilmente ser monitorizado e uniformizado.

Processos como os supracitados tendem a ter uma elevada complexidade de concessão, devido principalmente à elevada quantidade de dados a analisar para discernir estes processos, pelo que não existem ferramentas tecnológicas capazes de os tornar autónomos e por sua vez uniformes. Estes apresentam também uma imperativa necessidade de intervenção humana no processo, pois geralmente implicam uma interação entre colaboradores da organização, sejam eles do mesmo ou de diferentes sectores ou departamentos podendo mesmo estar em localizações físicas distintas.

Neste âmbito, pretende-se criar novos processos, otimizar processos já existentes, criar estatísticas, analisar padrões ou regras, fazer previsões e alterar processos já existentes através de estatísticas ou previsões futuras. Estas atividades são possíveis através da utilização dos vários tipos de ferramentas utilizadas nas organizações para comunicação entre colaboradores, sendo estas ferramentas de diversos tipos, como por exemplo, *Skype*, *Lync*, *Outlook*, *Google*, entre outras.

Com o crescimento exponencial da utilização dos dispositivos moveis, estes têm vindo a desempenhar um papel cada vez mais importante no seio das organizações, pois apresentam uma portabilidade e uma autonomia bastante superior ao computador convencional, tornando assim a comunicação mais eficiente e mais célere.

No que diz respeito à extração de dados de dispositivos móveis, com a finalidade de identificar e criar processos de negócio, ainda não existem ferramentas capazes de realizar estas tarefas. Desta forma pretende-se utilizar as funcionalidades dos dispositivos móveis como as agendas, contatos frequentes, calendários, aplicações de planeamento/gestão de tarefas do dia-a-dia, *Evernote*, entre outras, para dar resposta a esta necessidade.

A utilização destas ou outras aplicações supracitadas possibilitará a criação de processos de negócio através da análise do histórico existente nos dispositivos móveis, sendo este relativo às atividades desenvolvidas pelo utilizador. Deste modo podem então ser criados processos que são realizados com um determinado espaço temporal constante pelo utilizador, criando assim estes processos e otimizando os processos já existentes. Podemos tomar como exemplo, a necessidade de aprovar um determinado documento todas as segundas feiras, sempre da mesma forma, pode então ser criado um mecanismo capaz de o fazer, criando assim numa regra através de um reconhecimento de padrões efetuado.

Com o desenvolvimento desta ferramenta espera-se maior eficiência na comunicação entre os diversos setores de atividade das organizações, melhor utilização dos recursos, quer sejam recursos humanos, tecnológicos ou financeiros, maior capacidade de entendimento dos processos de negócio já existentes e dos futuros de modo a aumentar o nível de satisfação do cliente e aumentar o desempenho do negócio.

Enquadramento

Os processos organizacionais são fundamentais para o correto funcionamento e conseqüente crescimento a nível organizacional. Para fazer face à necessidade de ter processos de negócio operacionais bem delineados e otimizados é imperativo combinar o conhecimento das tecnologias de informação com os da gestão. (WMP Van Der Aalst, 2004)

Dentro desta área existem diversas subáreas do conhecimento e dos processos como por exemplo o *Process Mining* e o *Business Intelligence*, mas esta não é uma afirmação unânime no seio da comunidade científica pois existem autores que defendem que o *Business Intelligence* é a base do *Process Mining* e outros que defendem o contrário.

O *Process Mining* é uma área de investigação relativamente recente que se situa entre a aprendizagem máquina e o *Data Mining* por um lado e a modelação e análise. O principal objetivo do *Process Mining* é descobrir, monitorizar e melhorar processos de negócio operacionais reais, isto é, processos que não são previsíveis à partida, isto através da extração de conhecimento de *Event logs* disponíveis em praticamente todos os sistemas de hoje em dia. (WMP Van Der Aalst, 2011)

Será principalmente, a partir de técnicas de *Process Mining*, que incidirá o desenvolvimento de uma ferramenta capaz de discernir processos operacionais nas organizações através das ferramentas de comunicação já existentes no mercado de trabalho.

Metodologia

No passado as organizações tinham estruturas bastante rígidas e burocráticas, o que pressuponha uma comunicação vertical, mas com o passar do tempo foram alterando a sua filosofia e passaram a ter uma comunicação mais horizontal tornando assim a sua comunicação e a sua gestão mais flexível. A postura apresentada no modelo tradicional apresentava uma estrutura *top-down* o que provocava baixa qualidade e produtividade dos seus colaboradores devido ao elevado controlo e pouca confiança nos colaboradores.

Hoje em dia as organizações passaram a adotar uma estrutura *bottom-up*, menos focada no controlo e considerando o colaborador como um ativo e não como um custo, isto levou a que os colaboradores se sentissem parte da organização e consequentemente apresentassem melhores rácios de produtividade e qualidade.

Com o crescimento exponencial da tecnologia, esta assume um papel preponderante no dia-a-dia das organizações, desde as atividades operacionais até à comunicação entre os diversos departamentos ou setores passando pela partilha de conhecimento entre as diversas áreas do conhecimento existentes numa organização. Para efetuar uma gestão eficiente e alicerçar as capacidades tecnológicas e operacionais existentes, as organizações implementam sistemas de gestão de processos, mais conhecidos por BPMS – *Business Process Management Systems*.

Estes *Softwares* são direcionados à representação de processos de negócio e segundo (WMP Van Der Aalst, 2012) compreendem as seguintes fases do ciclo do BPM – *Business Process Management*:

- Análise dos dados;
- (Re) Desenho;
- Configuração e Implementação;
- Execução e Ajustamentos.

Esta é uma abordagem que vai de encontro às necessidades de grande parte das organizações e possibilita o desenvolvimento de uma solução genérica a nível organizacional, ou seja, possibilidade de ser aplicada em diferentes áreas de negócio.

Para o desenvolvimento desta ferramenta será utilizado o *Process Mining* como área específica do *Business Process Management*, sendo que esta tem como principais pressupostos descobrir, monitorizar e melhorar processos reais através da extração de conhecimento proveniente de *Event logs*. (WMP Van Der Aalst, 2013)

O desenvolvimento desta dissertação irá assentar numa metodologia *SCRUM*, que permitirá um desenvolvimento ágil, sendo criadas pequenas fases de entrega (*Sprint*), por forma a estar sistematicamente em sintonia com os diferentes intervenientes e a responder de forma rápida e eficaz a possíveis alterações do previamente planeado.

Em primeira mão, será desenvolvido um capítulo denominado estado da arte. Neste, irão ser desenvolvidas diferentes investigações, estas terão como base uma análise sobre a área dos processos de negócio, de modo a para adquirir conhecimentos gerais sobre a mesma, depois de adquiridos os mesmos, será desenvolvida uma investigação mais profunda acerca do *Process Mining*, pois será sobre esta área que incidirá em grande parte as referências para o desenvolvimento da ferramenta supracitada, esta investigação será tanto teórica como prática, pois existe a necessidade de ter uma elevada base teórica como fundamentação para o

desenvolvimento desta ferramenta, isto deve-se em grande parte, a esta dizer respeito uma área bastante recente na investigação.

Este capítulo será desenvolvido através da leitura e posterior análise de diferentes livros, artigos científicos e dissertações de mestrado ou doutoramento relacionados com tanto como o *BPM* como com o *Process Mining*.

Em contra ponto na parte mais prática será desenvolvida uma pesquisa acerca das ferramentas já existentes no mercado, de modo a poder observar as suas mais-valias e os seus pontos de melhoria, tendo em vista o desenvolvimento da ferramenta já apresentada, servirá também para saber se existem, ou não, no mercado ferramentas que possam ser uteis para o desenvolvimento pretendido.

Será também necessário analisar a existência ou não de ferramentas que sejam capazes de solucionar o problema apresentado, de modo a garantir a fiabilidade da motivação e da necessidade apresentada neste documento para o desenvolvimento desta ferramenta.

De seguida irá ser criado um plano de ação sobre os principais pontos a solucionar para o desenvolvimento do problema no seu todo, sendo este plano munido já de consistência teórica, fornecida pelo capítulo relativo ao estado da arte, e de consistência prática previamente elaborada sobre as ferramentas já existentes.

A fase seguinte irá ser a fase de desenvolvimento propriamente dita, pois é onde será criada a ferramenta proposta, tendo em conta a análise feita nos capítulos anteriores. Esta fase de desenvolvimento fará se acompanhar de um capítulo denominado caso de estudo, onde será aplicada a ferramenta, depois de desenvolvida, a um caso do dia-a-dia.

Por fim será elaborada uma validação para o caso de estudo, de modo a analisar a fiabilidade e a viabilidade fornecida pela ferramenta no contexto selecionado.

Resultados Esperados

O sistema de obtenção de processos informais através de ferramentas de comunicação utilizadas nas organizações pressupõe a utilização das mesmas de modo a descobrir e otimizar processos existentes no seio da organização.

Este pressupõe uma investigação aos conteúdos e componentes providenciados pelas ferramentas de comunicação, por forma a demonstrar que através da análise destes conteúdos, é possível identificar padrões existentes ao nível da comunicação informal no seio da organização, ou seja, será desenvolvida uma metodologia capaz de discernir processos simples, mas não menos importantes com vista à constante melhoria por parte das organizações. Por forma a validar esta

investigação será desenvolvido um protótipo capaz de validar e demonstrar que é possível aplicar em termos práticos as ilações retiradas da investigação desenvolvida.

A utilidade primordial estará centrada no desenvolvimento e otimização de processos informais de modo a tornar mais eficiente e mais fluida a comunicação entre os diversos colaboradores da organização, proporcionando também a monitorização deste tipo de processos ou atividades que hoje em dia são praticamente órfãs de um controlo formal.

Esta ferramenta possibilitará também às organizações uma maior agilidade na obtenção de novos processos de negócio e a otimizar outros já existentes.

Conclusões do Capítulo

Neste capítulo pretendeu-se fazer uma breve apresentação sobre a estrutura, o conteúdo e os objetivos desta dissertação de mestrado, com este espera-se que o leitor tenha apreendido de forma genérica o que será abordado ao longo do documento.

Estado da Arte

Introdução de Capítulo

Neste capítulo será desenvolvido um preambulo acerca de informação científica já existente nos dias de hoje, tendo em conta as principais áreas em que incide este estudo, o BPM, o *process Mining* e o *data mining*.

Resumo

Com este capítulo pretende-se abordar todos os conceitos científicos que serão utilizados ao longo desta dissertação, por forma a providenciar uma contextualização à investigação a desenvolver.

Esta abordagem terá como pontos fulcrais o *Data Mining*, onde serão abordadas as suas raízes históricas, a sua definição, os seus principais conceitos e por fim alguns dos algoritmos existentes que apresentem relevância para a investigação, outro ponto a ser abordado será o *Business Process Management*, à semelhança do que é apresentado no *Data Mining*, será elaborada uma abordagem história acerca do tema e depois serão apresentados os principais e mais relevantes conceitos para a investigação, por fim serão apresentados, para além da sua evolução histórica, os conceitos, as metodologias e as ferramentas relativas ao *Process Mining*, fundamentais para a compreensão e desenvolvimento desta investigação.

Data Mining

Evolução Histórica

O crescimento que tem sido verificado ao nível do universo digital nos últimos anos tem potenciado a evolução do *Data Mining*. Inicialmente não era comumente aceite na comunidade científica, devido em grande parte ao fato de apresentar dados em tempo real que não eram estatisticamente aceites, mas com o passar do tempo esta disciplina começou a ser aceite pelos seus métodos e pelas suas diversas aplicações. (D.Hand, H. Mannila and P.Smyth, 2001) (E. Alpaydin, 2010) (I.H. Witten and E. Frank, 2005) (M. Bramer, 2007) (T.M. Mitchell, 1997)

Data mining começou por ser utilizado na análise da relação entre produtos em supermercados, lojas de compras *online*, cafés, lojas de retalho, entre outras. Com esta análise é possível saber os produtos que os utilizadores mais compram, quais os que compram ao mesmo tempo que outros, quais os que compram num determinado período do ano, entre outras estatísticas que possam ser analisadas tendo em conta o histórico de compras dos utilizadores. (D.Hand, H. Mannila and P.Smyth, 2001) (E. Alpaydin, 2010)

Com o passar do tempo, passou a ser utilizado também noutras áreas da sociedade que não o retalho, sendo essas, a área da banca, dos seguros e do setor público. (Aalst, W. M. P. van der., 2011)

O que é?

Segundo (D.Hand, H. Mannila and P.Smyth, 2001), *Data Mining* é a análise que é feita a um conjunto de dados, com a finalidade de encontrar relações e padrões inesperados, bem como sintetizar os dados por forma a tornar a informação perceptível e útil para o utilizador.

Dados Input

Como referido no capítulo anterior, *Data Mining* é a análise que é desenvolvida em relação a um determinado conjunto de dados, sendo que esse conjunto, geralmente está representado sobre a forma de *Data Set*. (Aalst, W. M. P. van der., 2011)

Estes *Data Sets* podem ser compostos por dados de diversos tipos e de diversos âmbitos, temos como exemplo a tabela 1 apresentada em baixo. Esta apresenta um *Data Set* que poderia ser utilizado numa análise de *Data Mining*, em que o caso de estudo seria as encomendas que são efetuadas por pessoa num café.

As colunas da tabela são os tipos de pedidos que podem ser feitos no café, enquanto que as linhas são o número relativo ao tipo de produto que foi encomendado por pessoa, isto é, uma pessoa encomendou um “Capuccino” e um “Bolo”, outro encomendo um “Expresso” e assim sucessivamente.

Capuccino	Expresso	Ristretto	Bolo
1	0	0	1
0	1	0	0
0	0	1	1
0	1	0	1

Tabela 1 - Data Set acerca dos pedidos num café. Dados apresentados meramente exemplificativos.

Cada *Data Set*, é composto por instâncias e variáveis, sendo que as instâncias são as linhas, estas também podem assumir outras denominações como por exemplo, entidades, casos ou objetos, e as variáveis são as colunas, que também podem ser denominadas por atributos ou características. (Aalst, W. M. P. van der., 2011)

Tendo em conta o exemplo apresentado na tabela 1, a instância é a encomenda efetuada por pessoa e as variáveis são “Capuccino”, “Expresso”, “Ristretto” e “Bolo”.

No que diz respeito às variáveis, estas podem ser divididas em dois grupos: (Aalst, W. M. P. van der., 2011)

- Variáveis categóricas – Estas, são facilmente enumeráveis, isto é, apresentam um conjunto limitado de possibilidades, Temos como exemplo as variáveis *booleanas*, pois estas só podem assumir o valor *true* ou *false*. Estas podem ainda ser divididas em dois sub grupos:
 - Variáveis ordenáveis – Estas, apresentam uma lógica intrínseca na sua ordenação. Temos como exemplo, uma pauta de notas de uma unidade curricular, onde pode ter resultados finais como, “desistiu”, “reprovou” ou “passou”;
 - Variáveis nominais – Estas, não apresentam uma lógica de ordenação, isto é, não têm uma correlação direta. Temos como exemplo, um conjunto de cores ou o nome dos países da união europeia.
- Variáveis numéricas – Estas, são facilmente ordenáveis, mas não são fáceis de enumerar. Temos como exemplo, a idade ou a altura de uma pessoa.

Pré-Processamento

Geralmente, antes de ser aplicada qualquer tipo de técnica de *Data Mining* são efetuados alguns pré-processamentos. Estes, podem ocorrer por diversos motivos, como por exemplo, quantidade de recursos reduzida face às necessidades enunciadas, dados apresentados deficitários ou inadequados ao caso de estudo em questão, entre outros. Contudo, são fundamentais para uma correta e eficiente análise, pois fazem com que seja reduzida a complexidade a nível das dimensões a ser analisadas posteriormente. (Aalst, W. M. P. van der., 2011) (D.Hand, H. Mannila and P.Smyth, 2001) (E. Alpaydin, 2010)

Um pré-processamento pode ser, a remoção de uma coluna, que não apresenta relevância para a análise, a remoção de uma instância que se encontra repetida, a remoção de uma instância que se encontra corrompida, isto é, alguma coluna podia estar preenchida de forma incorreta, entre outros tipos de pré-processamentos que podem ocorrer. (Aalst, W. M. P. van der., 2011) (D.Hand, H. Mannila and P.Smyth, 2001) (E. Alpaydin, 2010)

Dados Output

Depois de efetuada a análise será devolvido ao utilizador o resultado final, sendo que esse resultado será apresentado na forma de regras de associação, conjuntos de dados semelhantes, estruturas em árvore, gráficos, equações e padrões. (Aalst, W. M. P. van der., 2011) (E. Alpaydin, 2010)

Técnicas de *Data Mining*

As técnicas de *Data Mining* podem ser divididas em dois grandes grupos, o grupo das técnicas de aprendizagem supervisionada e o grupo das técnicas de aprendizagem não supervisionada. (D.Hand, H. Mannila and P.Smyth, 2001)

Aprendizagem Supervisionada

No grupo das técnicas de aprendizagem supervisionada, temos aquelas que são desenvolvidas com base em exemplos, isto é, é efetuada uma análise a um conjunto de dados de teste, ou treino e de seguida são refletidos modelos ou funções que foram encontrados nesse conjunto de dados. (D.Hand, H. Mannila and P.Smyth, 2001) (Aalst, W. M. P. van der., 2011)

Este tipo de análise pode incorrer em dois tipos de erros, *underfitting* e *overfitting*. No caso do *underfitting* o grupo de teste, ou treino pode ser demasiado genérico e desta forma, produzir modelos que não sejam capazes de satisfazer as necessidades pretendidas, bem como levar a uma elevada complexidade no desenvolvimento do procedimento que dá origem a estes modelos. Em relação ao *overfitting*, este pode acontecer quando o grupo de teste, ou treino dá origem a um modelo que está demasiado bem preparado para o *data set* providenciado, o que faz com que possam ocorrer algum tipo de erros no caso de existirem algumas necessidades que não se encontravam espelhadas no *data set* apresentado. (Aalst, W. M. P. van der., 2011) (D.Hand, H. Mannila and P.Smyth, 2001)

Dentro do grupo das técnicas de aprendizagem, estas ainda podem ser divididas em dois tipos de técnicas distintas, a classificação e a regressão. A diferença que reside na escolha de uma ou outra técnica prende-se com o tipo da variável de análise, isto é, se é uma variável de análise numérica, ou uma variável de análise categórica. (Aalst, W. M. P. van der., 2011) (E. Alpaydin, 2010) (I.H. Witten and E. Frank, 2005)

A classificação utiliza variáveis categóricas de análise e tem como principal objetivo classificar uma instância com base na variável em estudo. (Aalst, W. M. P.

van der., 2011) (D.Hand, H. Mannila and P.Smyth, 2001) Por exemplo, se tivermos o *data set* apresentado na tabela 2 apresentada em baixo, e a variável em estudo fosse o “Resultado Final”, era necessário saber qual a principal diferença entre o resultado final “Passou” e o “Reprovou”, podíamos então observar que os alunos que tinham obtido o resultado final “Passou”, tinham como característica o valor de todos os restantes atributos maiores que dez.

Matemática	Português	Inglês	Resultado Final
10	12	10	Passou
9	11	14	Reprovado
8	10	13	Reprovado
10	14	15	Passou

Tabela 2 - *Data Set* acerca da análise de uma pauta de estudantes de uma determinada escola, onde para os alunos terem um resultado final de "passou" tinham que ter as três disciplinas positivas, caso contrário teriam "chumbou". Dados apresentados meramente exemplificativos.

A regressão utiliza variáveis numéricas e tem como principal objetivo encontrar uma função que forneça uma estimativa o mais próxima possível do desejável, isto é, como o menor erro possível. (D.Hand, H. Mannila and P.Smyth, 2001)

As principais técnicas de regressão utilizadas são a regressão linear, a regressão geral de modelos e as redes neuronais. (D.Hand, H. Mannila and P.Smyth, 2001)

Aprendizagem Não Supervisionada

A aprendizagem não supervisionada não necessita de algo ou alguém que lhe providencie as respostas, ou seja, neste tipo de abordagens a rede terá que ser capaz de se auto ensinar tendo em conta as relações, os padrões e as regras que consegue extrair do conjunto de dados que lhe foi fornecido. (Aalst, W. M. P. van der., 2011) (E. Alpaydin, 2010) (M. Bramer, 2007)

Em contra ciclo com o que acontece com a aprendizagem supervisionada, a não supervisionada não apresenta única e exclusivamente dois tipos variáveis, ou seja, os dados não são divididos entre variáveis de resposta ou variáveis previsíveis. (Aalst, W. M. P. van der., 2011) (E. Alpaydin, 2010) (M. Bramer, 2007)

Este tipo de aprendizagem pode ser apresentada em dos tipos distintos, o *clustering* e o reconhecimento de padrões. (Aalst, W. M. P. van der., 2011)

No que diz respeito ao *clustering*, esta não apresenta uma definição consensual na comunidade científica, daí existirem diversos algoritmos com este denominador comum. (Estivill-Castro, Vladimir, 2002) Contudo o objetivo de todos estes algoritmos é encontrar, no conjunto de dados fornecidos, grupos de instâncias que sejam similares. (Aalst, W. M. P. van der., 2011) (E. Alpaydin, 2010) (I.H. Witten and E. Frank, 2005)

O *clustering* e a classificação, aparentam ter as mesmas características no que diz respeito à sua conceção, mas o primeiro não se foca numa determinada variável de resposta, que é o que acontece com a classificação, mas sim numa instância no seu todo. (Aalst, W. M. P. van der., 2011) (I.H. Witten and E. Frank, 2005) (E. Alpaydin, 2010)

Os dois algoritmos de *clustering* mais utilizados são o *k-means clustering* e o *agglomerative hierarchical clustering*. (Aalst, W. M. P. van der., 2011)

O *clustering* não está diretamente ligado ao *process discovery*, contudo pode ser uma ferramenta bastante útil para o pré-processamento, que é um dos passos do *process mining*, fazendo com que os casos similares fiquem agrupados e dessa forma facilitar o desenvolvimento de modelos por grupo, criando assim modelos parciais. (G. Greco, A. Guzzo, L. Pontieri, and D. Saccà, 2006) (A.K.A de Medeiros, A. Guzzo, G. Greco, W.M.P. van der Aalst, A.J.M.M. Weijters, B. van Dongen, and D. Sacca, 2008) (R.P.J.C. Bose and W.M.P. van der Aalst, 2009)

O reconhecimento de padrões apresenta como principal objetivo a procura de regras lógicas existentes no conjunto de dados fornecido, isto é, se “A igual B” e “B igual a C” então “A igual a C”. (Aalst, W. M. P. van der., 2011) (E. Alpaydin, 2010)

O algoritmo de reconhecimento de padrões mais utilizado é o *association rule mining*. (Aalst, W. M. P. van der., 2011)

Aprendizagem por Árvores de Decisão

A aprendizagem por árvores de decisão é uma técnica de aprendizagem supervisionada que é bastante útil para classificar instâncias com base em variáveis previsíveis. Esta é desenvolvida tendo por base uma variável categórica de resposta, isto é, permite que o resultado da árvore de decisão seja a classificação final de uma determinada instância. (E. Alpaydin, 2010) (M. Bramer, 2007) (I.H. Witten and E. Frank, 2005) (Aalst, W. M. P. van der., 2011)

Uma árvore de decisão é composta por (Aalst, W. M. P. van der., 2011) :

- Nós folha - São aqueles que correspondem aos valores possíveis das variáveis de resposta, sendo estes os nós que não têm mais nenhuma ramificação posterior;
- Nós - São aquelas que apresentam uma ou mais ramificações, são os nós que têm as variáveis previsíveis, sendo estes nós relativos aos atributos de cada instância;
- Nó inicial – É o nó que não tem nenhuma ramificação precedente; e
- Arcos – São a ligação entre os diferentes nós e apresentam o resultado de cada atributo.

Este tipo de abordagem é bastante genérica e apresenta algumas particularidades como por exemplo, o fato de ser complexo decidir quando parar de adicionar nós à árvore, restrições ao nível do limite de profundidade e de largura possíveis de serem explorados e elevada complexidade de escolha dos atributos a serem analisados. Para fazer face a estes desafios existem funções e algoritmos que permitem definir o limite de profundidade e/ou de largura a serem explorados bem como os atributos que devem de ser utilizados. (E. Alpaydin, 2010) (M. Bramer, 2007)

O fato de ser uma abordagem com elevado grau de especificidade faz com que para cada caso seja necessário criar uma árvore de decisão concreta, de modo a ser possível obter um modelo capaz de responder às questões que se pretende. (Aalst, W. M. P. van der., 2011) (I.H. Witten and E. Frank, 2005)

As árvores de decisão não estão diretamente ligadas com o *process discovery*, contudo podem ser aliadas a técnicas de *data mining* e desta forma possibilitarem a análise dos pontos de decisão previamente desenvolvidos com base em algoritmos de *process mining*, sendo a variável de resposta o caminho desde o nó inicial até ao nó folha cuja profundidade é máxima e sendo os atributos os nós não folha. (Aalst, W. M. P. van der., 2011)

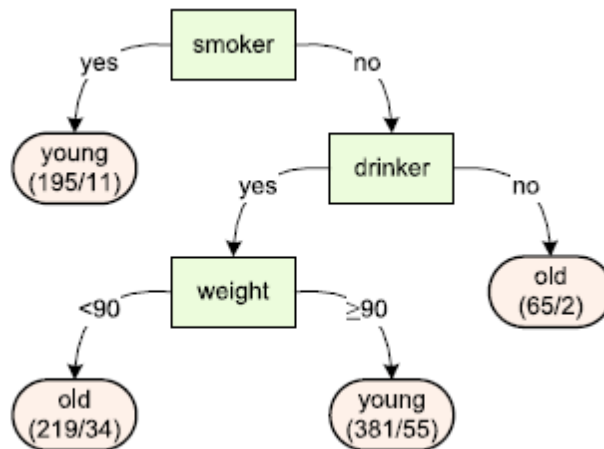


Figura 1 - Exemplo de uma árvore de decisão (Aalst, W. M. P. van der., 2011)

K-Means Clustering

O *k-means clustering*, como referido nos capítulos supracitados, é um algoritmo utilizado para fazer *clustering* de instâncias em grupos com características similares. (E. Alpaydin, 2010) (I.H. Witten and E. Frank, 2005)

Na figura 2, apresentada em baixo, podemos observar um exemplo da aplicação do algoritmo *k-means*, sendo que neste exemplo o conjunto de dados apresentado tem apenas duas características, o peso e a altura. Sendo cada um dos pontos os valores relativos à idade e à altura de uma determinada pessoa, ou seja, cada um dos pontos é uma instância. (Aalst, W. M. P. van der., 2011)

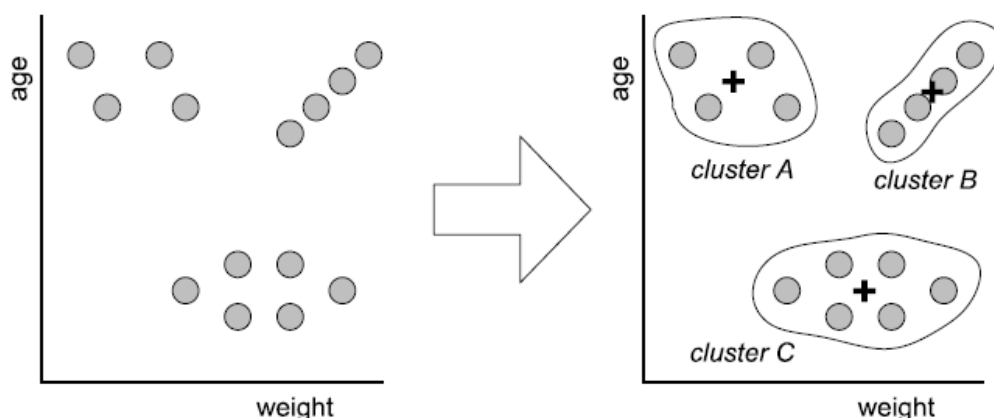


Figura 2 - Exemplo de *Clustering* com duas variáveis em estudo, peso e altura (Aalst, W. M. P. van der., 2011)

Como podemos observar na figura 2, existe uma grande diversidade de pontos ao longo do gráfico, mas é possível agrupar estes pontos, pois apesar de não serem todos iguais, apresentam características similares, na gráfica isto pode ser percebido tendo em conta as coordenadas de cada um dos pontos. (Aalst, W. M. P. van der., 2011)

Para que este agrupamento seja possível torna-se necessário primeiramente definir os centros através dos quais serão analisadas as distâncias - representado na figura 2 pelo símbolo "+" - sendo estes definidos com base em heurísticas preexistentes. Posteriormente a este processo irão ser calculadas as distâncias existentes entre cada uma das instâncias. Para este método pode ser utilizado, por exemplo, o método das distâncias euclidianas. Por fim serão agrupadas as instâncias que apresentarem uma menor distância entre si e o ponto central. (Aalst, W. M. P. van der., 2011) (M. Bramer, 2007) (I.H. Witten and E. Frank, 2005)

Sendo que este algoritmo pode apresentar diversas variâncias dependendo da forma de implementação e do caso de estudo apresentado. (Aalst, W. M. P. van der., 2011) (M. Bramer, 2007) (I.H. Witten and E. Frank, 2005) (E. Alpaydin, 2010)

Agglomerative Hierarchical Clustering

Como o *k-means* o *agglomerative hierarchical clustering* também apresenta como resultado final do seu processamento um agrupamento das instâncias fornecidas no *data set*, contudo a forma como se chega a este resultado é diferente da que é utilizada no *k-means*. (E. Alpaydin, 2010) (M. Bramer, 2007)

Este algoritmo de *clustering* pode ser dividido em duas abordagens distintas (I.H. Witten and E. Frank, 2005) (D.Hand, H. Mannila and P.Smyth, 2001) (E. Alpaydin, 2010):

- Abordagem *top down* – Esta, leva a que o primeiro estágio do agrupamento sejam a totalidade das instâncias existentes no *data set*, podemos ver este exemplo na figura 3, apresentada em baixo, depois serão efetuadas divisões em agrupamentos mais reduzidos até chegar ao ponto em que existe um grupo para cada instância; e
- Abordagem *bottom up* – Esta, funciona de forma oposta à abordagem *top down*, ou seja, o ponto de partida será um grupo por cada instância existente, sendo depois agrupados em grupos de diversas instâncias até que todas as estejam agregadas no mesmo grupo, podemos observar um exemplo desta abordagem na figura 4.

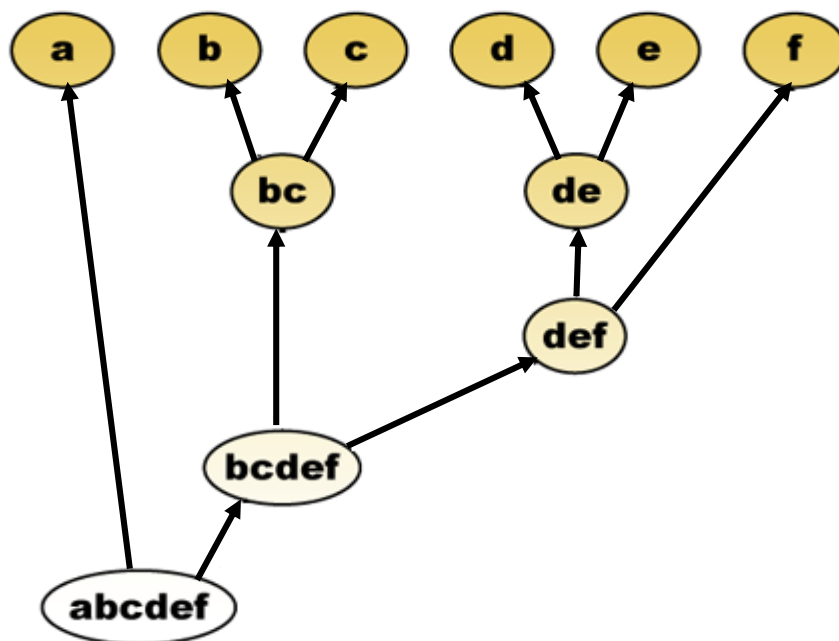


Figura 3 - Exemplo de uma abordagem *top down* do algoritmo *agglomerative hierarchical clustering*. Adaptado de (Yong Joseph Bakos, 2014)

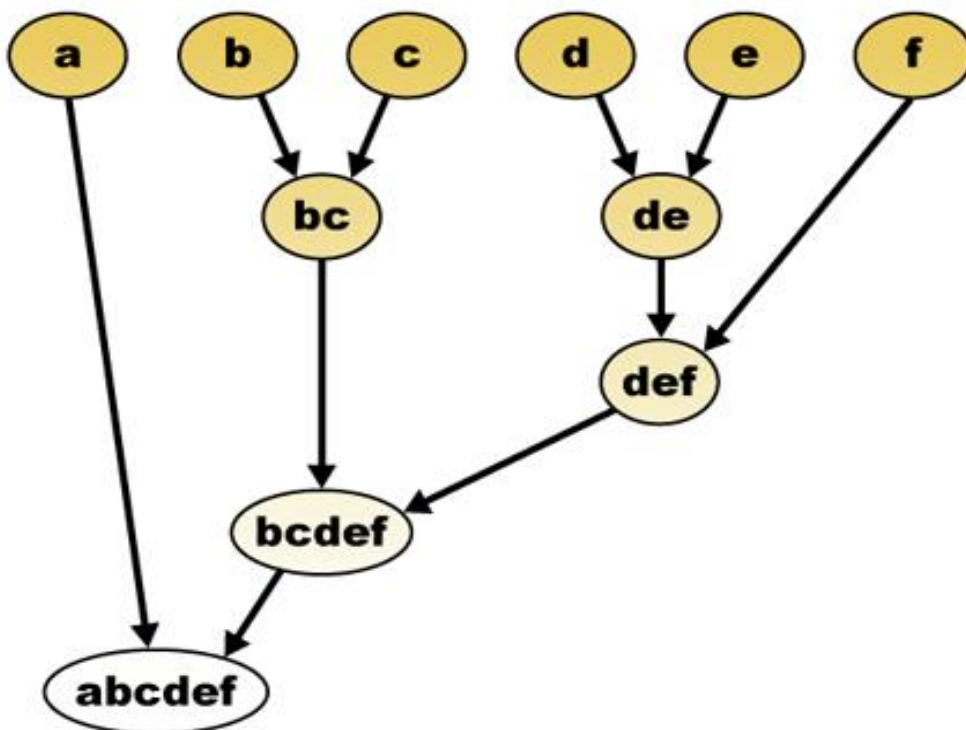


Figura 4 - Exemplo de uma abordagem *bottom up* do algoritmo *agglomerative hierarchical clustering*. Adaptado de (Yong Joseph Bakos, 2014)

Association Rule Learning

Como referido anteriormente, as árvores de decisão podem ser utilizadas para prever o valor de uma variável de resposta. O algoritmo de *association rule learning* apresenta o mesmo objetivo, ou seja, encontrar determinadas regras ou padrões no *data set* providenciado, sendo que a grande diferença destes algoritmos reside no fato de este último não se basear numa única variável de resposta, mas sim em todo o conjunto de dados apresentados. (Aalst, W. M. P. van der., 2011) (E. Alpaydin, 2010) (D.Hand, H. Mannila and P.Smyth, 2001)

O principal objetivo do *association rule learning* é encontrar regras na forma “*IF A THEN B*”, sendo o “A” o antecedente e o “B” o conseqüente, contudo estas também podem ser representadas na forma de “ $A \Rightarrow B$ ”. Por exemplo podemos ter regras como, “*IF 1<2 AND 2<3 THEN 1<3*”. Sendo que geralmente estas regras são constituídas por variáveis categóricas. (Aalst, W. M. P. van der., 2011) (I.H. Witten and E. Frank, 2005)

A validação das regras que são geradas por este tipo de algoritmos tendem a apresentar uma elevada complexidade, pois se o caso de estudo apresentar um *data set* com poucas instâncias e poucos atributos, é possível validar os resultados de forma manual, mas no caso de existirem conjuntos de dados com um número de instâncias bastante assinaláveis torna-se necessário utilizar métricas para efetuar esta validação. (I.H. Witten and E. Frank, 2005) (M. Bramer, 2007)

No que diz respeito ao algoritmo de *association rule learning* existem três métricas que são frequentemente utilizadas. Estas são o *support*, a *confidence* e a *lift*. (Aalst, W. M. P. van der., 2011) (E. Alpaydin, 2010) (D.Hand, H. Mannila and P.Smyth, 2001)

De seguida serão apresentadas as métricas tendo em conta as suas características e as suas formas de cálculo. Sendo que para estas fórmulas de cálculo irão ser utilizadas as seguintes variáveis: (Aalst, W. M. P. van der., 2011)

- “ N_x ” - Número de instâncias que estão relacionadas com x;
- “ N_y ” - Número de instâncias que estão relacionadas com y;
- “ $N_{x \cap y}$ ” – Número de instâncias que estão relacionadas tanto com x como com y;
- “N” – Número total de instâncias.

O *support* indica a aplicabilidade das regras geradas, isto é, o número de instâncias por cada uma variável que correspondem ao conjunto do antecedente - conseqüente. Sendo que uma regra com um valor de *support* elevado é mais capaz de fazer face às necessidades. (Aalst, W. M. P. van der., 2011)

$$support(X \Rightarrow Y) = N_{X \wedge Y} / N$$

Figura 5 - Fórmula de cálculo da métrica *support*. (Aalst, W. M. P. van der., 2011)

A *confidence* indica a viabilidade da regra, isto é, a capacidade que a regra tem de garantir que quando o antecedente existe também irá existir o consequente. Quando existe um alto valor de *confidence* a regra é mais viável do que nos casos em que o valor é mais baixo. (Aalst, W. M. P. van der., 2011)

$$confidence(X \Rightarrow Y) = N_{X \wedge Y} / N_X$$

Figura 6 - Fórmula de cálculo da métrica *confidence*. (Aalst, W. M. P. van der., 2011)

A *lift* avalia a correlação existente entre o antecedente e o consequente, isto é, se por exemplo o valor for superior a 1, podemos afirmar que o antecedente e o consequente estão relacionados positivamente, isto é, no caso de valor ser 3, por exemplo, temos que o antecedente e o consequente aparecem mais três vezes juntos do que quando são independentes. No caso de o valor ser inferior a 1, denota que o antecedente e o consequente estão relacionados negativamente, o que faz com que sejam mais interessantes as regras com altos valores de *lift* do que quando são mais reduzidos. Contudo, este valor só é utilizado no caso de já serem conhecidos os valores de *support* e de *confidence*. (Aalst, W. M. P. van der., 2011)

$$lift(X \Rightarrow Y) = \frac{N_{X \wedge Y} / N}{(N_X / N) (N_Y / N)} = \frac{N_{X \wedge Y} N}{N_X N_Y}$$

Figura 7 - Fórmula de cálculo da métrica *lift*. (Aalst, W. M. P. van der., 2011)

Este tipo de algoritmos apresentam como principais entraves à sua concretização, o número de regras geradas que não apresentam qualquer interesse para o caso de estudo em questão e o tempo de processamento para serem geradas as regras de um determinado caso, este fato pode acontecer quando existe um elevado número de instâncias e atributos. (E. Alpaydin, 2010) (M. Bramer, 2007) (D.Hand, H. Mannila and P.Smyth, 2001) (I.H. Witten and E. Frank, 2005) (Aalst, W. M. P. van der., 2011)

As técnicas de *association rule learning* estão diretamente relacionadas com o *process discovery*, pois também este percorre o *event log* com o objetivo de encontrar regras ou padrões, contudo a primeira não toma em consideração a ordem das

atividades e não apresenta a capacidade de criar um modelo de processo genérico. (Aalst, W. M. P. van der., 2011)

Sequence Mining

O principal objetivo do *sequence mining* é vislumbrar sequências frequentes de padrões ou regras. Por exemplo, no caso de termos um conjunto de dados relativo aos pedidos que são feitos por um cliente durante um determinado período temporal num café, podemos encontrar pedidos iguais, de dia para dia, ou pedidos que sejam consequentes, isto é, às segundas-feiras o cliente pede um café, mas às terças já pede um sumo e um bolo, e através destas regras ou padrões é possível criar um modelo sobre os pedidos do cliente. (Aalst, W. M. P. van der., 2011) (R. Srikant and R. Agrawal, 1996)

Para que seja possível a aplicação deste algoritmo, é necessário que as instâncias do conjunto de dados apresentado tenham algumas características específicas. Estas prendem-se com a necessidade de todas terem atributos como o número de instância, que deve ser único, o tempo em que aconteceu a instância, ou seja, o *timestamp*, e deveram de estar ordenadas por número de sequência. (Aalst, W. M. P. van der., 2011)

Uma sequência é frequente se o padrão está contido num subconjunto do conjunto de dados predefinidos para o caso de estudo. (Aalst, W. M. P. van der., 2011)

Podem existir um número de regras ou padrões infinitos, contudo se uma sequência é frequente, então as suas subsequências também o serão, e desta forma poderão ser criadas regras na forma de $A \Rightarrow B$, onde “A” é o padrão e “B” é uma extensão do padrão. Regras estas que depois poderão ajudar a analisar os resultados e a desenvolver modelos, pois, se por exemplo, o support desta regra apresentar um valor elevado podemos afirmar que este padrão também tende a acontecer em maior número. (R. Srikant and R. Agrawal, 1996) (Aalst, W. M. P. van der., 2011)

Episode Mining

Em conformidade com o *sequence mining*, também o *episode mining* é uma abordagem *Apriori*, que tem como principal objetivo descobrir episódios frequentes. (H. Mannila, H. Toivonen, and A.I. Verkamo, 1997)

Um episódio é uma ordem parcial de um conjunto de acontecimentos, ou seja, uma ordem através da qual foram executadas as atividades expressas no *event log*. (R. Srikant and R. Agrawal, 1996) (Aalst, W. M. P. van der., 2011)

Para que um conjunto de atividades possa ser considerado episódio é necessário que este ocorra numa determinada janela temporal e que cada um dos nós possa conter eventos ordenados tendo em conta a realização das atividades.

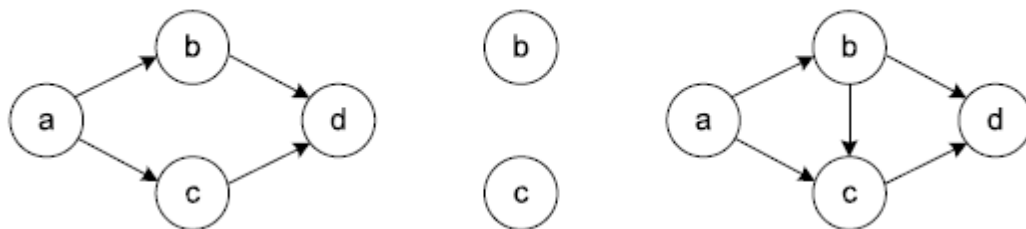


Figura 8 - Exemplo de três episódios distintos. (Aalst, W. M. P. van der., 2011)

Na figura 8, podemos observar três exemplos de episódios, onde por exemplo no primeiro grafo, é possível vislumbrar que existem quatro atividades que ocorrem naquela determinada ordem. Mas para um determinado conjunto de atividades ser um considerado um episódio não é obrigatório que as atividades sejam sequenciadas, pois como está demonstrado no segundo exemplo da figura apresentada em cima, existem duas atividades que ocorrem em simultâneo, mas são consideradas um episódio pois existe uma ordem para que estas aconteçam em paralelo. (Aalst, W. M. P. van der., 2011)

Como no *sequence mining*, no *episode mining* também poderão ser geradas um conjunto infinito de regras ou padrões, contudo podemos utilizar a mesma metodologia utilizada no primeiro, ou seja, se é possível afirmar que se um episódio é frequente, então os seus sub episódios também o serão, e desta forma poderemos gerar regras na forma " $A \Rightarrow B$ ", onde " A " é um sub episódios de " B ". (Aalst, W. M. P. van der., 2011)

Tanto o *sequence mining*, como o *episode mining* podem ser considerados variações do *association rule learning*, pois a grande diferença entre estes reside no fato de os primeiros tomarem em consideração o tempo e a ordenação dos eventos, contudo ambos estão relacionados diretamente com o *process discovery*. (Aalst, W. M. P. van der., 2011)

Apesar de existirem muitas semelhanças entre os algoritmos de *process mining* e o *episode mining* e/ou *sequence mining*, também existem algumas diferenças. Estas diferenças residem no fato de só os padrões são considerados, ou seja, não existe o desenvolvimento de modelos, o foco dos algoritmos é o comportamento das instâncias

e não o desenvolvimento de modelos, os modelos não conseguem modelar escolhas nem ciclos e por fim, tanto o *episode mining* como o *sequence mining* não conseguem resolver de forma apropriada o problema da concorrência. (Aalst, W. M. P. van der., 2011)

Qualidade dos Modelos Resultantes

As técnicas apresentadas nos capítulos anteriores não podem ser utilizadas para as tarefas mais importantes do *process mining*, sendo estas tarefas o *process discovery*, o *conformance checking* e o *process enhancement*. Apesar disto, não é apenas no *data mining* que existe uma complexidade elevada para efetuar a validação dos resultados, no *process mining* a validação dos modelos resultantes também é uma tarefa bastante complexa. Contudo a experiência já existente no *data mining* é de extrema utilidade para avaliar os modelos gerados pelas técnicas de *process mining*. (Aalst, W. M. P. van der., 2011)

Medição da Performance do Classificador

Como foi apresentado no capítulo das árvores de decisão, existem algumas decisões bastante complexas a tomar quando se está a desenvolver uma árvore de decisão para aprendizagem supervisionada, sendo estas a seleção dos atributos, o momento onde deve ser terminada a exploração e os valores onde devem de ser efetuada as ramificações. (D.Hand, H. Mannila and P.Smyth, 2001) (E. Alpaydin, 2010) (M. Bramer, 2007) (Aalst, W. M. P. van der., 2011)

Contudo a grande dificuldade prende-se com o fato de saber como é que pode ser avaliada a performance de uma árvore de decisão. Este fato torna-se ainda mais relevante devido à existência de um número considerável de abordagens e à elevada dificuldade de analisar as instâncias que não são discerníveis ao olho humano no *data set*. (Aalst, W. M. P. van der., 2011) (E. Alpaydin, 2010)

Para efetuar uma análise ao conjunto de dados existente no *data set* podemos utilizar uma matriz de confusão, sendo esta uma tabela utilizada para avaliar a performance dos algoritmos de aprendizagem supervisionada. (M. Bramer, 2007) (I.H. Witten and E. Frank, 2005) (D.Hand, H. Mannila and P.Smyth, 2001)

Existem um sem número de medidas de performance com base na matriz de confusão, para apresentar algumas dessas medidas, será utilizada a tabela apresentada na figura abaixo como exemplo. (D.Hand, H. Mannila and P.Smyth, 2001) (I.H. Witten and E. Frank, 2005) (Aalst, W. M. P. van der., 2011)

		<i>predicted class</i>		
		+	-	
<i>actual class</i>	+	<i>tp</i>	<i>fn</i>	<i>p</i>
	-	<i>fp</i>	<i>tn</i>	<i>n</i>
		<i>p'</i>	<i>n'</i>	<i>N</i>

(a)

Figura 9 - Matriz de confusão para duas classes. (Aalst, W. M. P. van der., 2011)

Na figura 9, temos uma matriz de confusão que espelha duas classes distintas, a classe “+” (positiva) e a classe “-“ (negativa). Podemos então observar os quatro resultados possíveis numa classificação, sendo estes, o “tp” que são os verdadeiros positivos, isto é, as instâncias que foram corretamente classificadas como positivas, o “fn” que são os falsos negativos, isto é, as instâncias que foram classificadas como negativas mas deviam ter sido classificadas como positivas, o “fp” que são os falsos positivos, isto é, as instâncias que foram classificadas como positivas mas deveriam ter sido classificadas como negativas, e por fim o “tn” que são os verdadeiros negativos, isto é, as instâncias que foram corretamente classificadas como negativas. (Aalst, W. M. P. van der., 2011)

Podemos ainda observar na figura 9 alguns somatórios, como são o “p”, que é a soma do número de instâncias são positivas, o “n” que é a soma do número de instâncias negativas, o “p'”, que é o número de instâncias que foram classificadas como positivas, o “n'” que é o número de instâncias que foram classificadas como negativas, e por fim o “N” que é a soma total do número de instâncias presentes num *data set*. (Aalst, W. M. P. van der., 2011)

Através destes somatórios será possível calcular as métricas que são apresentadas na figura 10, sendo estas por sua vez calculadas tendo como base as fórmulas apresentadas na figura supracitada. Nestas temos, o “erro” que é o número de instâncias que foram classificadas de forma incorreta, a “exatidão” que é o número de instâncias que foram bem classificadas, o “rácio de positivos” que é o número de instâncias positivas que foram classificadas como positivas, o “rácio de negativos” que é o número de instâncias negativas que foram classificadas como positivas, a “precisão” que é o número de instâncias apresentadas que são relevantes, e por fim

a “rechamada” que é o número de instâncias relevantes que são recuperadas. (Aalst, W. M. P. van der., 2011) (D.Hand, H. Mannila and P.Smyth, 2001) (I.H. Witten and E. Frank, 2005)

<i>name</i>	<i>formula</i>
error	$(fp+fn)/N$
accuracy	$(tp+tn)/N$
tp-rate	tp/p
fp-rate	fp/n
precision	tp/p'
recall	tp/p

Figura 10 - Métricas para medir a performance de um classificador. (Aalst, W. M. P. van der., 2011)

Validação Cruzada

As diferentes métricas existentes para avaliar a performance que utilizam a matriz de confusão, apresentada no capítulo anterior, utilizam o mesmo *data set* para validar os dados que foi utilizado para ensinar a árvore de decisão, isto é, unicamente podemos analisar as instâncias que são discerníveis pelo olho humano, podendo assim originar um problema de *overfitting*. (Aalst, W. M. P. van der., 2011)

Este fato pode ser vislumbrado por exemplo, no caso de haver um *data set* com os números de aluno, neste caso podemos definir que cada nó folha será responsável por cada um dos números de aluno, desta forma é relativamente fácil fazer a validação, mas não será possível validar a performance expectável de um novo número de aluno, pois não existe qualquer tipo de informação relativamente a esse novo aluno. (Aalst, W. M. P. van der., 2011)

Para fazer face a este desafio é necessário utilizar estimativas, por forma a poder prever com alguma precisão as instâncias que não perceptíveis ao olho humano, isto é, que não estejam contempladas no conjunto de dados fornecido. Com a finalidade de estimar estas instâncias, pode ser utilizada a validação cruzada, esta consiste na separação do *data set* inicial em conjunto de teste e conjunto de treino. Onde o conjunto de treino será utilizado para ensinar o modelo ao passo que o

conjunto de teste será utilizado para avaliar o modelo criado, tendo por base os exemplos não existentes no conjunto de treino. (Aalst, W. M. P. van der., 2011)

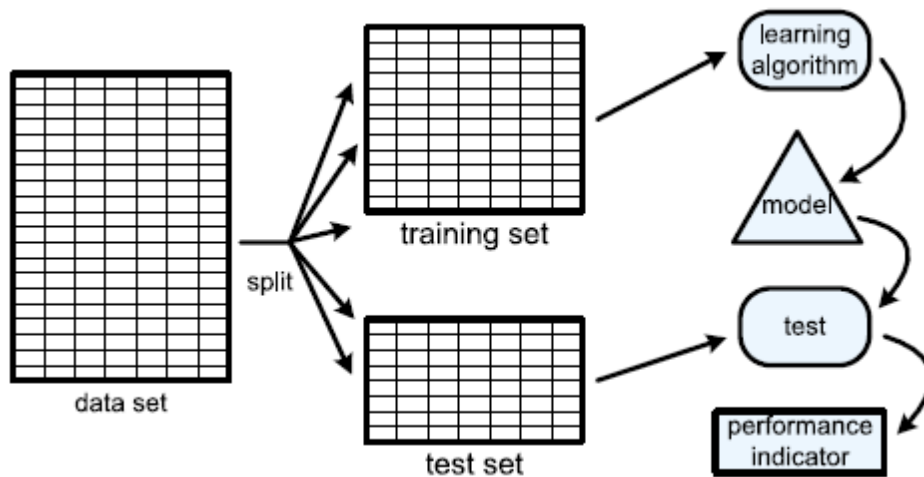


Figura 11 - Exemplo de funcionamento da validação cruzada. (Aalst, W. M. P. van der., 2011)

A validação cruzada não é restrita para a classificação, também pode ser utilizada para outras técnicas de *data mining*. A condição necessária para ser possível a sua aplicação é a existência de métricas capazes e validar a performance dos resultados. (Aalst, W. M. P. van der., 2011)

Desafios na Análise da Qualidade dos Resultados

A avaliação da qualidade dos resultados provenientes da aplicação de uma técnica de *data mining* está longe de ser acessível e consensual, este fato deve-se em grande parte à forma como o problema é posto, pois este só providencia uma solução tendo em conta os exemplos que lhe tenham sido dados. (Aalst, W. M. P. van der., 2011)

Para fazer face a este desafio existe a necessidade de ter em conta fatores externos, definir regras, de modo a limitar a exploração das árvores de decisão, e utilizar estratégias já implementadas por alguns algoritmos por forma a criar limitações ao número de soluções. Contudo, e apesar de estas técnicas servirem de guia para a procura de processos, também levaram a que sejam ignoradas algumas soluções que poderiam ser boas. (Aalst, W. M. P. van der., 2011)

Business Process Management

Evolução Histórica

Evolução das organizações

No início do século XX surgiram uma elevada quantidade de teorias organizacionais emergentes, levando assim ao aparecimento de um novo campo doutrinário, denominada ciência das organizações. Este, apresenta como principais objetivos o estudo da organização e suas funções implícitas (Henri Fayol, 1930), a organização científica do trabalho (Taylor, Frederick W., 1911) as tarefas dos gestores e do executivo (Chester I. Barnard, 1938), a sua estruturação formal-burocrática (Weber, M., 2007) e a abordagem da gestão e dos seus conceitos base. (Peter Drucker, 1942)

Em meados do século XX, como consequência do crescimento económico proveniente da segunda guerra mundial e da necessidade das populações afetadas pela mesma, especialmente a população europeia e a japonesa, quererem atingir o mais depressa possível um nível de vida de qualidade, surge a necessidade de desenvolver novas metodologias e abordagens organizacionais, tendo em conta a forma como estas se relacionam com o ambiente que as rodeia (Lorsch, J. and Lawrence, P., 1967) (Kast, E. & Rosenzweig, E. & Johnson, A., 1967), tendo em conta a tomada de decisão (March, J. & Simon, H., 1958), a necessidade de aumentar a humanização organizacional (McGregor, D., 1960), (Maslow, A., 1943) (Hezberg, F., 1959) e o aparecimento de novas técnicas como o *Marketing* e a qualidade.

No final do século XX, através das potencialidades cada vez maiores da informática e das crescentes inovações tecnológicas associadas às organizações temos assistido à criação de novas metodologias e abordagens nas organizações. Começou então a surgir a necessidade de desenvolver estruturas organizacionais para o futuro (Alvin Toffler, 1985), (Peter Drucker, 2001), racionalizar recursos e providenciar maior relevo na análise das contingências relativas ao contexto organizacional (Pfeffer, J. and G. R. Salancik, 1978) (Hannan, M.T. and J. Freeman, 1989) (Porter, Michael E., 1996) (Thomas J. Peters and Robert H. Waterman, 1982), desenvolver novas abordagens ao nível da gestão (Porter, Michael E., 1991) (Mintzberg, H., 1994) e analisar diferentes perspetivas organizacionais, nas quais se destacam as perspetivas cultural, sociológica e organizacional. (Mintzberg, H., 1994), (Galbraith, J., 1995), (Weick, K., 1979) (Handry, C., 1990)

O crescente aumento da competitividade das organizações potencializada pela globalização e pelo crescimento exponencial de novas tecnologias, levou as

organizações a focarem-se na relação com cliente (Stefanou, C. J., Sarmaniotis, C. & Stafyla, A., 2003), esta impulsionada pelo desenvolvimento do *Marketing* como campo doutrinário, desenvolvendo assim um novo tipo de organizações, denominadas organizações orgânicas, fazendo assim com as organizações se tornassem mais capazes de fazer face a mudanças inesperadas bem como uma maior facilidade e rapidez na tomada de decisão (O'Connor, M.K. & Netting, F. E., 2009), levando-nos assim ao princípio de que as organizações são organismos vivos e crescem tendo em conta as necessidades dos seus clientes. (Mole, J., 2003)

Nos tempos mais recentes as organizações têm vindo a sofrer alterações no que diz respeito à forma como são geridas, passando de uma perspectiva operacional para uma perspectiva centrada em processos (Munstermann, B., Eckhardt, A., & Weitzel, T., 2010), isto leva-nos à cada vez maior importância dos processos nas organizações.

Processos de Negócio

Como referido no capítulo anterior, as organizações atualmente utilizam uma gestão assente em processos de negócio (Munstermann, B., Eckhardt, A., & Weitzel, T., 2010), mas estes podem variar de organização para organização, pois cada organização é única e tem as suas especificidades.

Por um lado, um processo pode ser visto como a representação de uma operação de uma organização, isto é, o meio através do qual executa as tarefas para alcançar os seus objetivos. (Bizagi, 2014) Podemos definir então um processo de negócio como uma representação ou transformação, ou seja, um evento onde são recebidos *inputs* e são retornados *outputs*, sendo que nas organizações estes são apresentados na forma de serviços ou de produtos. (Bizagi, 2014), (Shewhart, A., 1931) Esta definição assenta numa ótica de processos de produção, ou seja, de algo específico para uma determinada atividade. (Shewhart, A., 1931)

Por outro lado, podemos analisar um processo como sendo um processo de controlo e não de produção, neste caso um processo é um conjunto de atividades realizadas através de uma sequência lógica com o objetivo de criar bens ou serviços com valor reconhecido pelos seus clientes. (Hammer, M. & Champy, J., 1994) (Graham, M. & Lebaron, M., 1994) (Gonçalves, J., 2000) (Henry J. Johansson et al. , 1993)

Os processos de negócio podem ser divididos em três tipos, que são (Gonçalves, J., 2000) (Harmon, P., 2007) (Davenport, T., 1993) (Hammer, M. & Champy, J., 1994):

- **Processos de Gestão:** São processos com foco principal na gestão e na relação existente entre a estrutura de gestão e a estrutura operacional, estes incluem entre outros métricas e análises de desempenho;

- Processos Operacionais: São os processos responsáveis pelo core do negócio, isto é, são focados fundamentalmente na organização e são responsáveis pela coordenação dos vários departamentos da organização;
- Processos de Suporte: São processos responsáveis por providenciar todas as informações e todos os recursos necessários para desenvolver os processos internos da organização.

Nos dias correntes já existem alguns processos de negócio implementados nas empresas como por exemplo, recrutamento e seleção, reclamações, planeamento estratégico, abordagens de vendas, processamento salarial, entre tantas outras tarefas executadas periodicamente. (Bizagi, 2014)

Na figura 1, temos um exemplo sobre a forma como os processos de negócio estão enraizados e como estão a ser aplicados nas organizações.

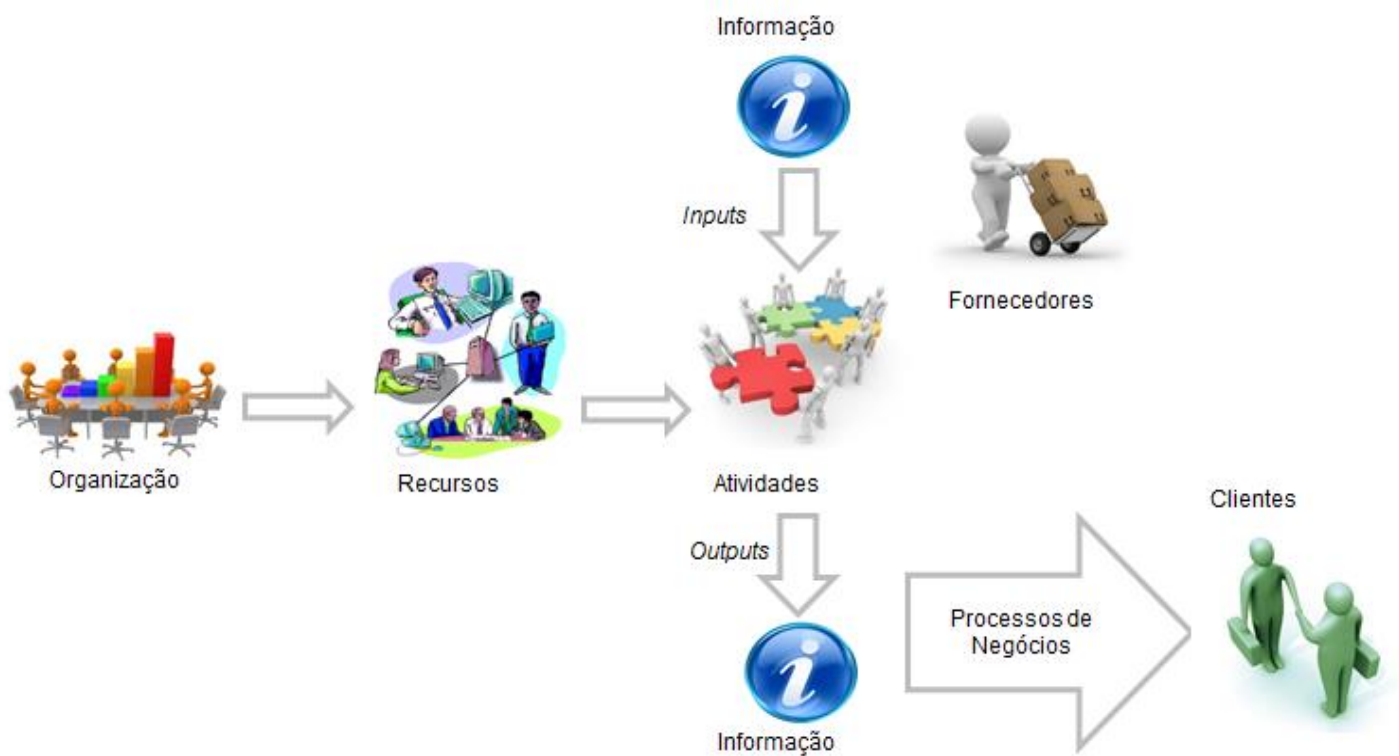


Figura 12 - Processo de Negócio

O que é?

O BPM é um conceito bastante complexo e para qual existe um conjunto de definições bastante vasto, temos por um lado a definição proposta por (WMP Van Der Aalst, 2004) e (M. Weske, 2007), que diz que o BPM é a combinação do conhecimento oriundo das tecnologias de informação com os da gestão e a aplicação do mesmo a processos de negócio operacionais, podemos ter a definição proveniente do (*Business Dictionary*), que é um dicionário de termos empresariais, que defende que o BPM é a atividade realizada por empresas ou organizações para identificar, avaliar e melhorar os seus processos de negócio, alicerçando a tecnologia, o BPM pode agora ser gerido de forma eficaz, através de *software* desenvolvido tendo em conta as métricas e as políticas determinadas pela empresa. Esse tipo de ação é essencial para as empresas que procuram melhor eficiência nos seus processos.

Evolução do BPM

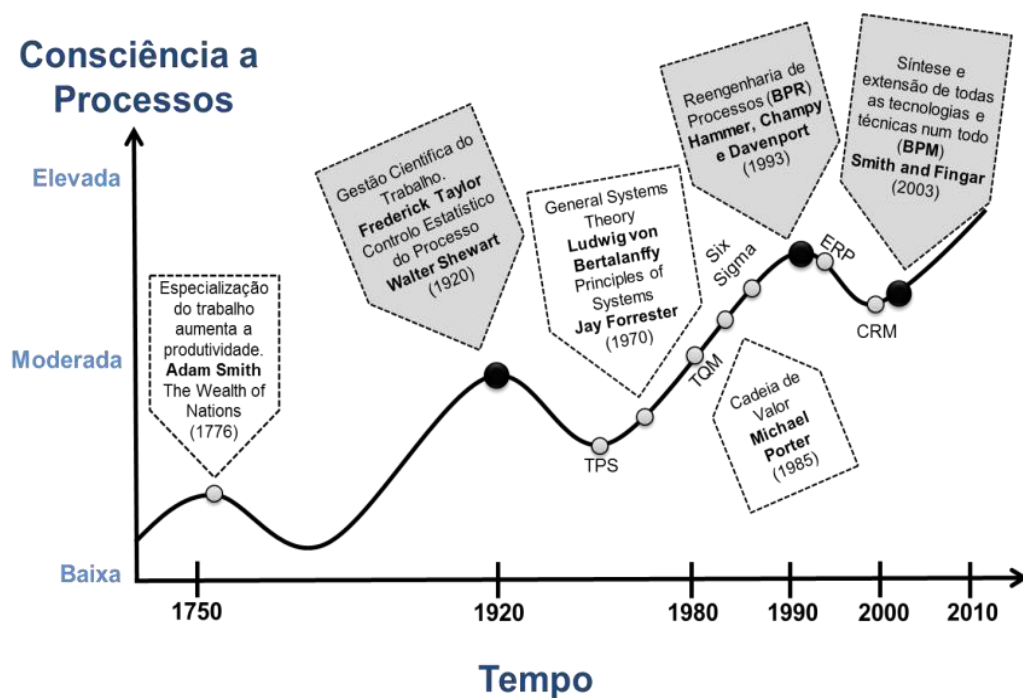


Figura 13 - Análise histórica da evolução dos sistemas de BPM (José António Sena Pereira, 2013)

Os sistemas informáticos, bem como a restante tecnologia sofreram uma elevada quantidade de inovações e transformações ao longo dos tempos, esta pode ser visualizada de forma gráfica na figura 13, onde é possível perceber a evolução destes sistemas desde os seus primórdios, ou seja, 1750 até o ano de 2010. (Aalst, W. M. P. van der., 2012) (Aalst, W. M. P. van der., 2012)

Estas evoluções tiveram diversas bases e diversos atores preponderantes, estes atores podem vir dos mais diversos quadrantes da sociedade e as suas obras não são tão recentes como a área de estudo do BPM, entre eles ressalva-se a influência de (Smith, A., 2006) que entre 1723 e 1790 identificou as vantagens da divisão do trabalho, (Taylor, Frederick W., 1911) que iniciou o estudo da gestão como uma ciência e (Ford, H., 1863 – 1947) que introduziu a produção em massa, através destas ideias/teorias podemos observar pontos comuns com os sistemas de BPM. (Aalst, W. M. P. van der., 2012)

Esta evolução teve início em meados do século XVIII com as ideias acerca da divisão do trabalho na indústria transformadora, que visavam potenciar as capacidades de cada um dos trabalhadores em particular com vista à redução do

tempo necessário a desenvolver a tarefa bem como a eficácia com que esta era desenvolvida. (Adam Smith, 1776)

Cerca de um século depois, (Taylor, Frederick W., 1911), considerado o fundador da administração científica, propôs a utilização de métodos científicos tendo como principal foco a eficiência operacional. (José António Sena Pereira, 2013)

Tendo em conta as sinergias do século XX, no que diz respeito à divisão do trabalho por Adam Smith, da especialização das tarefas por Frederick Taylor, existe uma grande melhoria no que diz respeito à produção industrial. Esta melhoria deveu-se principalmente a Henry Ford, que desenvolveu a primeira linha de montagem, por forma a aumentar a produção, e reduzir tempo e custo. (José António Sena Pereira, 2013) (Ford, H., 1863 – 1947)

Após o término da segunda guerra mundial, mais precisamente em 1950, a Toyota apresenta um novo sistema integrado, denominado TPS, *Toyota Production System*. Este, assentava em dois conceitos primordiais, o primeiro tinha como pressuposto a impossibilidade de utilizar uma determinada máquina, caso esta esta a produzir produtos defeituosos, o segundo assentava no conceito *Just-In-Time*, levando a que cada processo só produza o que é estritamente necessário para satisfazer as necessidades do processo seguinte. (José António Sena Pereira, 2013)

Depois do foco dado à qualidade, aos materiais e aos desperdícios de tempo, nos anos 70, apareceu um novo paradigma para o desenvolvimento de teorias, o *General System Theory*, este tinha como principal objetivo ser uma teoria para desenvolver novas teorias de forma dinâmica. (Ludwig von Bertalanffy, 1989) (José António Sena Pereira, 2013)

Em meados dos anos 80, aparece uma abordagem muito conhecida, e que ainda hoje é bastante utilizada, que é a cadeia de valor de Michael Porter, abordagem esta tinha como base a organização do trabalho a nível organizacional tendo em conta determinados objetivos previamente definidos. Sendo que esta cadeia de valor vai desde as funções mais primárias até às funções de topo. (Porter, Michael E., 1991) (José António Sena Pereira, 2013) (M. Weske, 2007)

Foi então durante os anos 80, que ocorreu uma mudança de paradigma, isto é, as organizações passaram a assentar a sua cadeia de valor em processos e não em funções, dando então foco ao TQM, ou seja, *Total Quality Management*, e ao Six Sigma. (Mintzberg, H., 1994) (José António Sena Pereira, 2013) (Hammer, M. & Champy, J., 1994) (Becker, J.; Kugeler, M.; Rosemann, M., 2003)

Nos anos 90, surge o BPR, *Business Process Reengineering*, tendo como principais bases a necessidade de repensar e redesenhar processos de forma abrupta

por forma a gerar melhorias ao nível da qualidade, custo, serviço e velocidade. (Davenport, T., 1993) (José António Sena Pereira, 2013)

A abordagem supracitada visava tornar as tecnologias de informação como um instrumento para coordenar e interligar tarefas e recursos e não apenas realizar tarefas autónomas e individuais. (José António Sena Pereira, 2013) (M. Dumas, Aalst, W. M. P. van der. and A.H.M. ter Hofstede, 2005)

Contudo, BPM e BPR, não são exatamente o mesmo conceito, pois o primeiro é mais prático, iterativo e incremental no ajuste dos processos de negócio, ao passo que o segundo existe uma alteração radical dos processos de negócio já criados. (José António Sena Pereira, 2013)

Ainda nos anos 90, surge o ERP, *Enterprise Resource Planning*, tornando-se no grande foco das organizações, pois era visto como a solução para todos os problemas, contudo, tal não era verdade, pois para além de existirem algumas debilidades no que diz respeito às questões relacionadas com os processos nas organizações, também não conseguiam torná-los mais eficazes e eficientes. (José António Sena Pereira, 2013) (Aalst, W. M. P. van der., 2012)

No início dos anos 2000, surgem os primeiros sistemas CRM, *Customer Relationship Management*, que mudam o foco para o cliente, fazendo assim com que a principal necessidade sejam os processos de *front-office* em vez dos processos de *back-office*. (José António Sena Pereira, 2013)

Hoje em dia o BPM assume-se como uma abordagem de processos que adquiriu sinergias de um passado mecanicista, capaz de desenvolver tarefas sequenciais e rotineiras com precisão. Todas estas sinergias possibilitaram que ao longo dos tempos fossem desenvolvidos métodos, técnicas e ferramentas para suportar o desenho, adoção, gestão e análise de processos de negócio, envolvendo pessoas, organizações, aplicações, documentos ou outras fontes de informação. (José António Sena Pereira, 2013) (M. Dumas, Aalst, W. M. P. van der. and A.H.M. ter Hofstede, 2005) (M. Weske, 2007) (Aalst, W. M. P. van der., 2012)

Atividades Chave do BPM

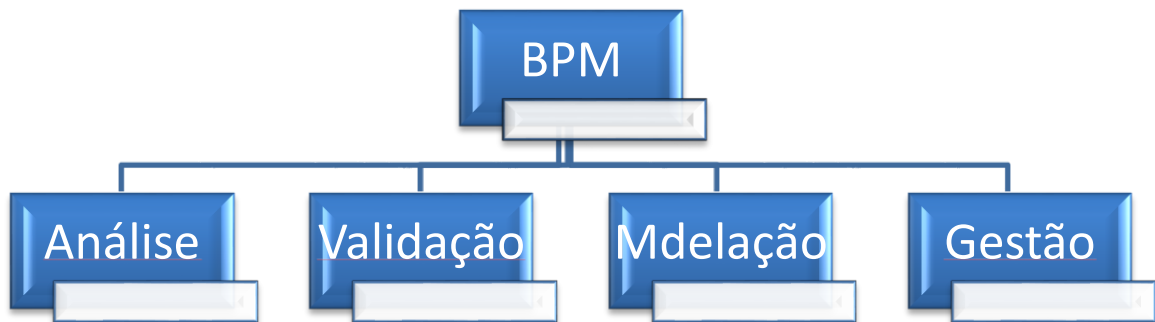


Figura 14 – Atividades chave do BPM, adaptado de (Aalst, W. M. P. van der., 2012)

Segundo (Aalst, W. M. P. van der., 2012) existem quatro atividades chave relacionadas com o BPM, como podemos observar na figura 14, estas são de extrema relevância pois são de uma forma genérica as tarefas que o sistema irá desenvolver.

Estas atividades são:

- Análise – Nesta atividade é desenvolvida uma análise a processos já existentes através de *process model's* ou de *event log's* com a finalidade de otimizar ou tornar mais eficientes e eficazes os mesmos;
- Validação – Nesta atividade são validados os processos já existentes e já modelados através da modelação previamente feita, ou seja, irá ser utilizada a modelação construída *A priori* para efetuar uma análise de modo a validar se o processo implementado está de acordo com o que foi previamente definido;
- Modelação – Nesta atividade será modelado o processo desejado, sendo esta atividade responsável pela análise dos requisitos de negócio e consequente adaptação do mesmo à realidade técnica;
- Gestão – Nesta atividade será efetuada a gestão e o controlo dos processos já em curso, isto é, possíveis ajustes ao processo ou realocação de recursos.

Representações gráficas BPM

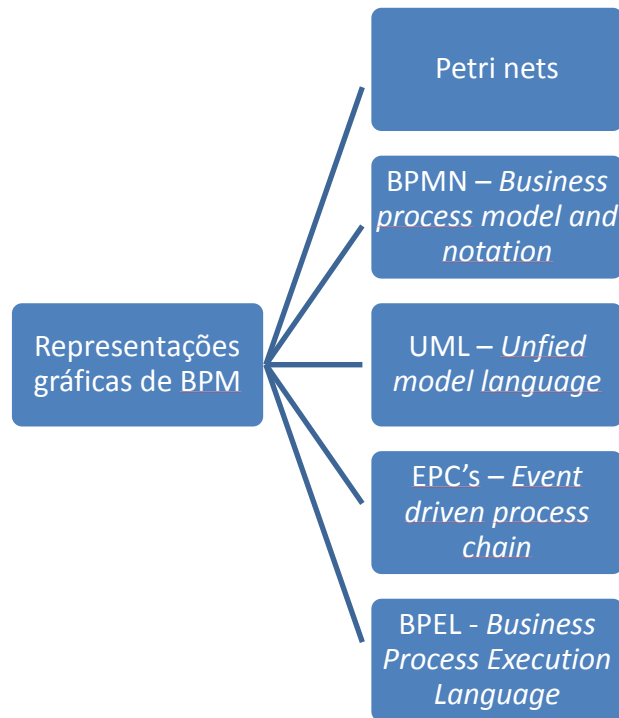


Figura 15 - Representações gráficas do BPM, adaptado de (Aalst, W. M. P. van der., 2012) (W. M. P. van der Aalst, A. H. M. terHofstede, B. Kiepuszewski & A. P. Barros, 2003)

O BPM pode ser representado de diferentes formas e através de diversas notações, na figura 15 é possível observar algumas dessas notações existentes, mas apesar de serem distintas na forma, apresentam como ponto de intersecção, o fato de todas apresentarem um processo de negócio por meio de atividades. (Aalst, W. M. P. van der., 2012) (W. M. P. van der Aalst, A. H. M. terHofstede, B. Kiepuszewski & A. P. Barros, 2003)

Apesar de todas as notações poderem ser utilizadas para representar um processo de negócio existem algumas que são mais apropriadas que outras para determinados casos reais, como é referido no estudo efetuado por (ZurMuehlen, M. and Recker, J., 2008) em que é demonstrado que quem desenvolve os processos de negócio só usa uma pequena fração das potencialidades existentes em linguagens mais elaboradas como o caso do BPMN.

Sistemas com BPM

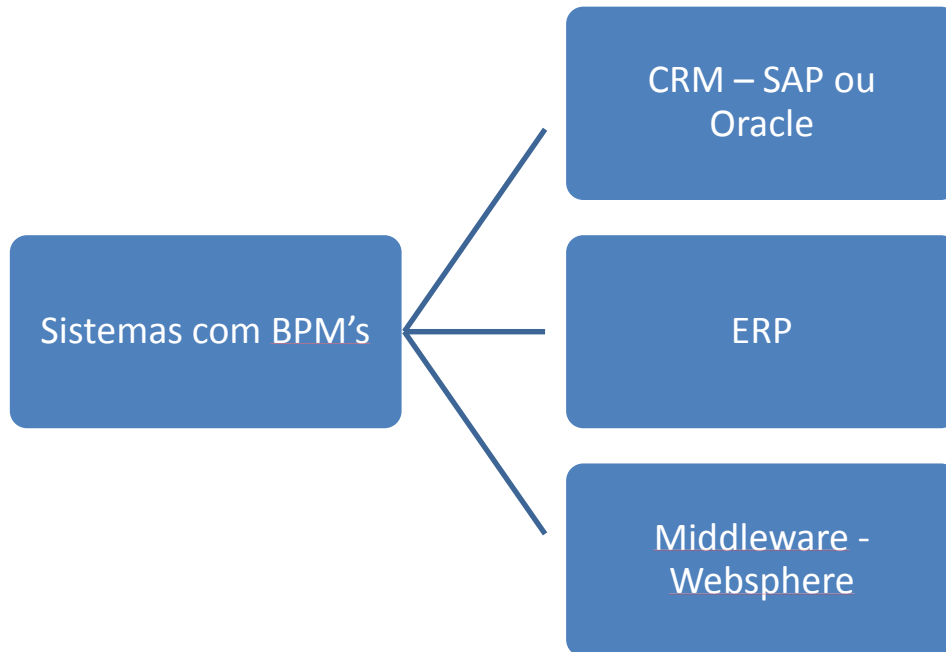


Figura 16 - Sistemas que têm tecnologia BPM, adaptado de (Aalst, W. M. P. van der., 2012)

Como referido anteriormente só na área da banca e na área dos seguros existem implementações de sistemas BPM “puras”, mas hoje em dia já existem sistemas montados que contêm BPM no seu interior bem como providenciam *workflows* para a criação e manutenção de processos de negócio. (Aalst, W. M. P. van der., 2012)

Através da figura 16, podemos observar como exemplo, alguns sistemas existentes atualmente que fornecem aos utilizadores a capacidade de gerir e criar os seus processos de negócio.

Ciclo BPM

Na figura 17 podemos observar o ciclo de vida do BPM, neste ciclo é nos possível visualizar as três principais fases do BPM, ou seja, a fase de desenho ou redesenho, a fase de implementação e configuração e a fase de execução ou manutenção.

A fase de desenho ou redesenho é onde é modelado o processo de negócio, ou no caso de este processo já tiver sido modelado anteriormente, serem desenvolvidos alguns ajustes que se tornaram necessários com o desenrolar do tempo.

A fase de implementação e ou configuração é onde ocorre a transformação do processo de negócio modelado para um sistema em execução, esta fase pode ser só colocado o sistema em execução ou pode ser necessário configurá-lo tendo em conta alguns parâmetros do negócio em questão. No caso deste processo já se encontrar em execução esta fase poderá ser bastante curta, pois trata-se única e exclusivamente de uma alteração pontual devido a alguma incoerência ou a um incremento de funcionalidades.

A fase de execução e manutenção é responsável pela validação do processo quando este está em execução, nesta o processo não é redesenhado nem é criado um novo *software*, é única e exclusivamente usada para alterar algo predefinido ou reconfigurar algum parâmetro que tenha sido alterado com o decorrer do processo ou então tenha sido incorretamente modelado.

Como está demonstrado na figura 6, os dados para modelar os processos de negócio podem ser analisados com base nos dados que geram ou através dos modelos que foram desenvolvidos para os processos.

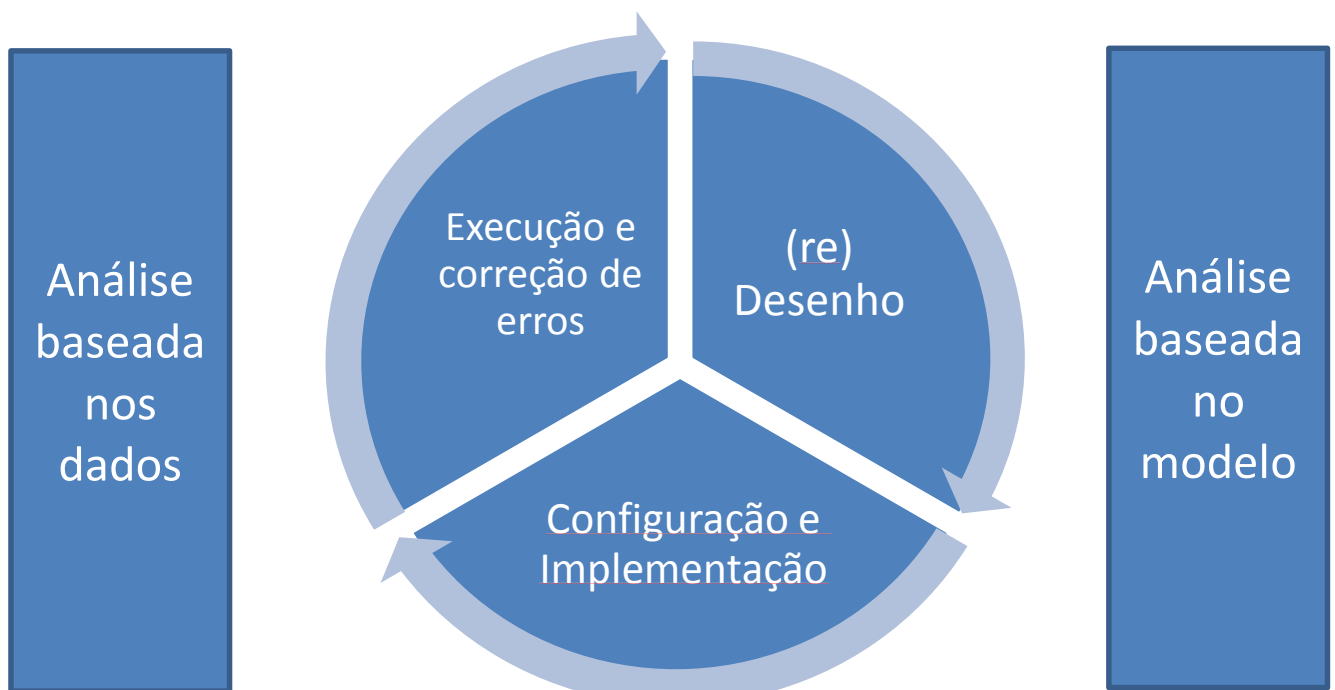


Figura 17 - Ciclo de vida do BPM, adaptado de (Aalst, W. M. P. van der., 2012) (Aalst, W. M. P. van der., 2004)

No caso da análise baseada em dados, estes são reunidos ao longo da execução do sistema, ou seja, são guardados todos os eventos para que posteriormente possam ser analisados e conseqüentemente encontrar possíveis falhas como *bottlenecks*, desperdícios ou desvios de performance, sendo esta análise feita enquanto o sistema se encontra em execução.

No caso da análise baseada em modelos, esta permite discernir se o sistema que está em execução está de acordo com aquele que foi modelado.

Process Mining

Evolução Histórica

Data Explosion

A exponencial evolução da tecnologia que temos assistido ao longo dos últimos anos, sendo ela tanto ao nível do *hardware* como do *software*, tem como grande impulsionadora a ciência da computação, tendo com base a lei de Moore de 1965, que diz que a cada ano a capacidade de processamento duplica em relação ao ano anterior, sendo esta análise baseada tanto o custo, pois passado um ano, por exemplo, um *chip* comprado por determinado valor terá um custo menor, como a performance, pois poderá haver outros com maior capacidade de processamento. (Aalst, W. M. P. van der., 2011)

Com estes avanços tecnológicos, existe cada vez maior dependência por parte das organizações, como por exemplo, entidades bancárias, seguradoras, agências de viagens, entre outras, devido ao fato dos seus dados terem passado de dados físicos, ou seja, de dados guardados tipicamente em papel, para dados digitais, que se encontram maioritariamente guardados em discos e servidores de grande porte. Mas não existe só uma grande dependência da parte das organizações, pois cada vez mais, através de *Smartphones* e *Tablets*, que hoje em dia se encontram acessíveis a uma grande quantidade de utilizadores particulares, fazem com que a quantidade de dados digitais seja gigantesca, pois estes utilizadores usam os seus aparelhos eletrónicos como repositório de todo o tipo de dados. (Aalst, W. M. P. van der., 2011) (NetCraft Ltd 2014, 2014) (David Sifry's musings, 2014) (Christine Alvarado, Jaime Teevan, Mark S. Ackerman and David Karger, 2013)

Esta gigantesca quantidade de dados digitais que existe hoje em dia no universo digital está desorganizada e não se encontra estruturada ou padronizada, o que leva a uma dificuldade acrescida por parte das organizações em lidar com este tipo de dados. Este é o grande desafio para as organizações de hoje em dia, onde necessitam de extrair informações e valor deste tipo de dados, que se encontram nos seus sistemas de informação. (Aalst, W. M. P. van der., 2011)

O crescimento exponencial da informação digital alinhado aos processos de negócio torna possível guardar e analisar eventos, estes podem ser de diversos tipos e de diferentes áreas da sociedade, como por exemplo, o levantar dinheiro de um ATM, um médico prescrever um medicamento, um utilizador solicitar a carta de condução, a entrega do IRS, entre outros. O desafio das organizações é conseguirem explorar de forma significativa os eventos dos seus sistemas de informação de modo a conseguirem encontrar, *bottlenecks*, informações relevantes, possíveis problemas, violação das políticas instauradas, formas de simplificar os processos instaurados.

(Aalst, W. M. P. van der., 2011) (Christine Alvarado, Jaime Teevan, Mark S. Ackerman and David Karger, 2013)

O que é?

Segundo W. M. P. van der. Aalst, o *process mining* é uma área de investigação emergente, que se posiciona entre a inteligência computacional e o *data mining* por um lado, e por outro, entre a modelação e a análise de processos. (Aalst, W. M. P. van der., 2011)

O *process mining* tem como principal objetivo descobrir, monitorizar e melhorar processos reais, ou seja, processos que existam e/ou que façam sentido para organizações. Sendo este feito através da extração de conhecimento de *event logs* existentes nos sistemas informáticos organizacionais. (Aalst, W. M. P. van der., 2011)

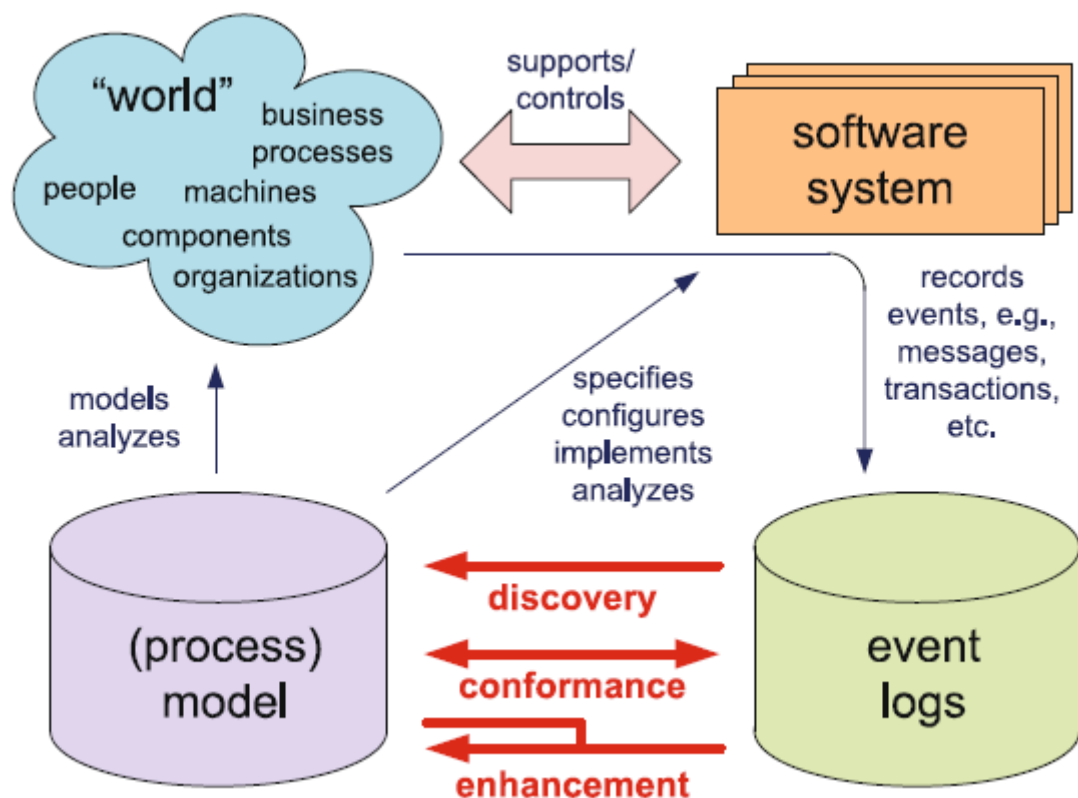


Figura 18 - Três principais tipos de *process mining*: *discovery*, *conformance* e *enhancement* (Aalst, W. M. P. van der., 2011)

Event logs

Hoje em dia, através dos sistemas de informação existentes nas organizações é possível aceder a uma quantidade gigantesca de eventos, podendo estes ser de diversos tipos, como por exemplo, transações, mensagens, recursos utilizados, tempo de execução, entre outros (como podemos observar na figura 19). O conjunto destes eventos, é denominado *event logs*.

Os *event logs* providenciam uma grande quantidade de informação detalhada sobre as ações que foram executadas. Estes podem guardar informações dos mais diversos sistemas, como WFM, BPM, ERP, SAP, CRM, entre outros.

Ao longo dos últimos anos, tanto o *event log*, como as técnicas de *data mining* têm maturado e hoje em dia já temos algumas técnicas capazes de extrair conhecimento de eventos. (Aalst, W. M. P. van der. and Schahram Dustdar, 2012)

Apesar da elevada quantidade e granularidade de informação existente nos *event logs*, muitas vezes esta não se encontra estruturada, isto é, por exemplo, os dados que queremos aceder encontram-se dispersos por diversas tabelas ou apresentam a necessidade de ser aprovados, neste caso os eventos existem mas o esforço para os extrair é bastante elevado. (Aalst, W. M. P. van der., 2011)

Tipos de Process Mining

Como podemos observar na figura 19, os diferentes tipos de *event logs* existentes na organização poderão dar origem a modelos de processos, esta transformação poderá ser efetuada através de três tipos distintos de *process mining*.

Estes são: (Aalst, W. M. P. van der. and Schahram Dustdar, 2012) (Aalst, W. M. P. van der., 2011)

- *Discovery* – Utilização de um *event log* e conseqüente criação de um modelo, isto sem a necessidade de aceder a nenhum tipo de informação *a-priori*. Existem diversas técnicas para a extração de modelos de processos através de um evento. Sendo este frequentemente utilizado como ponto de partida para efetuar análises aos processos organizacionais;
- *Conformance* – Comparação de um modelo de processo existente na organização e um *event log* do mesmo processo. Esta pode mostrar possíveis desvios sobre o que é real e o que foi modelado, fazendo com que se possa diagnosticar o porquê dessas diferenças. É muitas vezes utilizado para fins de auditoria;
- *Enhancement* – Utilização do modelo de processo e do *event log* que origina o mesmo, e estender ou melhorar o modelo através das observações

efetuadas aos eventos. Este é frequentemente utilizado para identificar *bottlenecks*, níveis de serviço e *performance*.

Perspectivas de análise do *Process Mining*

Podemos observar de quatro perspectivas diferentes os aspetos que o *process mining* analisa nas organizações. Estas perspectivas são: (Aalst, W. M. P. van der., 2011)

- *Control-Flow* – Esta é focada no *control-flow*, isto é, na sequência das atividades. Nesta, o *process mining* ajuda a encontrar a melhor sequência para as diferentes possibilidades existentes. Esta sequência pode ser expressa de diversas formas, como por exemplo, *Petri Nets*, BPMN e UML;
- Organizacional – Esta é focada nos recursos que se encontram no *event log*, como por exemplo, os atores que estão envolvidos no processo. Esta tem como principal objetivo definir o papel de cada ator na organização;
- Caso – Esta é focada nas propriedades de cada caso específico, pois estes podem ser caracterizados pelos valores dos seus elementos, como por exemplo, saber o número de produtos que foram encomendados numa encomenda;
- Tempo – Esta é focada no tempo de duração e na frequência com que ocorrem os eventos. Através desta análise é possível encontrar, *bottlenecks*, medir o nível do serviço, monitorizar a utilização de recursos, quer físicos, quer lógicos e prever o tempo de execução do processo quanto este é encontrado em *runtime*.

Muitas das atividades do *process mining* são desenvolvidas *off-line*, isto é, a análise dos processos é feita posteriormente à sua melhoria ou depois de ser melhor compreendido. Mas isto não significa que este só possa ser utilizado *off-line*, pois já existem técnicas que podem ser usadas *online*, isto é, por exemplo, a deteção de uma não conformidade no momento em que acontece o desvio, ou a previsão do tempo que irá demorar a executar um determinado caso em específico, sendo este feito, por exemplo, através da informação histórica recolhida previamente em casos com as mesmas características. (Aalst, W. M. P. van der., 2011)

Modelos de Processos e Realidade

Um dos grandes desafios da modelação, é desenvolver processos que se adequem à realidade, o que muitas vezes se torna uma tarefa bastante complexa, não pela dificuldade da modelação, mas sim pela elevada dificuldade de interpretação do modelo de negócio que se quer modelar, ou seja, a realidade.

O desafio da modelação é um dos elementos chave do *process mining*, pois existe a imperativa necessidade de criar uma relação bastante sólida entre o modelo de processo desenvolvido e a realidade, isto é, o que o *event log* consegue capturar. (Aalst, W. M. P. van der., 2011) Esta relação pode ser classificada utilizando a terminologia usada por David Harel, onde são usados os termos *Play-in*, *Play-out* e *Replay*. (D. Harel and R. Marely, 2003)

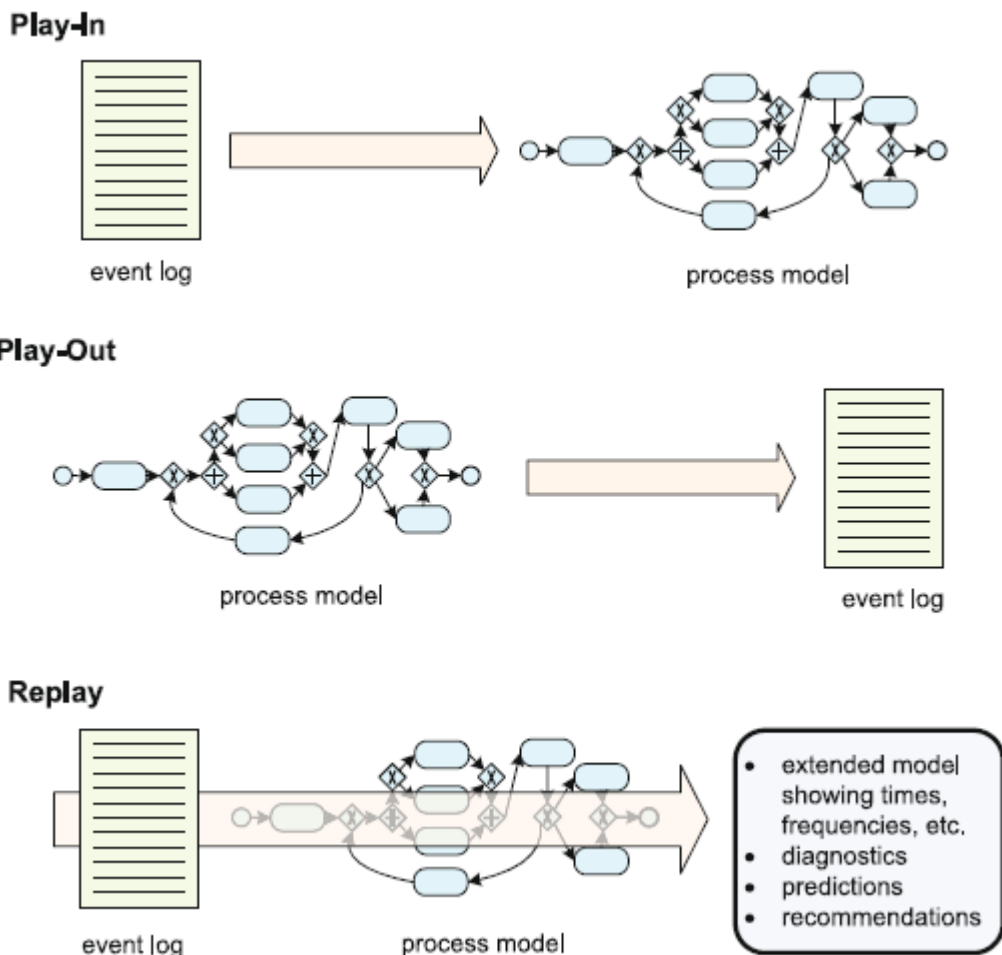


Figura 19 - Terminologias sobre a relação entre modelos de processos e a realidade (Aalst, W. M. P. van der., 2011)

Na figura 20 podemos observar uma representação das três diferentes formas de caracterizar a relação entre os modelos de processos e a realidade. Podemos então ter: (Aalst, W. M. P. van der., 2011) (D. Harel and R. Marelly, 2003) (C.W. Günther and Aalst, W. M. P. van der., 2007)

- *Play-out* – Este refere-se à utilização tradicional dos modelos de processos, isto é, é fornecido um modelo e através do mesmo é gerado o seu comportamento. Pode ser usado tanto para analisar como para *enactment* de processos de negócio. Uma *Workflow engine* pode ser vista como uma “*Play-out engine*”, pois monitoriza os casos de estudo, permitindo apenas o que está definido no modelo previamente concebido. A *Play-out engine* pode ser utilizada para fazer operações de *enactment*, isto através da utilização de ferramentas simulação, pois estas permitem simular diversas vezes o mesmo modelo e guardar intervalos de confiança e diversas outras estatísticas;
- *Play-in* – Este é o oposto do *Play-in*, isto é, o *input* é o comportamento e o objetivo é construir o modelo. O *Play-in* é associado muitas vezes à inferência. O processo *discovery* é um exemplo de uma técnica de *Play-in*. Grande parte das técnicas de *data mining* utilizam *Play-in*, isto é, um modelo é criado através da aprendizagem de um conjunto de exemplos;
- *Replay* – Esta utiliza um *event log* e um modelo de processo como *input*. O *event log* é “multiplicado”, ou seja, são criados outros com as mesmas características. Um *event log* pode ser replicado por quatro razões distintas:
 - Validar a conformidade – Possibilidade de analisar discrepâncias que existam entre o modelo do processo e os múltiplos *event logs* que foram replicados;
 - Expandir o modelo com frequências e informação temporal – Através da replicação do *event log* é possível analisar o número de vezes que cada step do modelo é utilizado, podendo assim detetar *bottlenecks*;
 - Construir *predictive models* – Isto é, através da repetição do *event log* e dos diferentes dados guardados em histórico, é possível prever por exemplo, o tempo estimado para terminar o processo;
 - Suporte operacional – O *Replay* não é exclusivamente utilizado *off-line*, isto é, com eventos guardados em histórico, pois este pode ser usado também *online*, onde é possível analisar desvios que aconteçam enquanto o processo está em execução.

Caminho ótimo nos modelos de processos

O *Replay* pode ser usado para caracterizar o caminho ótimo, isto é, o caminho que alia a eficácia à eficiência, existente em cada processo. Isto é possível através da multiplicação do mesmo *event log* e consequente execução dos diversos eventos que

foram replicados, fazendo assim com seja possível efetuar uma análise acerca das ligações mais propícias ente os diferentes caminhos passíveis de serem efetuados. (C.W. Günther and Aalst, W. M. P. van der., 2007)

Com a utilização desta técnica é possível, através do histórico existente, prever e “forçar”, o melhor comportamento a ser efetuado pelo processo, ou seja, o caminho ótimo. (Aalst, W. M. P. van der., 2011)

Overfitting vs Underfitting

Os algoritmos do *process mining* têm a necessidade de generalizar o comportamento que está contido no *event log*, ou seja, generalizar aquilo que é percebido como a realidade. Isto para não se tornarem inválidos no caso de existir um novo conjunto de observações. (Aalst, W. M. P. van der., 2011)

Um dos grandes desafios do *process mining* é o equilíbrio entre: (Aalst, W. M. P. van der., 2011)

- *Overfitting* – O modelo está demasiado “bem treinado” para o conjunto de observações, ou seja, é muito específico e muito dificilmente será uma solução viável no caso de existir um novo conjunto de observações;
- *Underfitting* – O modelo é bastante genérico, isto é, permite que sejam adotados comportamentos que não eram os esperados. Isto pode-se dever a um conjunto de observações bastante reduzido.

Mas este equilíbrio não apresenta uma forma linear e unanime de ser conseguido, pois pode depender de diversos fatores, como por exemplo: (Aalst, W. M. P. van der., 2012)

- Caso de estudo – Este pode apresentar características que não sejam possíveis de contemplar numa generalização, o que leva a que o modelo não esteja em conformidade com a generalidade do *event log*;
- Modelo incorretamente desenvolvido – Este pode estar a ter uma perspetiva da realidade incorreta, o que leva a um comportamento errático por parte do processo em relação ao esperado;
- *Event log* incorreto – Este pode dever-se ao fato de existir um desvio na realidade, ou seja, o que era esperado não é o que está realmente a acontecer.

Noise e Incompleteness

Para criar novos modelos de processos é necessário existir um *event log* representativo do comportamento do processo que se pretende modelar, ou seja, um conjunto de observações significantes acerca do mesmo. (Aalst, W. M. P. van der., 2011)

Apesar de em muitas situações existirem esses *event logs* necessários para a criação de modelos, estes podem apresentar algumas dificuldades ao nível da representatividade das observações. Estas dificuldades em grande parte estão relacionadas com: (Aalst, W. M. P. van der., 2011) (Aalst, W. M. P. van der., 2012)

- *Noise* – Isto é, as observações existentes representam um comportamento pouco frequente e não é possível tipificar um para o processo em questão. Apesar de não se possível tipificar um comportamento genérico para um modelo não significa que o *event log* esteja errado, de fato pode estar correto, pois está a perceber a realidade de forma correta, mas a realidade apresenta características demasiado específicas para ser generalizada. É importante separar *noise* de *logging* incorreto, pois o algoritmo irá efetuar a análise das observações não será capaz de distinguir um erro de um evento esporádico, para isto será sempre necessário a validação humana, quer antes, quer após o processamento;
- *Incompleteness* – Isto é, as observações existentes no *event log* são muito reduzidas, e como tal, não é possível criar um modelo subjacente ao histórico existente.

Análise de modelos

Existem duas abordagens distintas para efetuar uma análise em relação aos modelos de processos, estas são: (E.M. Clarke, O. Grumberg, and D.A. Peled., 1999) (Z. Manna and A. Pnueli, 1991) (R.J. van Glabbeek and W.P. Weijland, 1996) (H.M.W. Verbeek, T. Basten, and Aalst, W. M. P. van der., 2001)

- *Verificação* – Esta tem como principal foco, a validação da correta implementação, quer do sistema, quer do processo. O modelo de processo desenvolvido pode ser incorreto por diversos fatores, como por exemplo, um desvio relativo ao que estava especificado e ao que na realidade foi desenvolvido, a criação de *deadlocks*, isto é, a existência de uma sequência de passos que conduza a fim não desejado. Para detetar estes estes possíveis erros, já são conhecidas algumas técnicas capazes efetuar essa validação;
- *Análise da performance* – Esta tem como principal foco, a análise do tempo que o processo precisa para terminar, os tempos de espera, a utilização e o nível de serviço. Este tipo de análise pode ser dividida por três dimensões

distintas, sendo que cada uma delas pode ser medida através de diferentes KPIs. Estas dimensões são:

- Tempo – Quando efetuamos uma análise em relação a esta dimensão temos os seguintes pontos como métricas:
 - Tempo do processo – É a medida desde o tempo da criação do caso até ao seu término. Podemos efetuar uma média sobre os tempos de processamento, mas esta pode estar enviesada por diferentes fatores, pois numa dada altura, pode demorar 8 horas e noutras 20 horas, por esta razão é também importante saber a variância entre os diferentes tempos de processamento;
 - Tempo de serviço – É o tempo que se esteve realmente a solucionar o caso. Este tempo tendencialmente é medido por cada uma das atividades desenvolvidas, no caso de terem existido atividades concorrentes, assume-se o tempo da atividade com o tempo mais elevado;
 - Tempo de espera – É o tempo em que o caso se encontra à espera de recursos para se tornar disponível. Podemos tomar como exemplo, o tempo em que um cidadão está num determinado hospital à espera de ser atendido por um especialista;
 - Tempo de sincronização – É o tempo em que a atividade ainda não está completamente disponível, encontrando-se à espera de um agente externo para a libertar. Esta métrica é diferente da do tempo de espera, pois no caso do tempo de espera existe a necessidade de esperar por falta de recursos disponíveis, nesta o tempo de espera diz respeito à necessidade de um agente externo.
- Custo – Para efetuar uma análise de custo podemos utilizar um conjunto de modelos já existentes, como por exemplo:
 - Modelo ABC – Este é mais conhecido por, modelo de atividades baseadas no custo, onde os custos são analisados com base nas atividades planeadas;
 - *Time Driven* ABC – Este, tem muitas semelhanças com o modelo ABC, e apresenta como característica diferenciadora, o fato de calcular o custo pelo tempo que demora cada atividade a ser concluída;
 - *Resource Consumption Accounting* – Este, é um modelo recente mais recentes em comparação com os supracitados, e tem como base a gestão Alemã. Apresenta como principais características, o dinamismo, a flexibilidade e a integração total entre as atividades e o tempo.
- Qualidade – Frequentemente a análise da qualidade é desenvolvida com base no produto ou serviço que é entregue ao cliente. Esta qualidade pode ser medida através de questionários de satisfação

efetuados aos clientes ou por outro lado pode ser medida com base no rácio entre os produtos acabados e os produtos com defeito.

Estas duas abordagens permitem que os modelos de processos possam ser otimizados em relação, ao tempo, ao custo e qualidade, através da utilização de técnicas já existentes, como por exemplo, a simulação. (H.M.W. Verbeek, T. Basten, and Aalst, W. M. P. van der., 2001)

Modelos de Processos

Os modelos de processos são processos que têm a mesma natureza, isto é, que têm a mesma tipologia e o seu desenvolvimento é semelhante, o que leva a que possam ser usados em diferentes tipos de aplicações, e por isso podem ser colocados dentro do mesmo modelo. (Rolland, Colette and Pernici, C. Thanos, 1998) (Colette Rolland, 1994)

Tipicamente estes modelos são utilizados nas organizações para diminuir a entropia em tarefas que apresentam alguma periodicidade, pois estes apresentam um conjunto de regras/normas de como deve/deveria/poderia ser feito por parte de quem executa o processo. (Rolland, Colette and Pernici, C. Thanos, 1998) (P.H. Feiler and W.S. Humphrey, 1993)

Os principais objetivos dos modelos de processos são (Aalst, W. M. P. van der., 2011) (Rolland, Colette and Pernici, C. Thanos, 1998):

- Descrição do processo – Análise acerca do que acontece realmente durante o processo e conseqüentemente determinar medidas capazes de tornar o processo mais eficaz e mais eficiente;
- Determinar as etapas do processo – Definir como deve/deveria/poderia ser efetuado o processo, para isto, são estabelecidas regras/normas para desta forma reduzir a flexibilidade de quem efetua o processo, e desta forma ter um comportamento padronizado;
- Explicação do processo – Explicação acerca do funcionamento lógico e racional do processo, através do modelo de negócio em questão.

Os modelos de processos de negócio podem ter diversas utilizações, mas as principais são: (Aalst, W. M. P. van der., 2011)

- Conhecimento – Quando estes estão a ser desenvolvidos, o utilizador tem diferentes perspetivas do processo que está a modelar e pode adquirir maior conhecimento sobre o mesmo;
- Discussão – Os *Stakeholders* utilizam estes modelos para estruturarem as suas discussões acerca do processo em questão;

- Documentação – Os modelos de processos são documentados de modo a providenciar diretrizes a quem os irá utilizar e a certificar os mesmos, como por exemplo, ISO 9000 *Quality Management*;
- Verificação – Os modelos de processos são analisados por forma a tentar encontrar erros nos sistemas ou nos procedimentos, como por exemplo, *dedlocks*;
- Análise de *performance* – São usadas diferentes técnicas para perceber os fatores que podem influenciar o tempo de resposta ou o nível do serviço;
- Especificação – Estes modelos podem ser usados para descrever o PAISs antes da implementação e podem também ser usados como “contratos” entre quem o modelador e o utilizador;
- Configuração – Estes modelos podem ser usados para configurar o sistema em questão.

Estas diferentes formas de utilização de modelos de processos podem ser visualizadas também na figura 18, onde podemos observar em que fases do ciclo de vida do BPM acontecem.

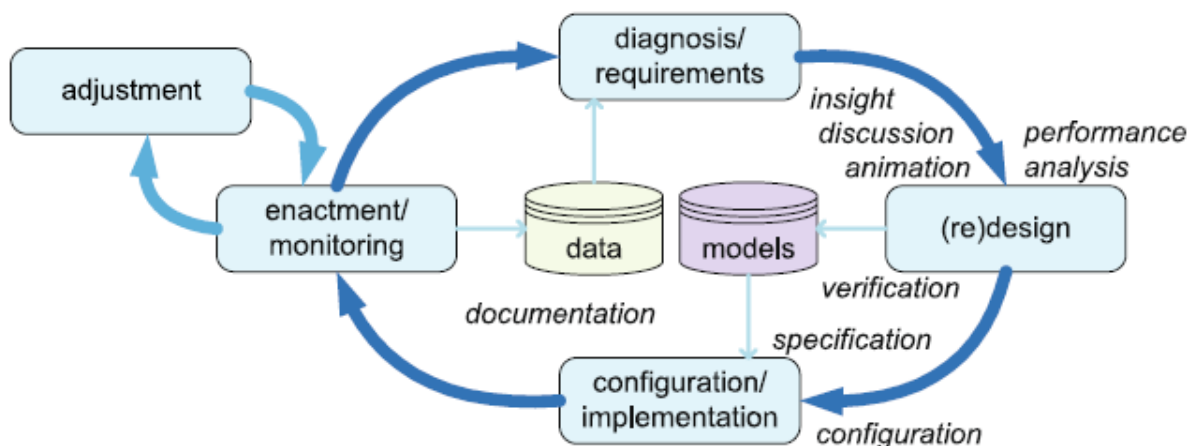


Figura 20 - Ciclo de Vida do BPM com as diferentes utilizações dos modelos de processos (Aalst, W. M. P. van der., 2011)

Tipos de Modelos

Os modelos de processos encontram-se divididos em dois tipos diferentes de modelos. Estes são: (M. Dumas, Aalst, W. M. P. van der. and A.H.M. ter Hofstede, 2005) (A.H.M. ter Hofstede, Aalst, W. M. P. van der, M. Adams, and N. Russell, 2010) (M. Weske, 2007) (Aalst, W. M. P. van der., 2011)

- Modelos formais – Estes são usados em *runtime* para analisar e *enactment*. São muito detalhados e de elevada complexidade, pois tem como principal foco os *stakeholders* ;

- Modelos Informais – Estes são usados para discutir ideias e melhorias, e para o desenvolvimento de documentação. São muitas vezes ambíguos e vagos, pois são apresentados como diagramas em *PowerPoint*.

O crescente interesse nos modelos de processos por parte das organizações, a abundância de eventos e a reduzida qualidade de modelos que são efetuados por utilizadores faz com que existe uma necessidade crescente de relacionar eventos de dados para desenvolver modelos de processos. Desta forma é possível descobrir novos modelos de processos e avaliar os modelos de processo já existentes. (Aalst, W. M. P. van der., 2011)

Desafios da Modelação

A modelação de processos através no BPM frequentemente tem como objetivo dar resposta a múltiplos pressupostos. Para além de poder ser expressa em diferentes tipos de notações, ou seja, em diferentes tipos de linguagens, pode dar resposta a diversas questões, como por exemplo, as responsabilidades de cada ator no processo, analisar a complacência, fazer uma previsão acerca da *performance* através da simulação e configurar sistemas. (Aalst, W. M. P. van der., 2011) (A.H.M. ter Hofstede, Aalst, W. M. P. van der, M. Adams, and N. Russell, 2010)

Apesar das capacidades existentes atualmente para o desenvolvimento de modelos viáveis, nem sempre é fácil de conseguir este desenvolvimento, e tipicamente podem ocorrer alguns erros, como por exemplo: (A.H.M. ter Hofstede, Aalst, W. M. P. van der, M. Adams, and N. Russell, 2010) (Aalst, W. M. P. van der., 2011) (Aalst, W. M. P. van der, H.T. de Beer and B.F. van Dongen, 2005)

- O modelo descrever uma realidade idealista – Este pode acontecer quando o modelador apresenta a tendência para se concentrar no suposto comportamento normal ou desejado, levando assim a ter um modelo que só está focado numa determinada percentagem das observações possíveis. Os modelos que são desenvolvidos por seres humanos apresentam frequentemente uma elevada subjetividade para que sejam fáceis de entender;
- Incapacidade para capturar o comportamento humano – Os modelos geralmente são constituídos por algoritmos e modelos matemáticos e por seres humanos. Quando existe a necessidade de modelar múltiplos processos que se encontram expostos a múltiplas características ou propriedades, os seres humanos podem revelar uma inadequação à capacidade de observação, isto deve-se em grande parte à dificuldade de distribuir a sua atenção por diversos processos, pois em vez de estar focado em todos os pormenores relativos a um determinado processo tem a sua atenção dispersa;

- O modelo encontra-se no nível de abstração errado – Este depende bastante da questão, ou temática que se pretende responder, bem como os seus *inputs*, pois só através desses dois pontos é possível decidir o nível de abstração a ter. Este nível deve-se muitas vezes à dificuldade de encontrar resposta detalhada sobre a questão que se pretende, ou por outro lado a questão que se está a responder não é relevante ou até mesmo a incapacidade de obtenção, por diversos motivos, da resposta que é requerida;

Limitações da análise de modelos

Esta análise, baseada na verificação e *performance* tem uma elevada relação com a realidade, o que leva a que modelos que não tenham muito em comum com a realidade não façam sentido ser analisados deste prisma. Isto deve-se ao fato de, por exemplo, mesmo que este satisfaça todos os pressupostos que tinham sido delineados, este pode ser inútil para a organização. (Aalst, W. M. P. van der., 2011)

Grande parte destas limitações situam-se entre o fato de o desenvolvimento ser efetuado por seres humanos e a realidade. (Aalst, W. M. P. van der., 2011)

Diferenças entre Data Mining e Process Mining

Como foi demonstrado em capítulos anteriores existem muitas semelhanças entre o *process mining* e o *data mining*, contudo também existem algumas diferenças. (Aalst, W. M. P. van der., 2011) (D.Hand, H. Mannila and P.Smyth, 2001) (E. Alpaydin, 2010) (I.H. Witten and E. Frank, 2005)

Na tabela abaixo são apresentadas essas diferenças.

<i>Process Mining</i>	<i>Data Mining</i>
Os <i>input's</i> do <i>process mining</i> podem ser <i>event logs</i> , <i>audit trails</i> e eventos de tecnologias de informação.	Os <i>input's</i> das técnicas de <i>data mining</i> são tabelas com diferentes tipos de dados.
Tem em conta a ordem dos eventos, isto é, o conjunto de dados utilizado para análise tem que conter um <i>timestamp</i> e um número sequencial tendo em conta a data de criação do evento.	Não tem em consideração a ordem dos eventos apresentados no <i>data set</i> .
A perspectiva do <i>process mining</i> não reside nos padrões existentes nos dados mas sim nos processos contidos nos dados.	Utilização de vistas multidimensionais, mais conhecidos por cubos, o que permite agregar diferentes níveis dimensionais, levando assim a que seja possível efetuar <i>drilled down</i> e <i>drilled up</i> sobre os dados.
Utilização para encontrar processos relacionados no conjunto de dados disponibilizado, permitindo assim ter uma visão sobre as atividades existentes nos diversos processos e podendo ainda inferir desvios do processo, melhorias de performance e <i>bottlenecks</i> .	Utilização com a finalidade de encontrar padrões num elevado número de dados.
O principal objetivo é encontrar informação relativamente aos processos de negócio.	

Tabela 3 - Diferenças entre o *process mining* e o *data mining*. (Aalst, W. M. P. van der., 2011) (D.Hand, H. Mannila and P.Smyth, 2001) (M. Bramer, 2007) (E. Alpaydin, 2010) (I.H. Witten and E. Frank, 2005)

Ferramentas de Suporte

O *Process Mining* como referido anteriormente, é uma metodologia bastante recente dentro dos processos de negócio, (Aalst, W. M. P. van der., 2011) e desta forma ainda não apresenta um vasto leque de ferramentas que apliquem/disponibilizem os conceitos de forma genérica e fiável.

Hoje em dia existem duas aplicações que possibilitam a aplicação destes conceitos e metodologias, contudo, estas ainda estão numa fase embrionária, o que faz com que tenham ainda algumas limitações, principalmente no que diz respeito à forma como os dados são carregados das fontes.

As aplicações são o *Prom*, que neste momento está na versão 6.4, e o *XESame*, que neste momento está na versão 1.2. (Hverbeek, 2104) (Jbuijs, 2014)

O *Prom*, é uma framework de código aberto que implementa algoritmos de *process mining*. Esta, apresenta uma plataforma para utilizadores e para *developers*. (Hverbeek, 2104)

Esta aplicação, permite modelar processos de negócio através da utilização de um ficheiro com eventos, sendo que este ficheiro deve de respeitar determinados padrões, para que possa ser consumido pelo *Prom*. (Hverbeek, 2104)

A criação do ficheiro supracitado, ou seja, do ficheiro de eventos a ser utilizado pelo *Prom* pode ser criado através do *XESame*.

O *XESame* é uma aplicação desenvolvida para utilizadores que permite a extração de eventos de diferentes tipos de fontes, ou seja, bases de dados, ficheiros excel, entre outros, e posterior criação de *Event Log's*. (Jbuijs, 2014)

Como referido no anteriormente, estas aplicações apresentam algumas limitações.

No que diz respeito ao *Prom*, este necessita de um ficheiro de eventos altamente padronizado, sendo que este não apresenta um guia para a sua construção e denota uma elevada complexidade. Este ficheiro de eventos, tem que ser gerado pelo *XESame*.

No que diz respeito ao *XESame*, ainda é uma aplicação muito recente e não suporta muitas fontes de dados, apesar de a documentação indicar que são suportados diferentes tipos de fontes, a realidade é que esta ainda não está estável o suficiente para fontes de dados complexas como bases de dados, e desta forma apresenta alguns erros aplicativos ainda não solucionados.

Em termos de ferramentas capazes de implementar uma abordagem de *process mining* de montante a jusante, tendo em conta a pesquisa desenvolvida não

foram encontradas praticamente nenhuma, visto ser uma ciência relativamente recente e que ainda se encontra pouco explorada. Contudo, da pesquisa que foi possível efetuar encontrei uma denominada *Fluxicon*. (Fluxicon Process Laboratories, 2014)

Esta, apresenta custos para o utilizador, e apesar da investigação não encontrei nenhuma possibilidade de obter uma licença de teste, logo não me foi possível analisar a aplicação, tendo em conta os seus pontos fortes e os pontos fracos.

Conclusões do Capítulo

Neste capítulo foram apresentados os principais conceitos que estão relacionados com o *Process Mining*, correlacionando o *Business Process management* e o *Data Mining*, pois o *Process Mining* é uma área que me grande parte derivou uma parte do *Business Process Management* e outra do *Data Mining*.

Após a apresentação destes conceitos é possível então iniciar a investigação tendo por base os conceitos teóricos, existindo assim a possibilidade de aplicar estes a um problema específico.

Metodologia

Introdução de Capítulo

Neste capítulo será apresentada a metodologia utilizada no desenvolvimento desta dissertação. Será também apresentado o planeamento do trabalho a ser realizado e a forma como esse trabalho irá decorrer.

Planeamento

Introdução

Este planeamento tem como finalidade, estabelecer o tempo destinado a cada tarefa, os recursos que as irão desenvolver e as respetivas etapas do desenvolvimento desta dissertação.

Descrição

Em baixo, está representado o planeamento sobre a forma de árvore. Neste, podemos observar que este planeamento está dividida em dez grandes fases, que em seguida passarei a enumerar e discriminar.

- **Elaboração da Proposta:** Esta fase contempla uma investigação acerca das principais temáticas a ser abordadas ao longo do desenvolvimento, sendo que findada esta investigação, será discriminado os principais pontos e objetivos a serem abordados ao longo desta dissertação;
- **Elaboração do Enquadramento:** Esta fase contempla a descrição dos principais pontos a serem abordados ao longo da dissertação de uma forma sintética e muito orientada para a descrição do problema a resolver, bem como a solução proposta para tal;
- **Elaboração do Estado da Arte:** Esta fase contempla uma investigação aprofundada acerca das três principais temáticas a serem abordadas ao longo da dissertação, entre elas, o *process mining*, o *business process management* e o *data mining*, bem como uma posterior revisão acerca de cada um dos seus temas;
- **Elaboração da Metodologia de Trabalho:** Esta fase contempla a descrição de todo o processo onde está baseado o caso de estudo desta dissertação, isto é, a forma como este será abordado, os principais desafios encontrados, as soluções propostas e a metodologia de trabalho utilizada;
- **Desenvolvimento do Protótipo:** Esta fase contempla o desenvolvimento de um protótipo, que terá como base o carregamento dos emails do *MS Outlook*, tratamento dos mesmos e posterior análise tendo em conta alguns procedimentos e algoritmos utilizados;
- **Desenvolvimento do Manual de Utilizador.** Esta fase contempla a elaboração de um manual que auxilie o utilizador na exploração e na utilização do protótipo previamente desenvolvido;
- **Elaboração do Relatório Final:** Esta fase contempla a agregação de todos os pontos previamente desenvolvidos até esta parte, de modo a criar um relatório final;

- **Elaboração da Apresentação Final:** Esta fase contempla o desenvolvimento da apresentação final que será feita tendo por finalidade, elucidar o júri acerca do processo que compôs o desenvolvimento deste trabalho;
- **Apresentação Final:** Esta fase contempla a apresentação previamente desenvolvida, bem como a discussão oral sobre os pontos que o júri entender necessários abordar.

Especificação da Tarefa	Nome da Tarefa	Duração	Início	Fim	Recurso
Agendamento Manual	Elaboração da Proposta da Dissertação				Jorge Martins
Agendamento Manual	Investigação acerca da temática <i>Process Mining</i>	12 dias	Domingo 1/5/14	Sábado 1/18/14	Jorge Martins
Agendamento Manual	Investigação acerca das ferramentas existentes no mercado sobre <i>Process Mining</i>	11 dias	Domingo 1/26/14	Sexta-feira 2/7/14	Jorge Martins
Agendamento Manual	Desenvolvimento do documento para a proposta	7 dias	Domingo 2/9/14	Segunda-feira 2/17/14	Jorge Martins
Agendamento Manual	Elaboração do Enquadramento da Dissertação				Jorge Martins
Agendamento Manual	Desenvolvimento dos objetivos	4 dias	Quarta-feira 2/19/14	Sábado 2/22/14	Jorge Martins
Agendamento Manual	Desenvolvimento do problema a ser resolvido	2 dias	Domingo 2/23/14	Segunda-feira 2/24/14	Jorge Martins
Agendamento Manual	Desenvolvimento da metodologia a utilizar	2 dias	Terça-feira 2/25/14	Quarta-feira 2/26/14	Jorge Martins
Agendamento Manual	Desenvolvimento dos resultados esperados	1 dia	Quinta-feira 2/27/14	Quinta-feira 2/27/14	Jorge Martins
Agendamento Manual	Elaboração do Estado da Arte				Jorge Martins
Agendamento Manual	Investigação acerca da temática <i>Business Process Management</i>	12 dias	Sábado 3/1/14	Domingo 3/16/14	Jorge Martins
Agendamento Manual	Investigação acerca da temática <i>Process Mining</i>	9 dias	Quarta-feira 3/19/14	Segunda-feira 3/31/14	Jorge Martins
Agendamento Manual	Investigação acerca da temática <i>Data Mining</i>	11 dias	Domingo 4/6/14	Sexta-feira 4/18/14	Jorge Martins

Agendamento Manual	Desenvolvimento dos módulos do estado da arte	9 dias	Domingo 4/20/14	Quarta-feira 4/30/14	Jorge Martins
Agendamento Manual	Elaboração da Metodologia de Trabalho				Jorge Martins
Agendamento Manual	Desenvolvimento do planejamento	3 dias	Quinta-feira 5/1/14	Sábado 5/3/14	Jorge Martins
Agendamento Manual	Desenvolvimento da abordagem ao problema	17 dias	Domingo 5/11/14	Sábado 5/31/14	Jorge Martins
Agendamento Manual	Desenvolvimento da solução para o problema	7 dias	Domingo 6/1/14	Sábado 6/7/14	Jorge Martins
Agendamento Manual	Desenvolvimento do protótipo				Jorge Martins
Agendamento Manual	Apresentação do protótipo	7 dias	Domingo 6/15/14	Sábado 6/21/14	Jorge Martins
Agendamento Manual	Investigação acerca das ferramentas a utilizar para desenvolver o protótipo	7 dias	Domingo 6/22/14	Sábado 6/28/14	Jorge Martins
Agendamento Manual	Criação do protótipo				Jorge Martins
Agendamento Manual	<i>Entrega 1</i>				Jorge Martins
Agendamento Manual	Desenvolvimento da Base de Dados	10 dias	Terça-feira 7/1/14	Sábado 7/12/14	Jorge Martins
Agendamento Manual	Testes	7 dias	Domingo 7/13/14	Sábado 7/19/14	Jorge Martins
Agendamento Manual	<i>Entrega 2</i>				Jorge Martins
Agendamento Manual	Carregar Emails	12 dias	Domingo 7/20/14	Sábado 8/2/14	Jorge Martins
Agendamento Manual	Testes	7 dias	Domingo 8/3/14	Sábado 8/9/14	Jorge Martins
Agendamento Manual	<i>Entrega 3</i>				Jorge Martins
Agendamento Manual	Estatísticas	12 dias	Domingo 8/10/14	Sábado 8/23/14	Jorge Martins
Agendamento Manual	Testes	7 dias	Domingo 8/24/14	Sábado 8/30/14	Jorge Martins
Agendamento Manual	<i>Entrega 4</i>				Jorge Martins
Agendamento Manual	Tratamento dos Eventos (<i>Mining</i>)	21 dias	Segunda-feira 9/1/14	Sábado 9/27/14	Jorge Martins
Agendamento Manual	Testes	7 dias	Domingo 9/28/14	Sábado 10/4/14	Jorge Martins

Agendamento Manual	Entrega 5				Jorge Martins
Agendamento Manual	Configuração	12 dias	Domingo 10/5/14	Sábado 10/18/14	Jorge Martins
Agendamento Manual	Testes	7 dias	Domingo 10/19/14	Sábado 10/25/14	Jorge Martins
Agendamento Manual	Desenvolvimento do Manual de Utilizador	7 dias	Sábado 11/1/14	Sábado 11/8/14	Jorge Martins
Agendamento Manual	Revisão do Relatório de Dissertação	15 dias	Domingo 11/9/14	Quinta-feira 11/27/14	Jorge Martins
Agendamento Manual	Entrega da Dissertação	1 day	Sexta-feira 11/28/14	Sexta-feira 11/28/14	Jorge Martins
Agendamento Manual	Elaboração e preparação da apresentação da Dissertação				Jorge Martins
Agendamento Manual	Apresentação da dissertação				Jorge Martins

Tabela 4 - Representação do planeamento sobre a forma de tabela

Abordagem

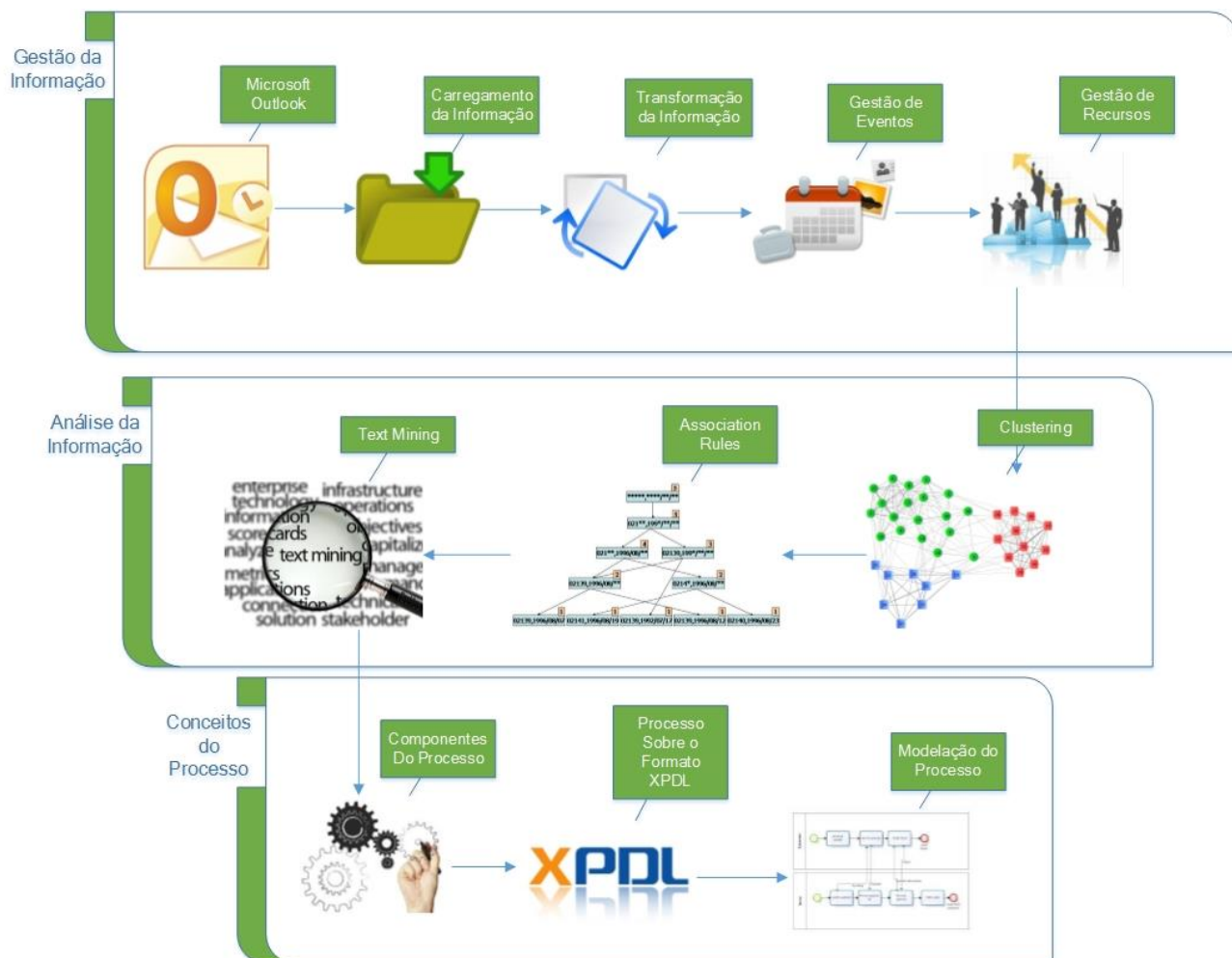


Figura 21 - Esquema sobre a abordagem desenvolvida ao longo da investigação

Na figura 21, apresentada em cima, pode ser observado um esquema acerca dos elementos que foram desenvolvidos ao longo desta investigação, bem como uma a sua posição no plano.

Estes elementos serão apresentados e detalhados no decorrer deste capítulo.

Outlook

Toda a análise apresentada em baixo tem como principal fonte as versões do *Microsoft Office* 2007 até ao 2013.

O *Microsoft Outlook* apresenta um conjunto bastante diversificado de funções e potencialidades, como por exemplo, o envio de emails, o agendamento de reuniões, a marcação de eventos, o guardar pequenas notas ou informações úteis, o guardar grandes quantidades de informação relativas a contatos, entre outras mais complexas e mais personalizáveis, tendo em conta as necessidades dos utilizadores.

Todas estas funções do *Microsoft Outlook* são denominadas por itens, ou seja, seja um email, uma reunião, um agendamento de um evento, uma nota, um contato, é considerado um item. Sendo que dentro do conjunto dos itens existem diferentes subconjuntos, e é nestes que se enquadram os diversos itens descritos atrás.

Estes subconjuntos de itens têm associadas propriedades e características específicas. Os mais relevantes e mais utilizados subconjuntos são:

- *Appointments* – Este subconjunto, apresenta todas as características e propriedades específicas relativas ao agendamento de eventos efetuado;
- *Contacts* – Este subconjunto, apresenta todas as características e propriedades específicas relativas à informação de contatos previamente guardada;
- *Notes* – Este subconjunto, apresenta todas as características e propriedades específicas relativas às mensagens de email enviadas ou recebidas;
- *Post* – Este subconjunto, apresenta todas as características e propriedades específicas relativas às notas que são criadas pelo utilizador;
- *Meeting* – Este subconjunto, apresenta todas as características e propriedades específicas relativas às reuniões previamente agendadas pelo utilizador ou ainda por agendar mas que se encontram à espera de resposta;
- *Task* – Este subconjunto, apresenta todas as características e propriedades específicas relativas às tarefas que se encontram criadas, ou assignadas ao utilizador.

Contudo, dentro destes subconjuntos ainda existem mais subdivisões de itens, sendo que estas são muito mais específicas. Nestas, podemos ter funções como regras de receção ou envio de emails, categorização dos emails recebidos, notificações, entre outras funções altamente personalizáveis pelo utilizador.

Apesar de existirem diferentes tipos de itens e de terem características distintas de subconjuntos de itens para subconjuntos de itens, estes também apresentam algumas propriedades semelhantes, como por exemplo:

- *Emissor* – Este, é o recurso que envia o item;

- Recetor – Este, é o recurso que recebe o item, sendo que este pode não ser único;
- Assunto – Este, é o título ou o resumo que caracteriza o item;
- Data de criação – Esta, é a data em que o item foi criado;
- Data de envio – Esta, é a data em que o item foi enviado pelo emissor para o recetor.

Estas propriedades semelhantes permitem assim uniformizar grande parte dos itens provenientes do *Microsoft Outlook*.

Carregamento da informação

Tendo em conta as distintas características existentes entre os itens, torna-se necessário encontrar uma forma de recolher a maior quantidade de informação possível dos diferentes itens existentes nas pastas dos utilizadores.

Para efetuar uma recolha dos dados de forma fiável é necessário criar estruturas diversas, que se adaptem a cada um dos diferentes tipos de itens, desta forma torna-se imperativo obter uma estrutura distinta para cada um dos subconjuntos de itens existentes.

A título exemplificativo é apresentada em baixo a estrutura criada para guardar os dados relativos aos *MailItems*:

- Id – Este, guarda o número responsável por identificar cada um dos itens deste tipo;
- Tipo de item – Este, guarda o tipo de item;
- Assunto – Este, guarda o assunto referente ao item;
- Número de sequência – Este, guarda um número identificador por tópico de interação, isto é, se existirem vários emails sobre o mesmo assunto, terão o mesmo número de sequência;
- Conhecimento – Este, guarda o nome das pessoas para as quais o item foi enviado, sendo que estas não os principais destinatários, mas sim, aquelas que devem de ter conhecimento sobre o assunto falado;
- Data de criação – Este, guarda a data em que o item foi criado;
- Data da última modificação – Este, guarda a data em que o item foi alterado pela última vez;
- Corpo da mensagem – Este, guarda todo o texto que está colocado no corpo do item, no caso de ser um email, é o corpo do email;
- Anexos – Este, no caso de existirem, guarda o nome e o tipo de ficheiro dos anexos enviados;
- Categoria – Este, guarda a categoria definida pelo utilizador para este tipo de item;

- Recetor – Este, guarda o nome, ou nomes, dos principais destinatários do item;
- Data de Envio – Este, guarda a data em que o item foi enviado;
- Data de receção – Este, guarda a data em que o item foi lido por parte do recetor;

Tratamento da Informação

Como referido anteriormente, o objetivo desta dissertação é descobrir processos de negócio existentes em emails trocados pelos mais diversos utilizadores através do *Microsoft Outlook*.

Através da utilização das características partilhadas pelos diferentes tipos de itens existentes no *Outlook*, tornou-se possível a criação de uma estrutura base. Sendo que a principal finalidade desta, é englobar todos os itens existentes nas diferentes pastas do *Outlook*, que apesar destes itens apresentarem uma origem distinta, contêm variáveis semelhantes, como por exemplo, as apresentadas em cima.

A estrutura criada é denominada por evento, este a nível conceptual é um email enviado de um recurso para outro, sendo que no caso de ser um email enviado para diversos recursos serão criados tantos eventos quanto o número de recursos a que o email se destinava. Um evento é composto por:

- Id – Este, é o identificador único de cada um dos eventos;
- Tipo de email – Este, é composto pela descrição de cada um dos diferentes tipos de itens já enunciados atrás;
- Data de criação – Esta, é composta pela data em que o emissor efetuou a criação do item;
- Data de envio – Esta, é composta pela data em que o item foi enviado do emissor para o recetor;
- Número de sequência – Este, é composto pelo identificador de item, isto é, cada item criado tem um valor sequencial associado, isto faz com que todas as interações que aconteçam entre os diversos recursos relativas a este tópico tenham o mesmo número de sequência;
- Emissor – Este, é composto pelo nome do recurso que criou e por sua vez efetuou o envio do item;
- Recetor – Este, é composto pelo nome do recurso que recebeu o item;
- Assunto – Este, é composto pela descrição do tópico ou assunto do item.

Com a utilização desta estrutura de eventos, possibilitada pela partilha de propriedades dos diversos itens do *Microsoft Outlook*, é possível também criar uma estrutura para guardar os recursos que intervêm na interação que existe constantemente por parte dos utilizadores, sendo que estes recursos serão os utilizadores que emitem e que recebem itens, sejam eles de que tipo forem.

A estrutura criada para guardar esta informação é denominada por recursos e é constituída por:

- Id – Este, é o indentificador único relativo a cada um dos recursos;
- Nome – Este, é composto pela descrição do nome do utilizador, isto é, o nome através do qual o utilizador está registado no domínio da aplicação.

Gestão de Eventos

Toda a informação recolhida do *Microsoft Outlook*, ou seja, todos os itens que foram carregados, serão posteriormente transformados em eventos, para que possam dar origem a um *event Log*.

Estes eventos são criados tendo por base algumas características, sendo estas características transversais a todos os itens. Estas, encontram-se dissecadas no capítulo anterior, bem como a apresentação da estrutura do evento.

Tendo em conta os padrões definidos em (Aalst, W. M. P. van der., 2011), apresentados anteriormente na secção estado da arte, acerca de como deve estar estruturado um *event log*, a estrutura do evento criado aplica as propriedades necessárias para ser definido como um *event log*.

As características primordiais para que seja um *event log*, prendem-se como fato de apresentar um identificador único por evento, para que desta forma seja possível identificar e ordenar rapidamente cada um dos eventos, uma data de criação do evento, permitindo assim a análise temporal de cada evento, um recurso de origem, isto é, o recurso que despontou o evento, sendo neste caso o recurso que efetuou o envio de um item, um recurso de destino, ou seja, o recurso a quem se destina o evento, sendo neste caso o recetor do item enviado.

Em baixo, na figura 21 podemos observar um exemplo de um *event log* criado através da extração dos itens provenientes do *Microsoft Outlook*, sendo que neste caso é apenas uma parcela do *event log* que foi criado, pois no seu todo apresenta uma quantidade bastante significativa de eventos.

Id	E mailType	CreationDate	SentOnDate	SequenceNumber	from	to	subject
9	Outlook.MailItem	5/29/2014 9:07	5/29/2014 9:08	7	Martins, Jorge	Martins, Rui M.	Mapa de horas Affinity
53	Outlook.MailItem	6/30/2014 12:12	6/30/2014 12:14	7	Martins, Jorge	Martins, Rui M.	Mapa de horas Affinity
55	Outlook.MailItem	7/1/2014 12:01	7/1/2014 12:03	7	Martins, Jorge	Martins, Rui M.	Aprovação de horas
56	Outlook.MailItem	7/1/2014 12:05	7/1/2014 12:05	7	Martins, Jorge	Martins, Rui M.	Mapa de horas Affinity
57	Outlook.MailItem	7/1/2014 13:33	7/1/2014 13:35	7	Martins, Jorge	Martins, Rui M.	Mapa de horas
159	Outlook.MailItem	8/29/2014 12:57	8/29/2014 12:59	7	Martins, Jorge	Martins, Rui M.	Mapa de horas de Agosto
161	Outlook.MailItem	8/29/2014 13:38	8/29/2014 13:38	7	Martins, Jorge	Martins, Rui M.	Mapa de horas de Agosto
212	Outlook.MailItem	9/30/2014 13:24	9/30/2014 13:26	7	Martins, Jorge	Martins, Rui M.	Aprovação Mapa de Horas de Setembro
458	Outlook.MailItem	5/29/2014 9:11	5/29/2014 9:11	7	Martins, Jorge	Martins, Rui M.	Mapa de horas Affinity
705	Outlook.MailItem	7/1/2014 12:05	7/1/2014 12:05	7	Martins, Rui M.	Martins, Jorge	Mapa de horas Affinity

Figura 22 - Exemplo de um event log

Como é possível observar no exemplo apresentado em cima, para além dos conceitos necessários para que seja um *event log* viável, existem outro tipo de informações que faziam sentido serem incluídas tendo em conta as particularidades dos itens existentes no *Microsoft Outlook*.

Para satisfazer então as necessidades impostas, foi adicionada mais informação ao *event log*, sendo esta, o tipo de item, para que desta forma existisse a possibilidade de associar cada um dos eventos a um diferente tipo de item, a data de envio, para que para além de já ser possível averiguar a data de criação do item, ser possível também saber a data em que o item efetivamente foi enviado, o número de sequência, que indica o número do tópico relativo ao evento, e por fim o assunto, visto ser esta a forma primordial de análise e agrupamento de cada um dos eventos.

Gestão de Recursos

Todos os itens recolhidos do *Microsoft Outlook*, para além de poderem ser expressos sob a forma de eventos, também apresentam outra particularidade que os torna viáveis à criação de *event logs*, que são os recursos, ou seja, os utilizadores que enviam e que recebem os itens.

Desta forma, em todos os itens, e consequentemente em todos os eventos existem dois recursos associados, isto é, um recurso será o emissor do evento, ou seja, quem criou ou evento e posteriormente o irá enviar, e o recetor do evento, que é o recurso que irá receber o evento.

É possível observar, a título exemplificativo, na figura 21 dois recursos distintos que interagiram num determinado conjunto de eventos.

Estes recursos são criados através da informação que é recolhida do *Microsoft Outlook*, sendo esta obtida tendo em conta os emissários e os recetores dos itens, ou seja, no caso de existir um email que é enviado para três recursos distintos, iremos

ter três eventos em tudo semelhantes, menos no que diz respeito ao recetor e ao identificador único do evento.

À semelhança do que se procede com os eventos, os recursos também apresentam uma estrutura onde são armazenados, sendo que esta está diretamente interligada com a dos eventos, pois só é possível um determinado recurso existir num evento se este existir na estrutura de recursos.

A estrutura desenvolvida para suportar os recursos apresenta as seguintes propriedades:

- Id – Este, representa o identificador único de cada um dos recursos;
- Nome – Este, representa o campo onde estará armazenado o nome do recurso.

Estes recursos provêm da análise efetuada a toda a informação recolhida do *Microsoft Outlook*, ou seja, de todos os itens que foram enviados foram retirados os emissores e os remetes de modo a identificar todos os recursos que faziam parte da rede de contatos.

Para além destes recursos se encontrarem armazenados numa estrutura criada para o efeito, estes também podem ser associados a papéis dentro da organização, isto é, por exemplo o recurso “A” poderá ser o gestor, ao passo que o recurso “B” poderá ser o comercial. Contudo, estes papéis atribuídos não são estáticos, ou seja, a qualquer momento podem ser alterados, o que permite que num determinado período temporal o recurso “B” seja o comercial, mas posteriormente poderá ser promovido e desta forma deixará de ter o papel de comercial atribuído para passar a ter o de gestor.

Na figura 22, apresentada em baixo, podemos vislumbrar a título exemplificativo uma parcela dos recursos previamente criados por forma a posteriormente serem consumidos e darem origem a eventos.

Id	Name
1	Martins, Jorge
2	Moreira, Carlos
3	Patronilo, Nuno
4	Santos, Ludovic
5	Martins, Rui M.
6	Marques, Antonio
7	Santos, Bruno
8	Costa, Jose
9	Alves, Nuno
10	Ricardo Desidério

Figura 23 - Exemplo dos recursos guardados com vista à criação do event log

Clustering

Como já referido anteriormente, os eventos apresentam particularidades mas também muitas semelhanças entre si. Estas, fazem com que seja possível descortinar grupos onde se podem englobar diversos eventos, mas apesar de serem diferentes apresentam características em tudo similares.

De forma a conseguir encontrar as propriedades que cada um dos grupos deve corresponder, torna-se necessário desenvolver uma análise a todos os tipos de item que são recolhidos do *Microsoft Outlook*.

Tendo em conta a análise desenvolvida sobre os dados recolhidos, isto é, sobre os itens armazenados, foi possível verificar que existiam algumas relações diretas entre determinadas propriedades.

No que diz respeito às propriedades que se correlacionam diretamente, temos em primeira instância o assunto do item, isto é, o tópico a que corresponde o item. No caso de existirem diversos eventos com o mesmo assunto, isto indica que esses eventos pertencem à mesma interação, ou seja, que todos esses eventos fazem parte de uma troca de emails acerca da mesma temática, ou de alguma tarefa periódica, ou de alguma reunião com espaço temporal definido e sistemático.

Posto isto, todos os eventos com o mesmo tópico, ou seja, com o mesmo assunto, estão associados através de um número que identifica o tópico. Este número é denominado como número de sequência. Podemos observar na figura 21, a título exemplificativo, que para aquele conjunto de eventos devido a todos apresentarem o mesmo tópico, ou tópicos cuja relação é muito estreita por se tratar do mesmo assunto com periodicidades distintas, apresentam o mesmo número de sequência, neste caso o número 7.

Como supracitado, o assunto não tem que ser necessariamente igual para fazer parte do mesmo grupo de análise, pois se o assunto apresentar uma semelhança de cerca de 80%, considera-se que o evento faz parte do mesmo grupo, isto porque como referido anteriormente, esta diferença pode significar, um mês diferente, um dia da semana diferente, um nome diferente na atribuição de uma tarefa, contudo e apesar destas pequenas diferenças, o âmbito continua a ser semelhante e como tal deverá fazer parte do mesmo grupo de análise.

Através da data de envio e de criação torna-se possível criar uma cronologia com os eventos que se encontram divididos dentro do mesmo grupo, levando assim à possibilidade de saber quando é que o primeiro evento aconteceu, quando é que aconteceu o último, e se nos restantes eventos é possível gerar um espaço temporal sistemático, ou seja, se os eventos acontecem num período de tempo fixo.

Esta divisão dos eventos por grupos com propriedades semelhantes permite tornar uma posterior análise mais fiável, pois desta forma quando existir a necessidade de encontrar regras no que diz respeito a quem envia e quem recebe os itens, ou o no que diz respeito ao seu conteúdo, esta já será real e adaptada ao seu âmbito real e não ao conjunto total dos eventos, que levaria a uma análise bastante envesada.

A opção por este agrupamento e pela utilização deste tipo de abordagem prende-se com o fato de o assunto ser uma propriedade transversal a todos os itens e desta forma poder aplicar a mesma característica a todos os eventos em análise, evitando assim erros de análise no que diz respeito à utilização de diferentes características no conjunto de dados recolhidos.

Rules

No capítulo anterior, está dissecada a metodologia utilizada para agrupar os eventos por *clusters*, com vista a obter grupos de eventos com propriedades semelhantes.

Posteriormente ao agrupamento dos eventos tendo em conta as suas propriedades transversais, torna-se vital efetuar uma análise acerca das regras que estão subjacentes ao envio e à receção de itens, ou seja, procurar possíveis regras que correlacionem os recursos entre si, tendo sempre em conta o grupo em que estão inseridos.

A abordagem desenvolvida para procurar estas possíveis regras foi baseada no algoritmo *Apriori*, sendo que como já foi desenvolvido no capítulo relativo ao *data mining*, este tipo de algoritmos assenta numa análise aos objetos existentes em bases de dados, podendo estas ser de diferentes tipos, onde se pretende encontrar os objetos que mais vezes ocorrem tendo em conta determinados critérios bem como os objetos que mais vezes ocorrem em simultâneo segundo os mesmo parâmetros. (H. Mannila, H. Toivonen, and A.I. Verkamo, 1997)

No que diz respeito aos itens recolhidos do *Microsoft Outlook*, a metodologia utilizado teve como principais pontos de atuação os recursos que se relacionavam e a forma como o faziam nos diversos eventos, sendo que esta análise é desenvolvida sempre tendo por base os grupos previamente criados de itens.

Para efetuar este tipo de análise torna-se imprescindível definir duas medidas fundamentais para a obtenção de resultados viáveis. Estas medidas são a confiança e o suporte. (H. Mannila, H. Toivonen, and A.I. Verkamo, 1997) (Aalst, W. M. P. van der., 2011) (D.Hand, H. Mannila and P.Smyth, 2001)

O suporte é um valor percentual que indica a proporção de objetos de um determinado tipo que existem no grupo de eventos analisado sobre o número total de objetos existentes no grupo, ou seja, por exemplo se existir um grupo onde tenha dez eventos, onde nove foram enviados pelo recurso “A” e um pelo recurso “B”, existirá 90% de suporte no que diz respeito ao recurso “A”, em contra ponto existirá 10% de suporte em relação ao recurso “B”. (Aalst, W. M. P. van der., 2011) (D.Hand, H. Mannila and P.Smyth, 2001) (H. Mannila, H. Toivonen, and A.I. Verkamo, 1997)

No que diz respeito à confiança, esta é também um valor percentual que desta feita é desenvolvida para averiguar quando é que no mesmo grupo de eventos existe o recurso “A” e o recurso “B” consecutivamente, sendo isto de de entre todo o tipo de combinações possíveis de objetos, ou seja, numa análise a quatro recursos, temos cerca de vinte e quatro possibilidades de emparelhamento entre recursos. (Aalst, W. M. P. van der., 2011) (D.Hand, H. Mannila and P.Smyth, 2001) (H. Mannila, H. Toivonen, and A.I. Verkamo, 1997)

Para a análise desenvolvida decidiu-se utilizar como valores padrões de procura um suporte igual a 60% e uma confiança igual a 80%. Estes foram os valores escolhidos, pois de entre a literatura observada foram os números mais consensuais no que diz respeito a uma análise robusta e fiável. Contudo, estes valores podem ser adaptados tendo em conta o âmbito em que se pretende desenvolver a análise. (Aalst, W. M. P. van der., 2011) (D.Hand, H. Mannila and P.Smyth, 2001) (H. Mannila, H. Toivonen, and A.I. Verkamo, 1997)

Posto isto, ao efetuar uma pesquisa ao longo dos eventos do mesmo grupo é possível descortinar se existem recursos que aparecem em 60% dos eventos, e desta forma serem viáveis tendo em conta o valor que foi definido para o suporte, e se os dois objetos existentes como recurso que envia o evento e o recurso que recebe o evento aparecem juntos em 80% de todas as combinações possíveis.

Tendo a título exemplificativo a figura 21, podemos observar que se fosse desenvolvida uma análise aquele conjunto de eventos poderiam ser obtidos como objetos que satisfaçam o valor do suporte requerido dois recursos, e no que diz respeito à confiança, estes recursos aparecem 100% das vezes juntos, o que faz com que também correspondam aos 80% de confiança predefinidos como o mínimo necessário para que o conjunto de objetos seja considerado uma regra.

Contudo, estas regras não tem em consideração a ordem pela qual os recursos aparecem, ou seja, não é possível tendo em conta única e exclusivamente a aplicação deste género de algoritmos saber quem é o emissor e o recetor do evento, pois os conjuntos não tem em conta a ordem sob a qual aparecem os objetos, ou seja, não validam se o recurso “A” aparece sempre antes do recurso “B” ou vice-versa. Isto, leva a que seja necessário fazer uma análise posterior para validar quem na realidade é que é o emissor e o recetor do evento.

Text Mining

Após agrupar os eventos tendo em conta as suas semelhanças em relação às propriedades comuns aos diversos itens provenientes do *Microsoft Outlook*, foi efetuada uma análise onde se pretende encontrar regras que possam existir entre os recursos que interagem entre si tendo em conta o âmbito previamente selecionado.

Contudo, esta análise não é suficiente para a obtenção de um processo de negócio válido, pois existe ainda a necessidade de analisar o corpo dos itens previamente guardados.

Esta análise revela-se fundamental pois, é através dela que será possível descortinar verdadeiramente as tarefas que são desenvolvidas no processo, bem como os responsáveis por elaborar cada uma dessas e por fim a periodicidade, no caso de esta existir, em que devem ser efetuadas cada uma das tarefas.

A análise terá como base o corpo dos itens, como referido anteriormente, mas também os possíveis anexos que estejam agrupados aos itens, para desta forma dar origem a um conjunto de tarefas.

A elaboração da análise terá como principais pressupostos três pontos fundamentais:

- Contar as palavras mais escritas - Ao desenvolver este tipo de abordagem é possível saber que palavras são mais utilizadas no conjunto dos eventos selecionados, permitindo assim com que seja viável obter um conjunto de ações ou expressões que sejam indispensáveis para o correto mapeamento dessas ações ou expressões para tarefas. Após a aplicação desta abordagem, e depois de estarem reunidas um conjunto com as palavras que mais vezes foram escritas no corpo das mensagens, torna-se necessário validar de forma mais específica cada uma dessas palavras, isto deve-se em grande parte ao fato de poderem existir uma grande diversidade de palavras que não apresentam uma correlação direta com as necessidades vinculadas nos eventos analisados, contudo apesar de não fazerem sentido para a análise pretendida não significa que não existam em grande quantidade no corpo dos itens analisados. Para contornar esta situação, optou-se por centrar esta abordagem em palavras que apresentem as seguintes características:
 - Tempos verbais – Através da restrição da análise aos tempos verbais é possível viabilizar as palavras que mais vezes estão expostas no corpo das mensagens como tarefas, isto em grande parte porque os tempos verbais são utilizados quando se pretende desenvolver uma ação;
 - Palavras com conotação temporal – Através da restrição da análise a palavras com conotação temporal, isto é, nomes dos meses ou dias de semana, torna-se possível criar um espaço temporal para a tarefa ou tarefas em causa, possibilitando assim a criação de processos com

tempos bem definidos, isto é, estabelecer que por exemplo um determinado processo deverá ser executado todos os meses, no último dia de cada mês.

- Analisar as palavras precedentes da palavra mais escrita – Ao desenvolver este tipo de abordagem torna-se possível criar um conceito para o objetivo final, que é a elaboração de um processo. Pois, através desta análise é possível saber qual a palavra que aparece mais vezes a preceder a palavra que mais vez aparece, isto leva a que exista a possibilidade de conhecer o conjunto de palavras mais utilizado, ou seja, a expressão mais utilizada. A título exemplificativo, podemos observar na figura 21, que o conjunto de palavras que mais vezes aparece é “Mapa Horas”;
- Analisar as palavras consequentes da palavra mais escrita – À semelhança da análise que é desenvolvida para saber as palavras precedentes da palavra mais vezes escrita, a análise ao conjunto de palavras consequentes à palavra que mais vezes está exposta é idêntica, levando assim a um conhecimento mais robusto sobre o processo real que se pretende modelar. Utilizando mais uma vez como exemplo a figura 21, podemos vislumbrar que ao englobar as três palavras obtém-se a expressão “Mapa Horas Affinity”, isto faz com que seja viável indicar um processo a modelar com o nome “Mapa Horas Affinity”.

Ao nível dos anexos, a análise tem como base diferentes critérios, pois neste caso, não era viável fazer uma análise única e exclusivamente ao nome do anexo, pois este pode apresentar variações nominais, contudo o objetivo poderá ser o mesmo.

Para fazer face a esta necessidade, foi utilizado como critério o tipo de anexo, ou seja, o tipo de ficheiro que está anexado, e determinadas palavras-chaves que possam existir ou não no anexo. As palavras-chaves utilizadas foram aquelas que mais vezes estavam expostas no corpo do item, os meses do ano e os dias da semana.

Após esta análise torna-se possível averiguar se determinado anexo é único naquele conjunto de itens, ou se em contra ponto, existe uma semelhança entre os diferentes anexos existentes no conjunto de eventos previamente selecionados. Esta semelhança pode-se prender por exemplo, com o simples fato de o mesmo ficheiro ser enviado todos os meses, contudo o nome do anexo é distinto, mas a diferença reside no nome do mês e não no âmbito.

No caso de o anexo estiver identificado como parte do âmbito dos eventos selecionados, então este também será considerado uma tarefa a ser modelada no processo.

Componentes do Processo

Após terem sido encontrados os eventos que apresentam o mesmo âmbito, ou seja, que fazem parte do mesmo grupo de eventos, os recursos que estão envolvidos nesses mesmos eventos tendo em conta os critérios já apresentados atrás, as tarefas descortinadas da análise desenvolvida ao conjunto dos itens selecionados, e por fim a periodicidade ou não das tarefas ou do processo no seu todo, torna-se possível atingir o objetivo final proposto, ou seja, o fornecimento de todos os componentes necessários ao desenvolvimento de um processo de negócio.

Torna-se então possível recolher todas as informações que foram analisadas independentemente e desta forma aplicá-las a componentes utilizados com a finalidade de desenvolver processos de negócios.

Desta forma podemos mapear as informações utilizando a seguinte abordagem:

- Papéis - Este componente, está diretamente relacionado com os recursos que anteriormente foram descortinados como os intervenientes nos eventos. Os recursos serão então os utilizadores responsáveis por elaborar cada uma das tarefas ou atividades que lhes forem atribuídas. Estes papéis não são necessariamente o nome dos recursos, mas sim o papel que lhes está atribuído dentro da organização, ou seja, apesar de análise ser desenvolvida tendo por base o nome dos recursos, estes encontram-se mapeados em funções dentro da organização, podendo estas funções ser de diferentes tipos tendo em conta o ramo em que se insere o núcleo comercial da companhia;
- Nome do processo – Este componente, está diretamente relacionado com a análise desenvolvida ao corpo dos itens, ou seja, a abordagem utilizada para averiguar a expressão mais utilizada ao longo do conjunto dos eventos em análise. Esta expressão será o nome do processo, bem como a forma como será percecionado pela organização;
- Tarefa ou atividade – Este componente, à semelhança do nome do processo está diretamente relacionado com a análise elaborada ao corpo dos itens, sendo que desta feita as tarefas ou atividades serão os tempos verbais encontrados em maior número ao longo de todos os eventos analisados, sendo que cada uma destas tarefas ou atividades estará associada a um papel, ou seja, a um recurso, sendo este o responsável pela sua correta execução tendo em conta as diretrizes previamente estabelecidas;
- Correlação entre tarefas ou atividades – Este, não é necessariamente um componente, pois apresenta-se como a forma através da qual as tarefas ou atividades estão sequenciadas, isto é, indicam que a tarefa “A” deve ser elaborada antes da tarefa “B”. Contudo, é fundamental para conhecer a correta sequencia que as tarefas ou atividades devem ter. Através da análise que foi desenvolvida às palavras precedentes e consequentes da palavra que mais

vezes está escrita no corpo do item, podemos obter uma sequenciação das tarefas ou atividades a serem desenvolvidas, pois sabemos que por exemplo, a palavra “X” aparece maioritariamente antes da palavra “Y”, o que faz com que a tarefa ou atividade “X” seja uma precedência da tarefa ou atividade “Y”;

- Periodicidade – Este componente, está diretamente relacionado com a análise que foi desenvolvida tendo por base das palavras-chaves, como o nome dos meses do ano e os dias da semana, e permite automatizar determinadas tarefas ou atividades que apresentem um espaço temporal de execução estático, ou num caso mais abrangente automatizar um processo no seu todo na situação de este única e exclusivamente ser despontado entre períodos de tempo fixos.

Através do mapeamento supracitado torna-se possível desenvolver processos de negócio informais através da análise dos itens recolhidos do *Microsoft Outlook*.

XPDL

Com toda a informação recolhida para a criação de novos processos de negócio torna-se necessário encontrar uma forma de agregar toda essa informação e transformá-la num fluxo, ou seja, modelar o processo.

Para desenvolver esta modelação de forma correta existe um sem número de ferramentas capazes de o fazer, algumas destas isentas de qualquer pagamento outras com custos para os utilizadores e para as organizações. Contudo, estas aplicações necessitam de intervenção humana para a consequente modelação dos processos, ou seja, estão dependentes de especialistas em modelação de processos capazes de transformar a informação recolhida por parte dos analistas organizacionais em modelos de processos.

Por forma a fazer face a estes desafios existe a necessidade de tornar esta modelação autónoma em relação ao utilizador, isto é, possibilitar a modelação do processo sem existir qualquer intervenção por parte do utilizador.

De modo a responder a este desafio e ultrapassar esta situação, decidiu-se utilizar uma tecnologia denominada *XPDL*, conhecida como *XML Process Definition Language*, que é uma abordagem padrão direcionada para o desenvolvimento de *workflows*, ou seja, que permite a modelação de processos de negócio. Esta abordagem, como referido atrás, é padrão, ou seja, pode ser utilizada em diversas aplicações concebidas para a criação de modelos, permitindo assim que através da criação de um ficheiro deste formato e consequente importação para uma aplicação previamente desenvolvida para a elaboração de processos de negócio, seja gerado um modelo. (wfmc.org, 2014)

Através da utilização do *XPDL*, torna-se possível utilizar toda a informação recolhida anteriormente, e consequentemente mapeada para uma visão direcionada a processos de negócio, e automatizar a modelação de novos processos de negócio tendo em conta toda a análise previamente desenvolvida.

Contudo, e apesar do *XPDL* ser um formato padrão para a modelação de processos de negócio, este apresenta determinadas peculiaridades de aplicação para aplicação, isto é, nem todos os ficheiros são corretamente importados por todas as aplicações desenvolvidas para o efeito.

Tendo em conta este desafio, a decisão passou por utilizar o Bizagi, que é uma aplicação utilizada para modelar processos de negócio, como ferramenta onde seria importado o *XPDL* desenvolvido. (bizagi.com, 2014)

Com primeira abordagem ao desenvolvimento de um ficheiro com as características necessárias para ser importado pelo Bizagi, foi necessário desenvolver uma análise a ficheiros previamente exportados da ferramenta, pois desta forma seria possível analisar os componentes que eram necessários à correta importação, bem como a sua forma de construção.

No capítulo anterior foi apresentada a forma como estavam mapeados todos os componentes necessários ao desenvolvimento de um processo de negócio, através desse mapeamento, tornou-se possível aplicar as componentes previamente descortinadas ao ficheiro *XPDL* a ser importado para o Bizagi.

A estrutura do ficheiro é estática, o que faz com que seja que processo for, o ficheiro apresentará sempre as mesmas características. Para os diferentes processos, as diferenças prendem-se única e exclusivamente com os componentes que estes necessitem, por exemplo, um determinado processo terá quatro tarefas repartidas por dois recursos, enquanto outro terá dez tarefas repartidas por quatro recursos.

Desta forma, é possível consumir o objetivo final, que se prendia com a necessidade de criar processos de negócio informais através da informação recolhida do *Microsoft Outlook*, isto porque para além de recolhida a informação necessária para o desenvolvimento dos processos também é possível modelá-los sem intervenção do utilizador.

Resultados

Os resultados finais, após a aplicação de toda a abordagem apresentada, assentam em novos ou otimização de processos de negócio, podendo estes estar modelados ou no estado de metadados, isto é, sobre o formato *XPDL*.

Contudo, estes resultados não são o fim da análise, definitivos, nem poderão ser vistos como um produto acabado e pronto a ser implementado a nível organizacional, pois toda esta análise pode sempre ser melhorada, quer seja a montante, isto é no carregamento e transformação da informação recolhida, como a jusante, no que respeita à análise de toda a informação e consequente descortinar de padrões.

Sendo que o melhoramento supracitado, deve de assentar principalmente no nível organizacional, isto é, tendo por base as necessidades específicas das organizações bem como as especificidades inerentes de cada um dos modelos de negócio que podem existir dentro da mesma organização.

Conclusões do Capítulo

Com este capítulo pretendeu-se dar a conhecer a abordagem e as metodologias aplicadas na investigação da temática proposta, esmiuçando ao detalhe todas as opções e decisões tomadas ao longo da investigação desenvolvida.

Protótipo

Introdução de Capítulo

No decorrer deste capítulo será apresentado o protótipo desenvolvido com a finalidade de demonstrar de forma prática toda a investigação desenvolvida, bem como apresentar um possível produto que seja fiável e possa ser utilizado pelas organizações.

Apresentação do Protótipo

Após a apresentação de toda a metodologia utilizada durante a investigação torna-se preeminente demonstrar que todas as opções tomadas e todos os pontos focados podem ser aplicados nas organizações existentes no mercado.

Desta forma, foi desenvolvido um protótipo onde são aplicados todos os conceitos já evidenciados no capítulo anterior.

Arquitetura

Por forma a demonstrar todos os componentes utilizados para o desenvolvimento do protótipo pode ser observada na figura 23, apresentada em baixo, uma arquitetura aplicacional que demonstra o funcionamento da aplicação desenvolvida, bem como os componentes que foram utilizados para o seu desenvolvimento.

Existem quatro grandes grupos no desenvolvimento da aplicação, a fonte dos dados, o desenvolvimento da aplicação, a estrutura que guarda os dados, ou seja, base de dados e por fim a modelação.

No que diz respeito à fonte dos dados, estes provêm unicamente do *Microsoft Outlook*.

Em relação ao grupo do desenvolvimento aplicacional, este é composto pela metodologia utilizada, ou seja, o modelo de três camadas, sendo que este modelo se define por diferenciar as camadas aplicacional, dados e interface, o tipo de tecnologia utilizada para desenvolver a aplicação, isto é, *WPF C#*, e por fim as bibliotecas auxiliares utilizadas.

O grupo relativo à base de dados contém apenas a base de dados desenvolvida para suportar a aplicação.

Por fim, o grupo destinado à modelação tem como componentes o tipo de ficheiros utilizado para a obtenção de processos de negócio, bem como a ferramenta utilizada para os modelar.

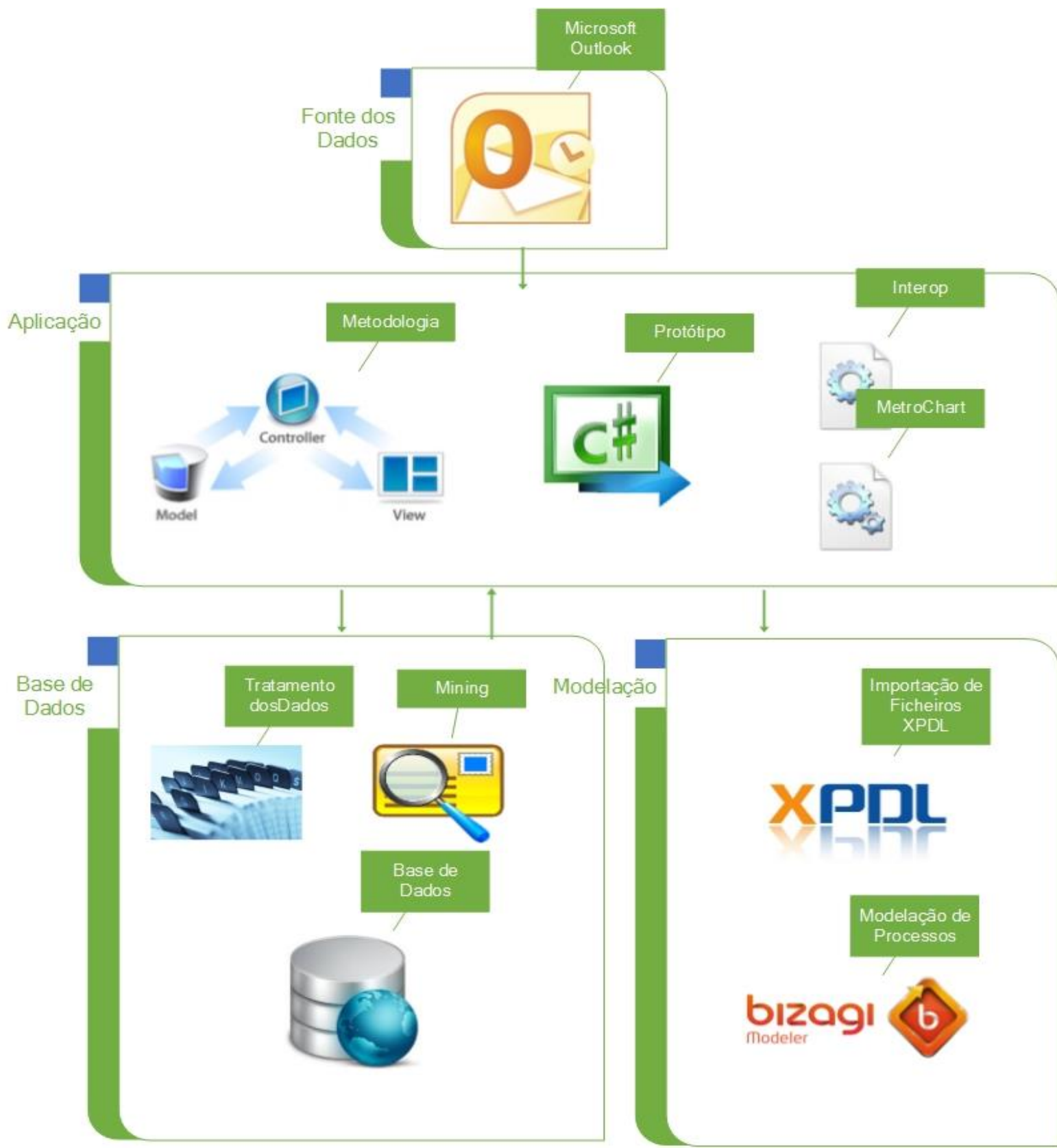


Figura 24 - Arquitetura aplicacional

Pré-Requisitos Aplicacionais

Este protótipo foi desenvolvido tendo por base diversas tecnologias *Microsoft*, pois foi desenvolvido através de *wpf c#*, assente na ferramenta *Microsoft Visual Studio 2013*, e utilizou como fonte de dados o *Microsoft Outlook 2013*.

Este não apresenta a necessidade de ser instalado, pois pode ser única e exclusivamente iniciado através de duplo clique no seu ficheiro executável.

Para que este funcione na plenitude das suas capacidades, o terminal onde é executado tem que satisfazer alguns critérios. Estes prendem-se com a necessidade da data do sistema estar formatada sob o tipo dos Estados Unidos da América, a pasta onde está contida a solução bem como o executável do protótipo devem de estar numa pasta denominada “Dissertação” na localização “C:\”, e por fim, deve ter instalada a *Framework .net 4.0* ou superior.

Modelo de Utilização

Tendo procedido à validação dos pré-requisitos explicitados no capítulo anterior, estão então reunidas todas as condições para poder ser utilizado o protótipo. De seguida será apresentado um conjunto de exemplos ilustrados com imagens, respeitante à forma como deve de ser utilizado o protótipo.

Carregamento e Tratamento da Informação

A figura a baixo apresentada, demonstra o ecrã inicial com que o utilizador se depara quando inicia a aplicação.

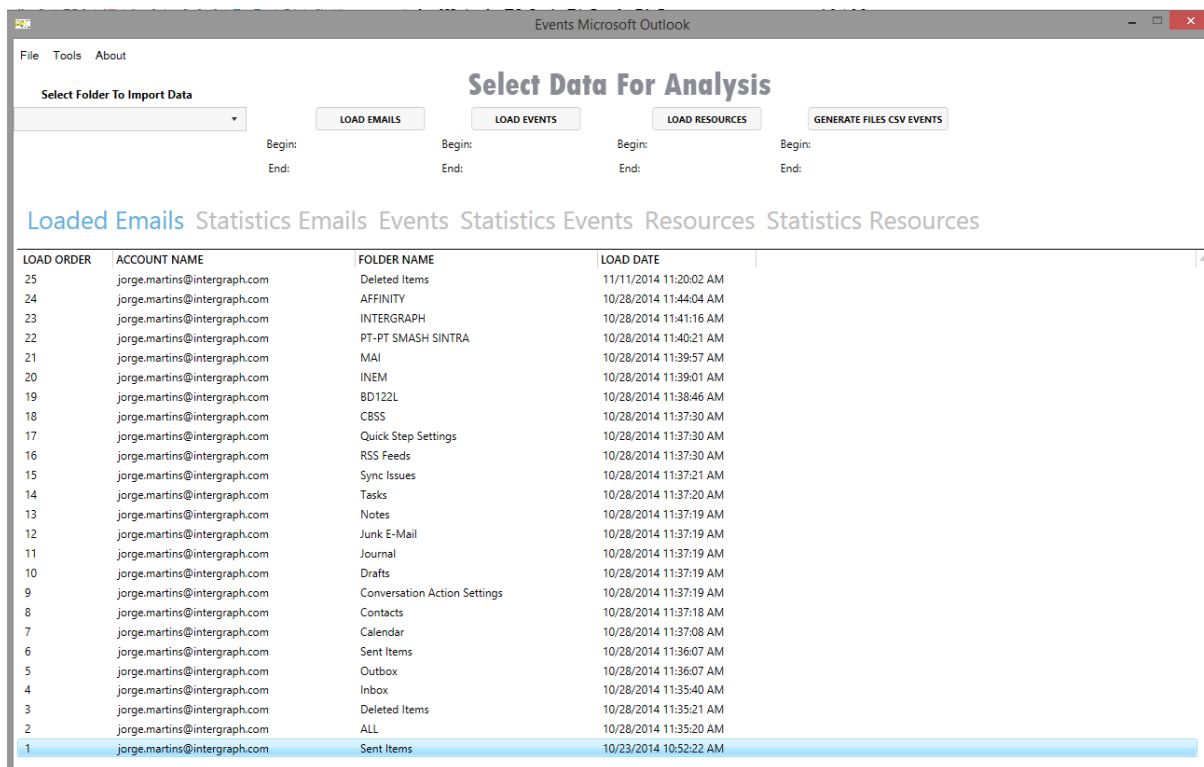


Figura 25 - Ecrã Inicial

Na figura 23, podemos vislumbrar todas as opções existentes na fase inicial do protótipo. Neste, é onde existe a possibilidade de carregar todos os itens que estão alocados no *Microsoft Outlook*, transformar os mesmos em eventos, descarregar os recursos existentes e por fim exportar tanto os eventos, como os recursos sob o formato csv, por forma a efetuar posteriores validações.

Para efetuar este carregamento, torna-se necessário selecionar a pasta que se pretende mapear, ou seja, selecionar a pasta do *Microsoft Outlook* cujos itens irão ser analisados. Contudo, poderá se escolhida a opção “ALL”, que permite descarregar todas as pastas existentes na conta *Microsoft* existente no terminal onde o protótipo de encontra em execução. Podemos verificar estas opções na figura 24.

Após a seleção de uma pasta a ser descarregada o utilizador deverá premir o botão “Load Emails” de forma a dar início ao carregamento.

Como podemos vislumbrar na figura 25, tendo sido concluído com sucesso o carregamento, será facultada ao utilizador o tempo inicial e o tempo final deste carregamento, e será acrescentada uma nova linha na tabela já existente.

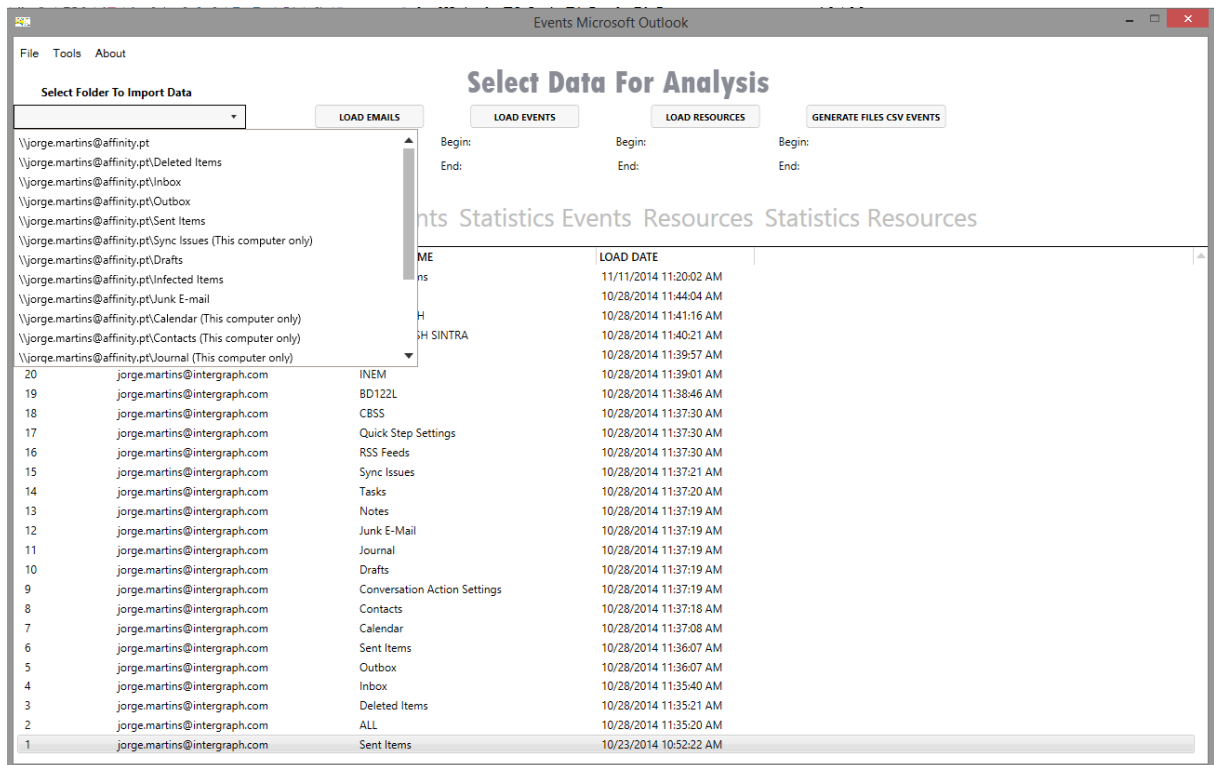


Figura 26 - Seleção da pasta a descarregar para a aplicação

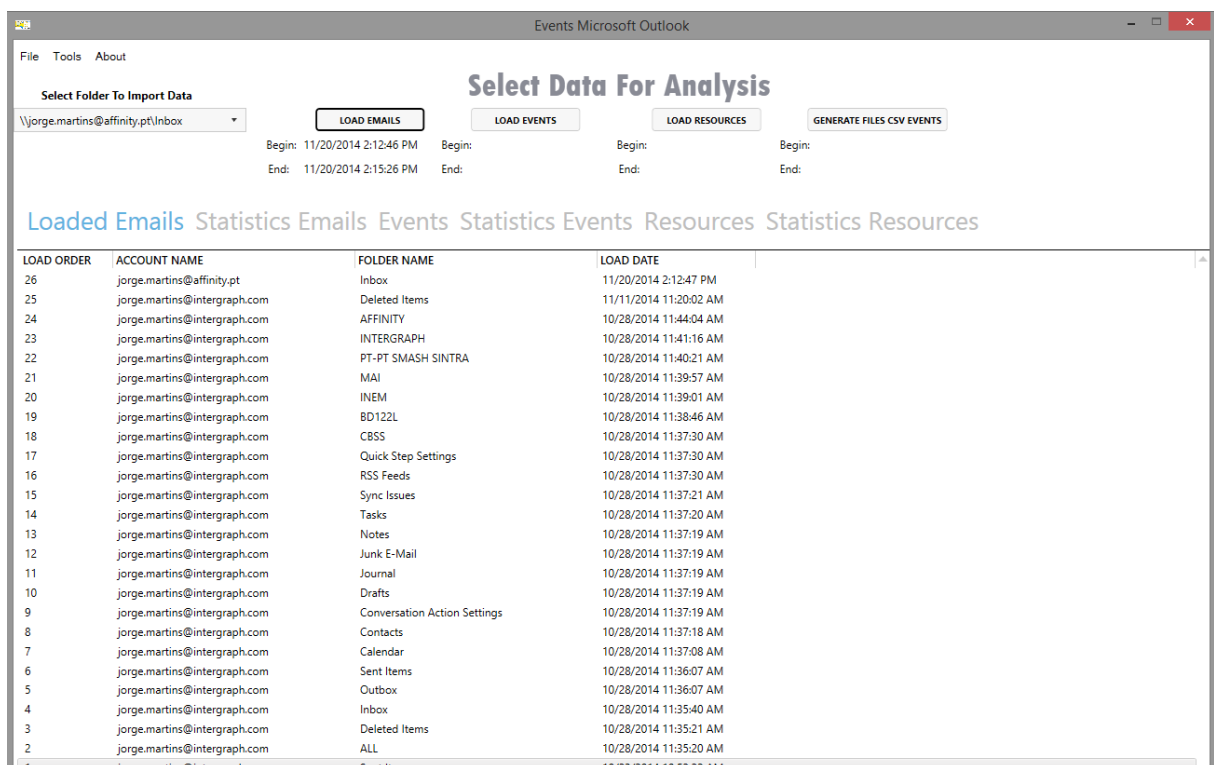


Figura 27 - Carregamento da informação selecionada efetuado com sucesso

Com o término do carregamento requerido o utilizador poderá observar as estatísticas referentes ao mesmo, sendo que esta informação poderá ser consultada premindo a linha correspondente na tabela responsável por guardar todos os carregamentos, ou premindo a segunda posição dos tabuladores existentes no ecrã, denominada “Statistics Email”.

Uma vez efetuada uma destas ações o utilizador será automaticamente enviado para o ecrã apresentado na figura 26. Onde poderá vislumbrar informações como o tempo de carregamento da informação na aplicação, o número de novos registos inseridos bem como o número de registos duplicados e por fim os tipos de itens que foram descarregados e conseqüentemente inseridos na aplicação.

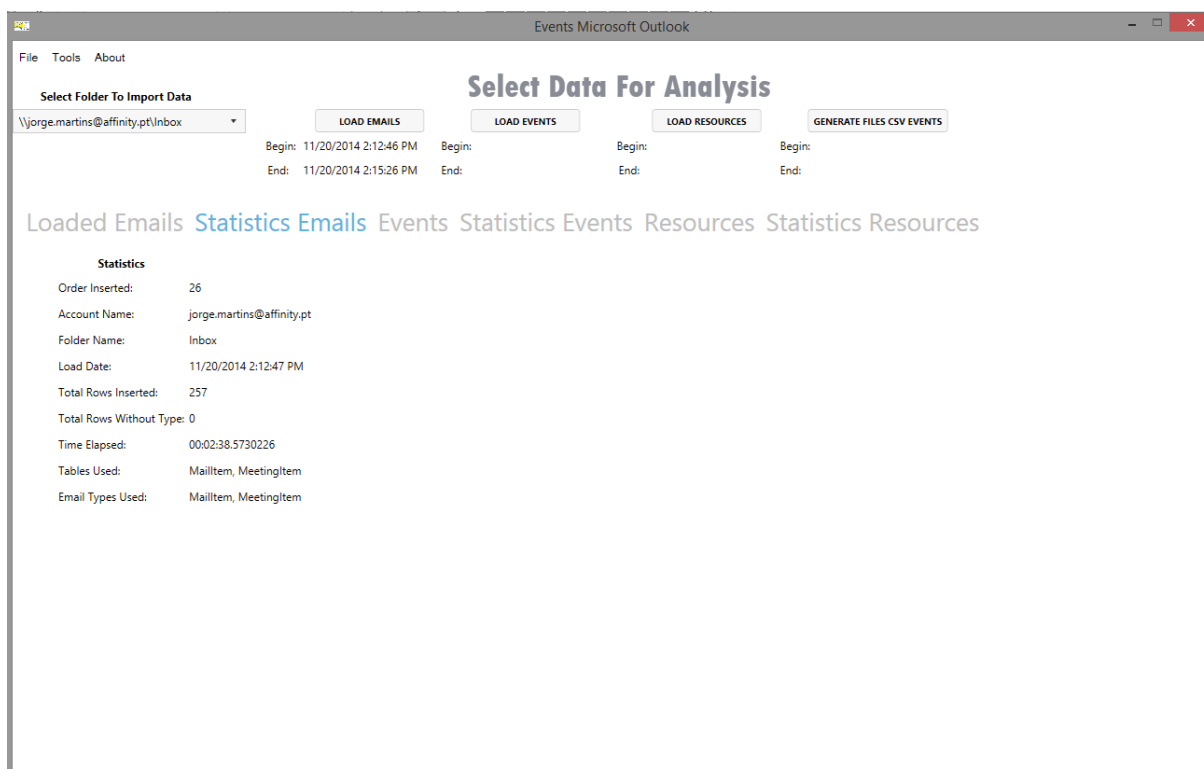


Figura 28 - Estatísticas relativas ao carregamento de itens

Depois de carregada a informação de forma correta para a aplicação torna-se possível criar os eventos relativos à informação que foi recolhida previamente. Para dar início a essa criação basta premir o botão “Load Events” e será de imediato despojado o mecanismo que permitirá transformar todos os itens existentes na aplicação em eventos.

Podemos então ver esse processo depois de terminado na figura 27, apresentada em baixo. Os eventos criados serão então demonstrados ao utilizador na tabela destinada para o efeito.

ID	EMAIL TYPE	CREATION DATE	SENT ON DATE	SEQUENCE NUMBER	FROM	TO	SUBJECT
341	Outlook.MailItem	5/13/2014 4:02:09	5/13/2014 4:02:09	357	Patronilo, Nuno	Martins, Jorge	Access to the bu-pss folder
342	Outlook.MailItem	5/13/2014 4:02:09	5/13/2014 4:02:09	357	Patronilo, Nuno	Salido, Fernando	Access to the bu-pss folder
343	Outlook.MailItem	5/13/2014 5:11:39	5/13/2014 5:11:38	9	Patronilo, Nuno	Martins, Jorge	
344	Outlook.MailItem	5/13/2014 6:25:26	5/13/2014 6:25:25	249	INEM	D-PT-P-INEM	Reenc. Introdução coordenadas GPS
345	Outlook.MailItem	5/13/2014 8:42:19	5/13/2014 8:42:18	249	Lourenco, Domingos	D-PT-PSS Everyone	Reenc. Introdução coordenadas GPS
346	Outlook.MailItem	5/14/2014 7:41:02	5/14/2014 7:41:01	249	Costa, Jose	Lourenco, Domingos	Reenc. Introdução coordenadas GPS
347	Outlook.MailItem	5/14/2014 7:41:02	5/14/2014 7:41:01	249	Costa, Jose	D-PT-PSS Everyone	Reenc. Introdução coordenadas GPS
348	Outlook.MailItem	5/14/2014 8:42:47	5/14/2014 8:42:46	358	Costa, Jose	D-PT-Everyone LIS	Casa para alugar
349	Outlook.MailItem	5/14/2014 10:47:51	5/14/2014 10:47:41	359	Johansen, Stig	Johnson, William C (Bill)	European SharePoint and Project Server Upgrade: Phas
350	Outlook.MailItem	5/14/2014 10:47:51	5/14/2014 10:47:41	359	Johansen, Stig	D-EURO-Everyone Europe	European SharePoint and Project Server Upgrade: Phas
351	Outlook.MailItem	5/14/2014 7:15:51	5/14/2014 7:15:50	124	Moreira, Carlos	Patronilo, Nuno	Código SET_KM
352	Outlook.MailItem	5/14/2014 7:15:51	5/14/2014 7:15:50	124	Moreira, Carlos	Martins, Jorge	Código SET_KM
353	Outlook.MailItem	5/15/2014 9:39:56	5/15/2014 9:39:28	135	Manuel Júlio Arrábida	Paulo Correia Lamego	Numeração bombeiros voluntários
354	Outlook.MailItem	5/15/2014 9:39:56	5/15/2014 9:39:28	135	Manuel Júlio Arrábida	Moreira, Carlos	Numeração bombeiros voluntários
355	Outlook.MailItem	5/15/2014 9:39:56	5/15/2014 9:39:28	135	Manuel Júlio Arrábida	D-PT-P-CBSS	Numeração bombeiros voluntários
356	Outlook.MailItem	5/15/2014 10:29:21	5/15/2014 10:29:21	248	INEM	D-PT-P-INEM	HelpDesk – Abertura de novo pedido com o nº 10342
357	Outlook.MailItem	5/15/2014 10:38:41	5/15/2014 10:38:41	125	Patronilo, Nuno	Costa, Jose	Desinstalar cad entities ...
358	Outlook.MailItem	5/15/2014 10:38:41	5/15/2014 10:38:41	125	Patronilo, Nuno	Martins, Jorge	Desinstalar cad entities ...
359	Outlook.MailItem	5/15/2014 11:06:51	5/15/2014 11:06:51	125	Costa, Jose	Patronilo, Nuno	Desinstalar cad entities ...
360	Outlook.MailItem	5/15/2014 11:06:51	5/15/2014 11:06:51	125	Costa, Jose	Martins, Jorge	Desinstalar cad entities ...
361	Outlook.MailItem	5/15/2014 11:07:11	5/15/2014 11:07:11	248	Lourenco, Domingos	Costa, Jose	HelpDesk – Abertura de novo pedido com o nº 10342

Figura 29 - Ecrã de visualização de eventos

À semelhança do que é possível observar em relação ao carregamento dos itens ao nível das estatísticas, é também possível analisar as estatísticas da criação de eventos através da seleção do menu “Statistics Events”.

Podemos observar estas estatísticas na figura 28, onde de um modo geral é visível o número de novos eventos criados e número de eventos que por motivo de já existirem carregados na aplicação não foram gerados de novo.

Depois de termos efetuado a correta criação do conjunto de eventos desejado, é possível agora carregar todos os intervenientes desses mesmos eventos, ou seja, os recursos.

Para efetuar o carregamento dos recursos para aplicação o utilizador terá de premir o botão “Load Resources” e de imediato será despontado o mecanismo responsável por descortinar os recursos existentes nos eventos previamente criados. É possível vislumbrar esses recursos após o carregamento na figura 29.

À semelhança das estatísticas apresentadas após a criação dos eventos, também é possível analisar as mesmas estatísticas, desta feita respeitantes aos recursos através da seleção do menu “Statistics Resources”. Esta informação pode ser visível na figura 30.

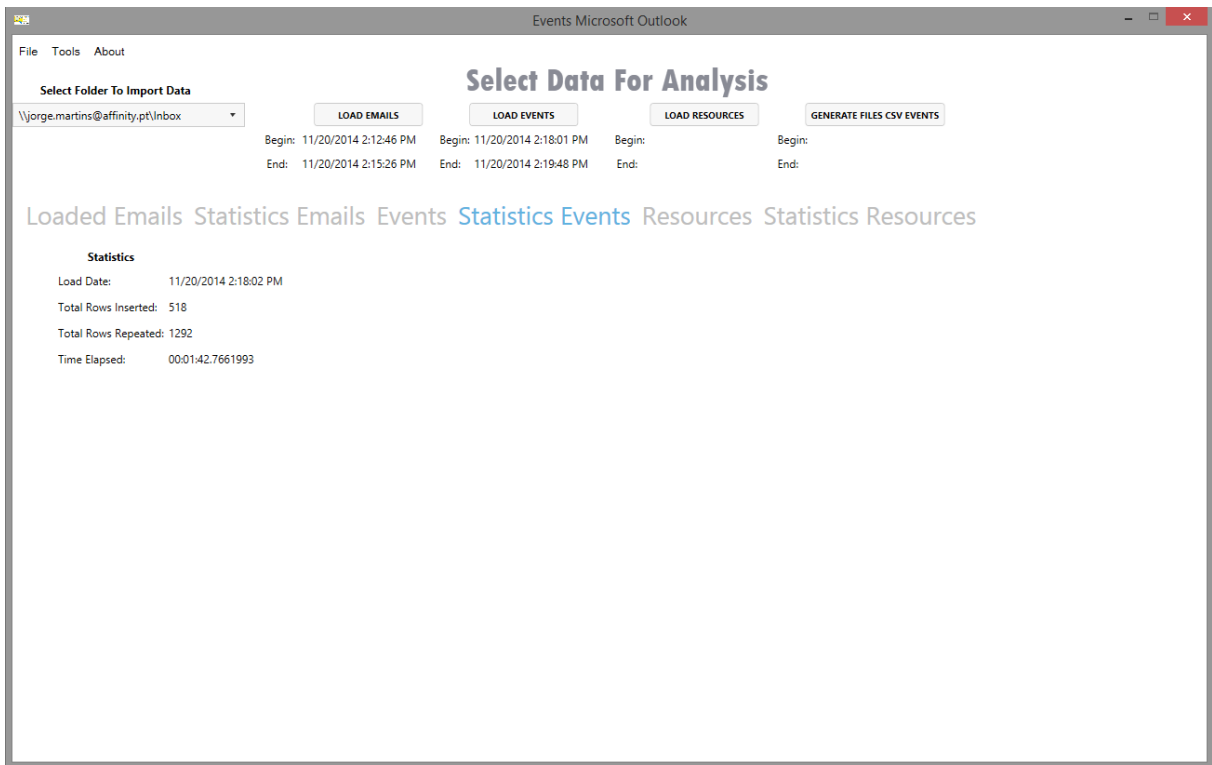


Figura 30 - Ecrã que apresenta as estatísticas relativas aos eventos criados

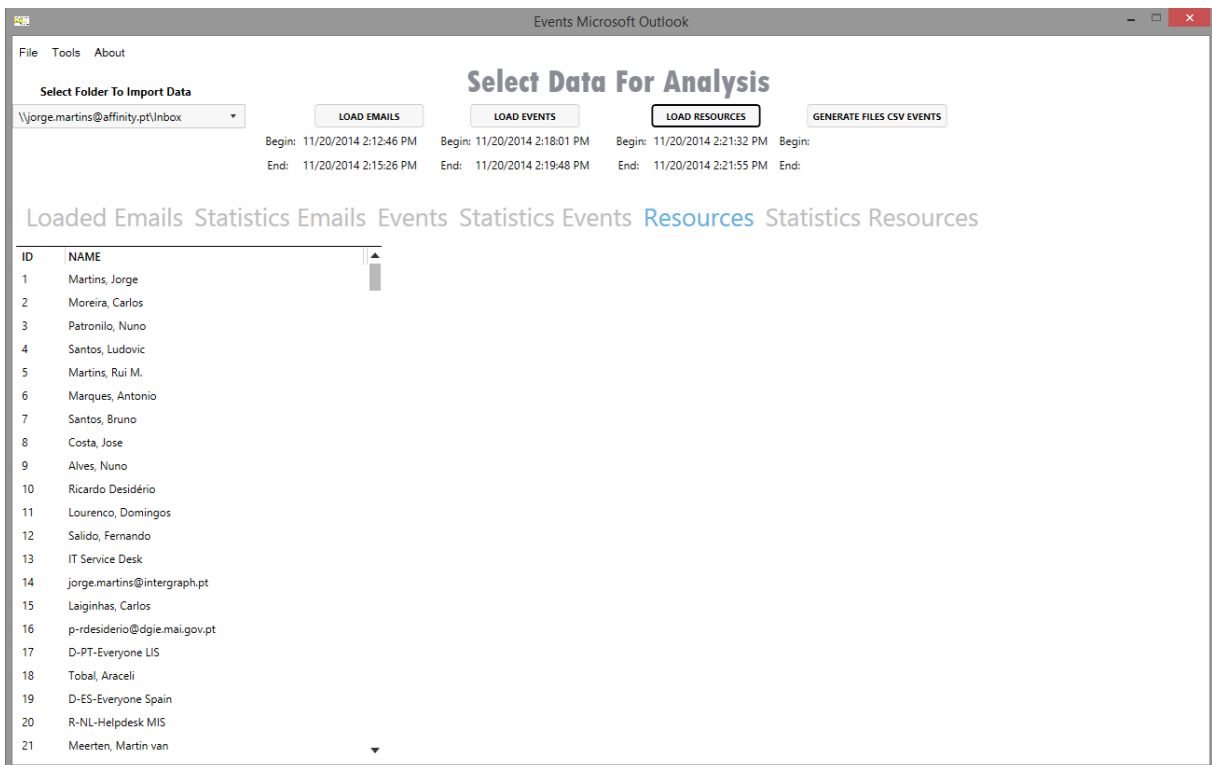


Figura 31 - Ecrã onde é possível observar os recursos previamente carregados tendo em conta os eventos criados

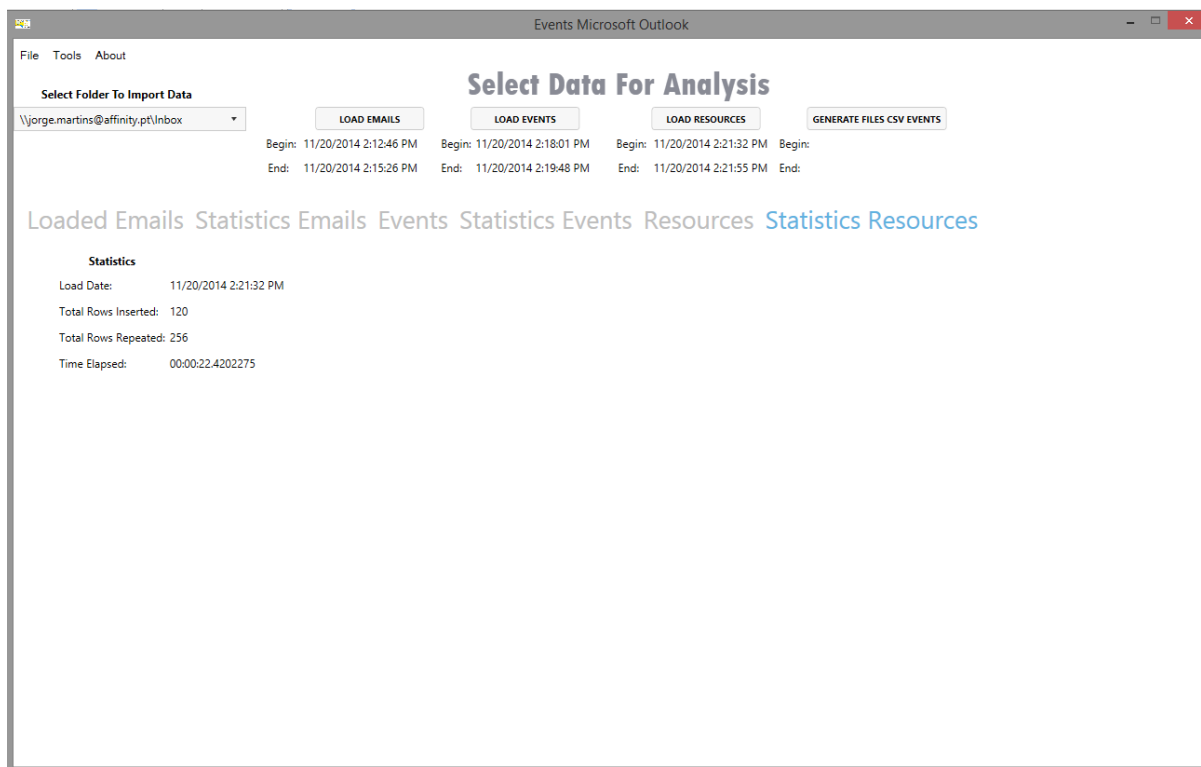


Figura 32 - Ecrã onde são apresentadas as estatísticas relativas aos recursos descortinados

Com a finalidade de poder validar o produto final de forma prática e eficaz é então possível gerar dois ficheiros com o formato csv, um com os eventos e outro com os recursos previamente gerados. Sendo que estes ficheiros terão como localização a pasta onde está colocada a aplicação e na subpasta “Documentos”.

Para desencadear a criação desses dois ficheiros o utilizador terá apenas de premir o botão “Generate Files CSV Events”, e de seguida será desencadeado o processo da sua criação.

Configurações

Depois de terminada toda a recolha e conseqüente tratamento de a informação selecionada torna-se necessário efetuar duas configurações fundamentais na aplicação. Estas podem ser encontradas no meu superior, como está demonstrado na figura 32.

Uma destas configurações pode ser observada na figura 33 e pode ser acedida quando o utilizador prime o meu “Configurations” no menu superior. Esta tem como finalidade selecionar o conjunto de eventos que pretende analisar, ou seja, definir os eventos que se encontram fora do âmbito de análise ou não apresentam relevância para o núcleo do negócio organizacional, e desta forma utilizar apenas os eventos que são importantes.

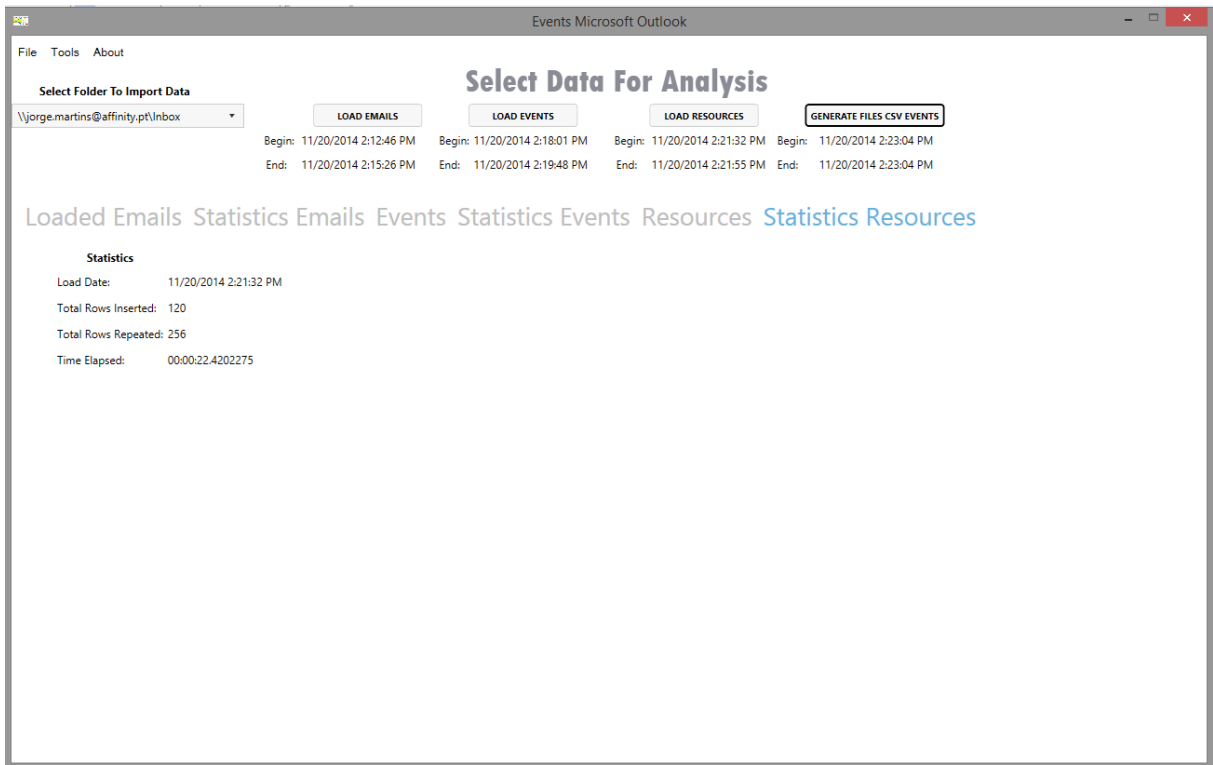


Figura 33 - Criação dos ficheiros de validação

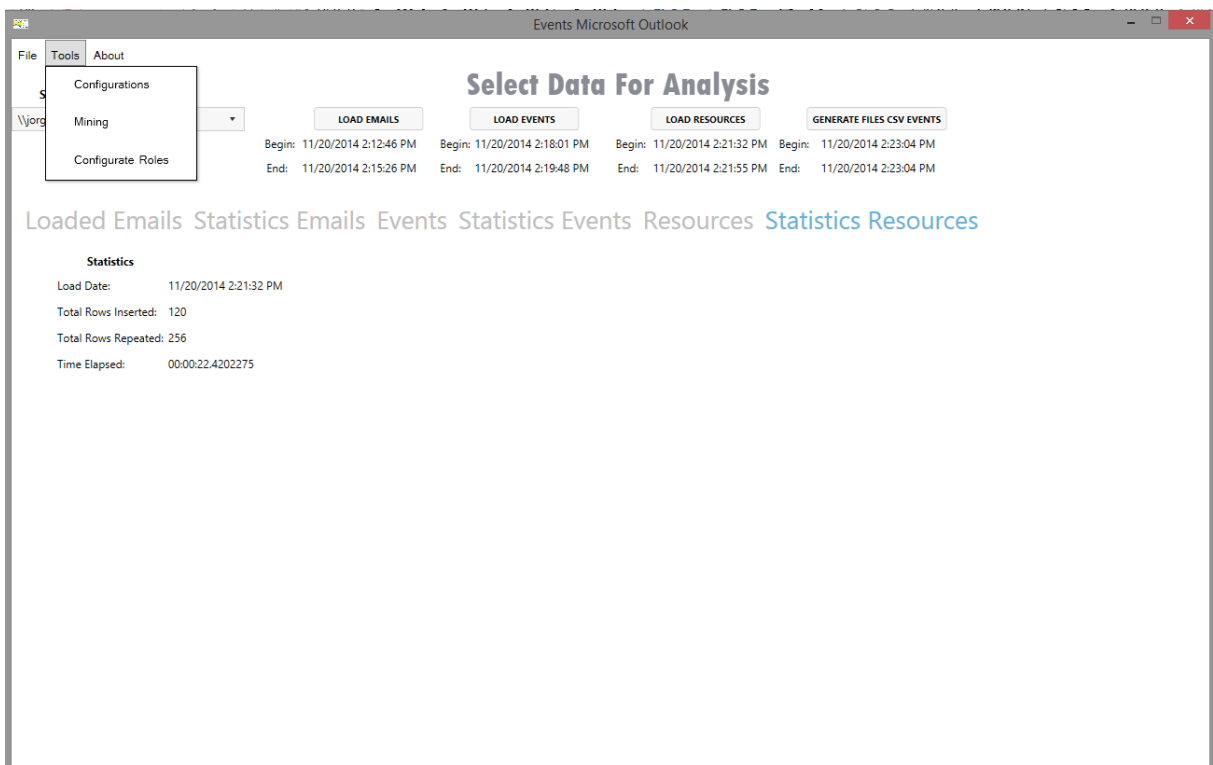


Figura 34 - Apresentação dos Menus de configuração bem como do ecrã utilizado para efetuar a análise dos eventos

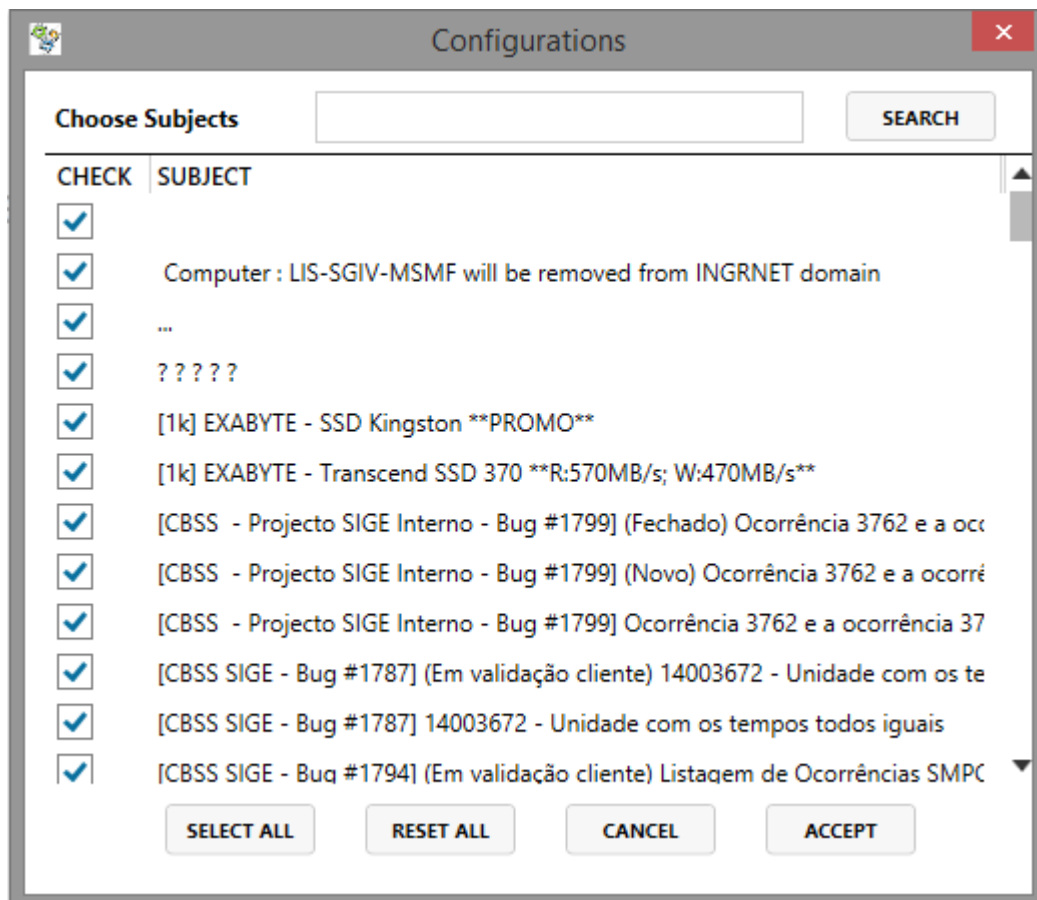


Figura 35 - Ecrã de configuração de eventos a serem analisados

Como podemos vislumbrar na figura apresentada em cima, para seleccionar os eventos pretendidos basta o utilizador seleccionar com um visto a caixa antes do assunto do evento. Para facilitar a procura dos eventos pretendidos o utilizador também poderá efetuar a pesquisa por algumas palavras-chave e dessa forma facilitar a buscar e conseqüente seleção dos eventos com relevância organizacional.

Após efetuada a seleção dos eventos a ser analisados, torna-se necessário seleccionar o papel dos diferentes intervenientes nas atividades da organização. Com vista a esta definição o utilizador poderá aceder no menu superior ao “Configure Roles” e desta forma gerir todas as categorias existentes na organização.

Podemos ver como desenrolar estas ações na figura 34, apresentada em baixo.

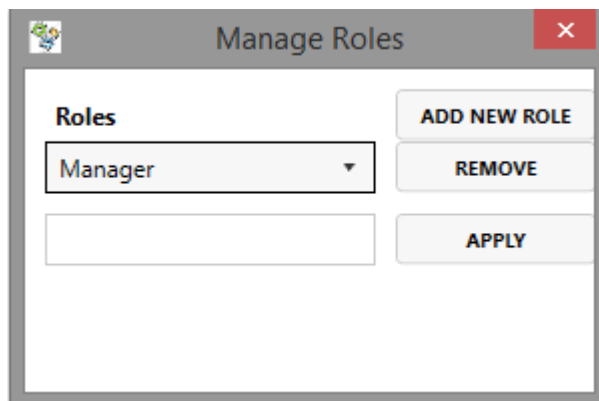


Figura 36 - Ecrã de gestão de *roles*

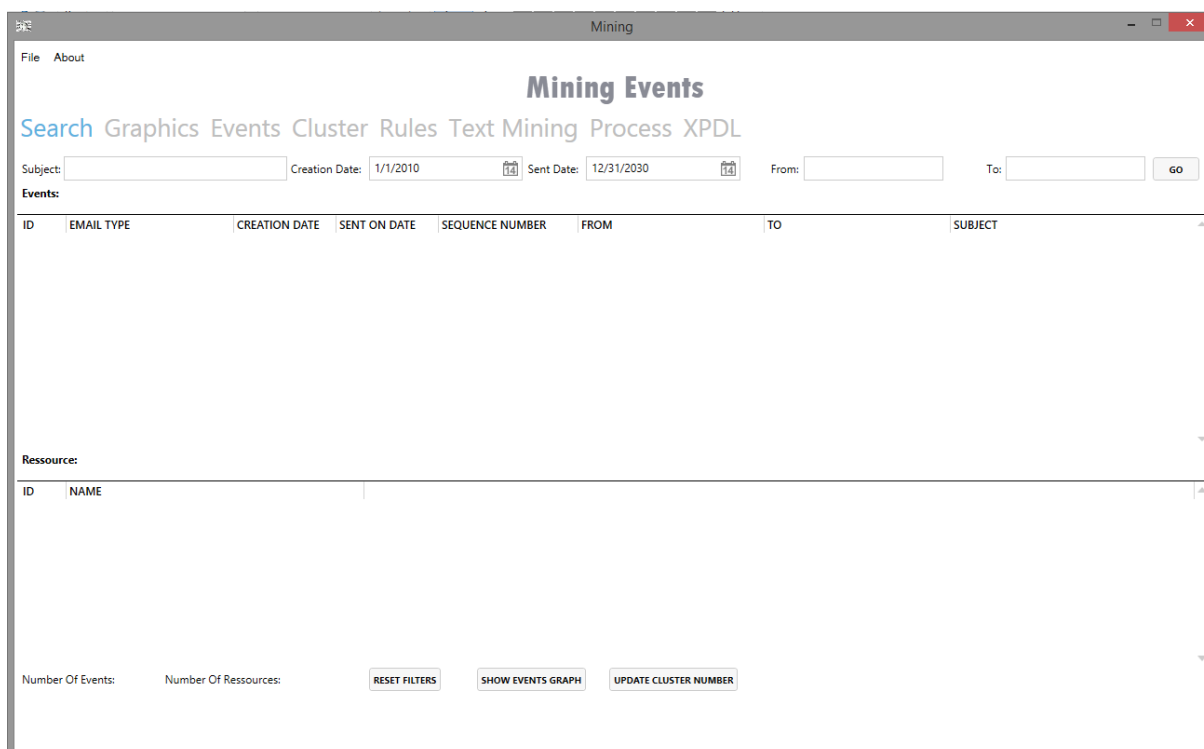


Figura 37 - Ecrã inicial relativo à análise de eventos

Mining

Como podemos vislumbrar na figura 32, existe uma opção denominada “Mining”, e é ao premir esta que o utilizador irá ter um ecrã como o apresentado na figura 35. Este ecrã apresenta todas as funcionalidades relativas à análise a ser desenvolvida aos eventos e recursos previamente carregados para a aplicação.

No primeiro menu apresentado na figura 35, denominado “Search” é possível observar os eventos que pretendemos analisar e no caso de existir alguma inconformidade, ser resolvida através da atualização dos eventos para que estes possam fazer parte do grupo a ser analisado. Esta pesquisa pode ser efetuada tendo em conta qualquer tipo de propriedade existe em cada evento.

Podemos observar na figura 36, um exemplo de uma pesquisa desenvolvida pelo utilizador com vista ao ultimar do grupo de eventos a ser analisado.

The screenshot shows the 'Mining Events' application window. At the top, there is a navigation menu with 'Search' selected. Below the menu, search filters are set: Subject: horas, Creation Date: 1/1/2010, Sent Date: 12/31/2030. A table of events is displayed with columns for ID, EMAIL TYPE, CREATION DATE, SENT ON DATE, SEQUENCE NUMBER, FROM, TO, and SUBJECT. Below the table, a 'Ressource:' section lists various resources like 'Martins, Jorge' and 'Martins, Rui M.'. At the bottom, there are buttons for 'RESET FILTERS', 'SHOW EVENTS GRAPH', and 'UPDATE CLUSTER NUMBER', along with summary statistics: 'Number Of Events: 78' and 'Number Of Ressources: 14'.

ID	EMAIL TYPE	CREATION DATE	SENT ON DATE	SEQUENCE NUMBER	FROM	TO	SUBJECT
9	Outlook.MailItem	5/29/2014 9:07:37	5/29/2014 9:08:42	7	Martins, Jorge	Martins, Rui M.	Mapa de horas Affinity
53	Outlook.MailItem	6/30/2014 12:12:33	6/30/2014 12:14:33	7	Martins, Jorge	Martins, Rui M.	Mapa de horas Affinity
55	Outlook.MailItem	7/1/2014 12:01:38	7/1/2014 12:03:21	7	Martins, Jorge	Martins, Rui M.	Aprovação de horas
57	Outlook.MailItem	7/1/2014 1:33:16 P	7/1/2014 1:35:01 P	7	Martins, Jorge	Martins, Rui M.	Mapa de horas
56	Outlook.MailItem	7/1/2014 12:05:43	7/1/2014 12:05:59	7	Martins, Jorge	Martins, Rui M.	Mapa de horas Affinity
159	Outlook.MailItem	8/29/2014 12:57:10	8/29/2014 12:59:41	7	Martins, Jorge	Martins, Rui M.	Mapa de horas de Agosto
161	Outlook.MailItem	8/29/2014 1:38:43	8/29/2014 1:38:56	7	Martins, Jorge	Martins, Rui M.	Mapa de horas de Agosto
212	Outlook.MailItem	9/30/2014 1:24:26	9/30/2014 1:26:10	7	Martins, Jorge	Martins, Rui M.	Aprovação Mapa de Horas de Setembro
458	Outlook.MailItem	5/28/2014 9:11:00	5/28/2014 9:11:08	7	Martins, Rui M.	Martins, Jorge	Mapa de horas Affinity

ID	NAME
1	Martins, Jorge
5	Martins, Rui M.
259	Inês Belém
260	Management
262	Office
280	AFFINITY IT
258	Jorge Martins

Figura 38 - Ecrã relativo à pesquisa desenvolvida aos eventos a serem analisados

Ao aceder ao menu “Graphics”, o utilizador poderá ter uma perspetiva acerca dos eventos que irão ser analisados sobre a forma de um gráfico de barras. Estes eventos estão separados em quatro grandes grupos de itens, fazendo assim com que o utilizador tenha a possibilidade de selecionar o tipo de itens que deseja vislumbrar. É possível observar estes gráficos na figura 37, sendo que neste exemplo, o tipo de itens selecionado foram os *MailItems*.

Depois de analisados os eventos de forma gráfica, podemos observar na figura 38, os eventos que irão fazer parte do grupo de observação. Contudo, pode ser também vislumbrada a informação referente aos anexos e ao corpo de cada um dos eventos, para tal, o utilizador terá que selecionar através de duplo clique o evento que pretende observar pormenorizadamente e de seguida irá ser apresentado um ecrã como o demonstrado na figura 39.

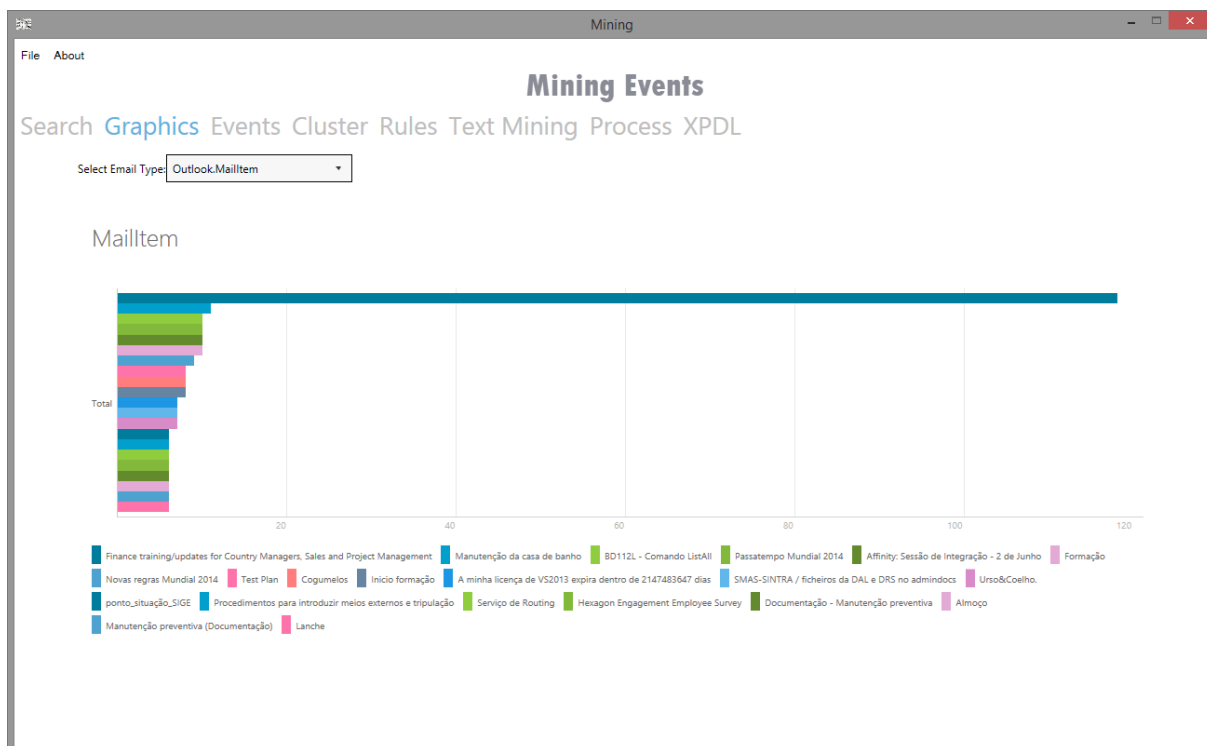


Figura 39 - Representação gráfica dos eventos

The figure shows a table of event details in the 'Mining Events' interface. The table has columns for ID, EMAIL TYPE, CREATION DATE, SENT ON DATE, SEQUENCE NUMBER, FROM, TO, and SUBJECT. The table contains 14 rows of data, with the last row (ID 705) highlighted in blue.

ID	EMAIL TYPE	CREATION DATE	SENT ON DATE	SEQUENCE NUMBER	FROM	TO	SUBJECT
9	Outlook.MailItem	5/29/2014 9:07:37	5/29/2014 9:08:42	7	Martins, Jorge	Martins, Rui M.	Mapa de horas Affinity
53	Outlook.MailItem	6/30/2014 12:12:33	6/30/2014 12:14:33	7	Martins, Jorge	Martins, Rui M.	Mapa de horas Affinity
55	Outlook.MailItem	7/1/2014 12:01:38	7/1/2014 12:03:21	7	Martins, Jorge	Martins, Rui M.	Aprovação de horas
57	Outlook.MailItem	7/1/2014 1:33:16 P	7/1/2014 1:35:01 P	7	Martins, Jorge	Martins, Rui M.	Mapa de horas
56	Outlook.MailItem	7/1/2014 12:05:43	7/1/2014 12:05:59	7	Martins, Jorge	Martins, Rui M.	Mapa de horas Affinity
159	Outlook.MailItem	8/29/2014 12:57:11	8/29/2014 12:59:48	7	Martins, Jorge	Martins, Rui M.	Mapa de horas de Agosto
161	Outlook.MailItem	8/29/2014 1:38:43	8/29/2014 1:38:56	7	Martins, Jorge	Martins, Rui M.	Mapa de horas de Agosto
212	Outlook.MailItem	9/30/2014 1:24:26	9/30/2014 1:26:10	7	Martins, Jorge	Martins, Rui M.	Aprovação Mapa de Horas de Setembro
458	Outlook.MailItem	5/29/2014 9:11:09	5/29/2014 9:11:08	7	Martins, Rui M.	Martins, Jorge	Mapa de horas Affinity
705	Outlook.MailItem	7/1/2014 12:05:01	7/1/2014 12:05:01	7	Martins, Rui M.	Martins, Jorge	Mapa de horas Affinity

Figura 40 - Ecrã relativo aos eventos que irão ser analisados posteriormente

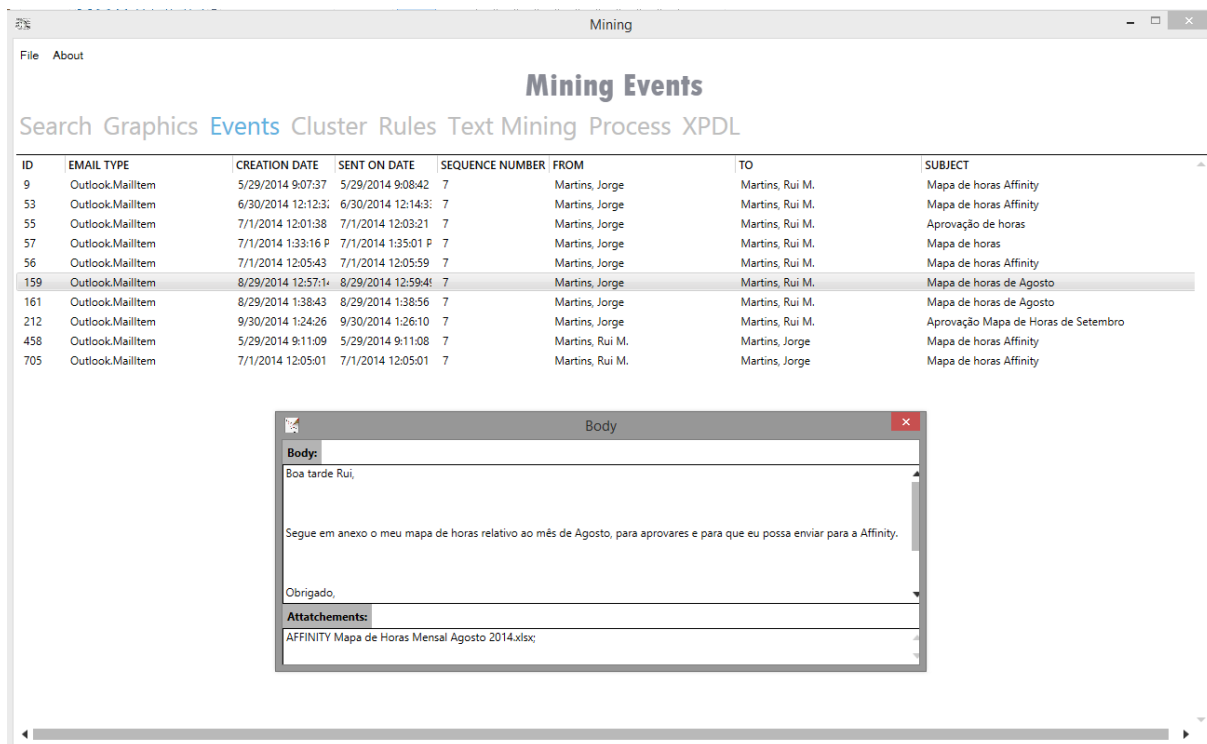


Figura 41 - Ecrã relativo à informação pormenorizada do evento

Após a verificação de todos os eventos a serem sujeitos a análise o utilizador pode seleccionar o separador denominado “Cluster”. Ao seleccionar este separador podemos iniciar o primeiro passo relativo à análise de eventos, pois será onde os eventos serão agrupados tendo em conta as suas características.

Através da figura 40 podemos observar os eventos agrupados depois de efetuada a análise às suas propriedades.

Depois de efetuado o agrupamento dos eventos, podemos aceder ao separador denominado “Rules”, onde irão ser analisadas as relações existentes entre todos os recursos envolvidos nos eventos analisados, sendo esta análise desenvolvida por cada grupo. Para esta análise o utilizador pode definir valores de confiança e suporte diferentes dos que se encontram estabelecidos.

Com a figura 41, podemos observar que foram encontrados três objetos que existem em grande quantidade nos eventos seleccionados, bem como uma relação entre dois desses objetos.

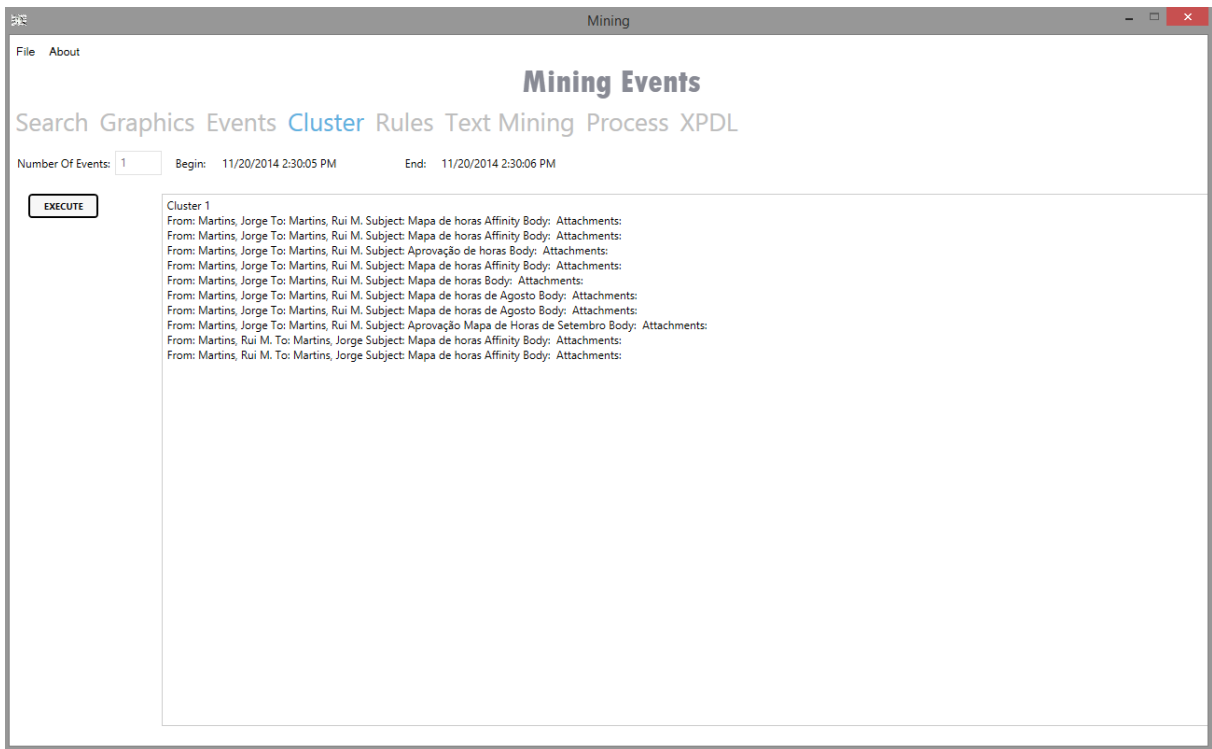


Figura 42 - Ecrã com o agrupamento dos eventos tendo em conta as suas propriedades

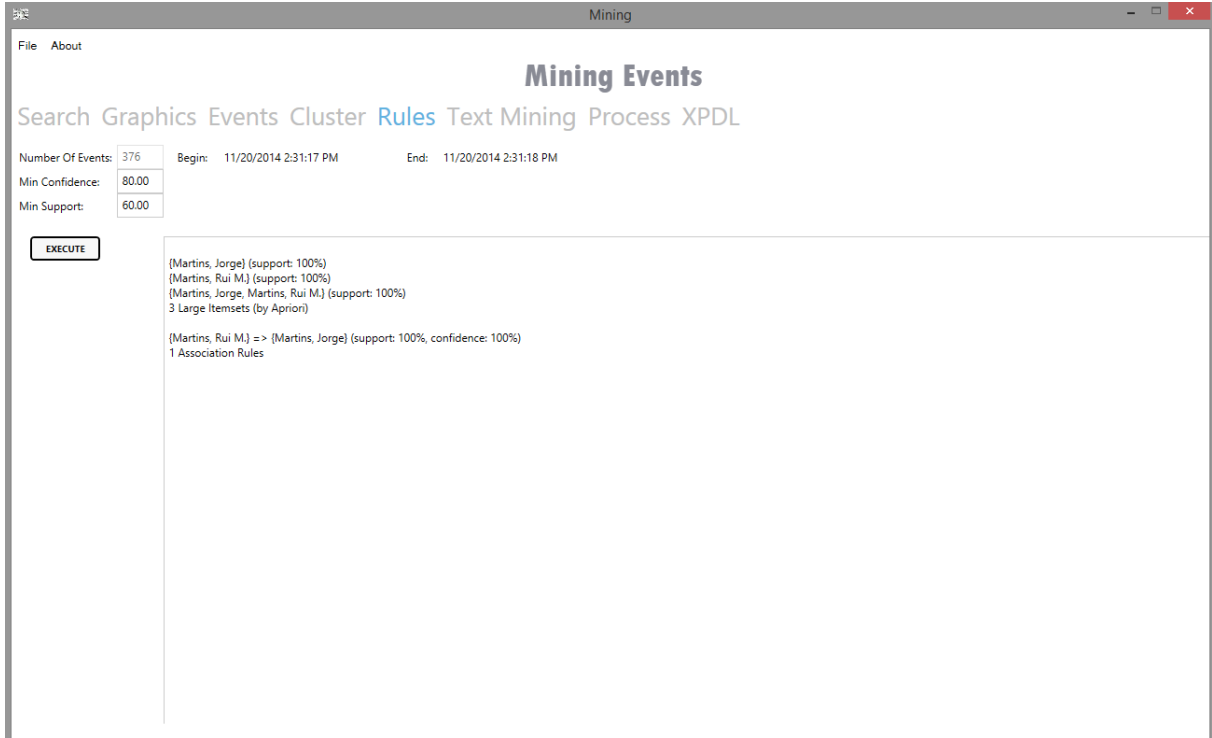


Figura 43 - Ecrã com as regras descartadas relativamente aos recursos envolvidos nos eventos selecionados

Após a análise das relações entre os recursos, o utilizador pode seleccionar o separador “Text Mining” para proceder à análise do corpo e dos anexos do evento. Depois de efetuada esta análise o utilizador poderá vislumbrar as palavras que mais são utilizadas pelos recursos nos eventos seleccionados.

Podemos observar essas *tags* na figura 42.

Tendo efetuada todas as análises necessárias aos eventos seleccionados, o utilizador poderá aceder ao separador denominado “Process” e obter todas as informações necessárias para criar um novo processo.

Essas informações estão apresentadas na figura 43, aqui o utilizador pode seleccionar as que são relevantes para o âmbito requerido. Depois de validadas as informações obtidas o utilizador pode premir o botão “Generate XPDL” e desta forma preparar-se para criar um ficheiro como o formato *xpdl*, que por sua vez dará origem à modelação do processo de negócio em questão.

Contudo, para a criação deste ficheiro é necessário o utilizador ainda validar alguns pontos fundamentais à modelação. Estes, prendem-se com a necessidade de validar se os papéis atribuídos aos recursos estão de acordo com os estipulados a nível organizacional, como podemos observar na figura 44, e se as tarefas atribuídas a cada um dos papéis vão de encontro com as perçecionadas pelo utilizador, podendo observar um exemplo desta seleção na figura 45.

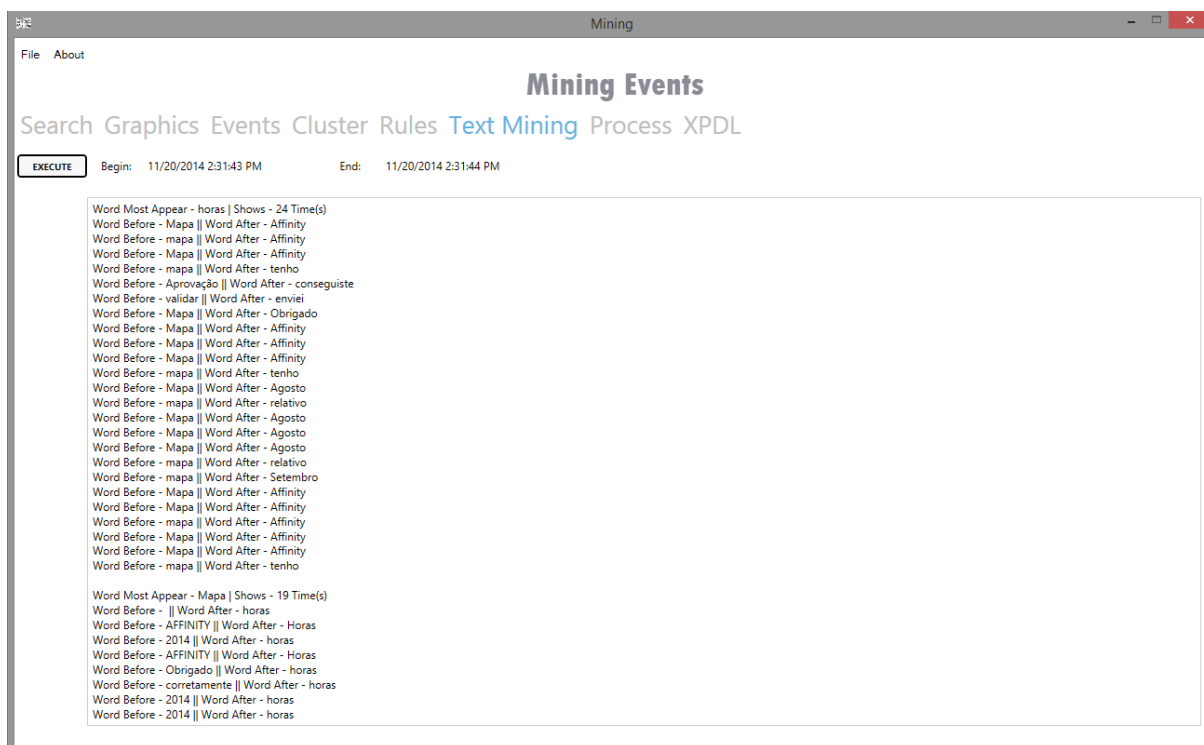


Figura 44 - Ecrã relativo à análise desenvolvida ao corpo e anexo dos eventos seleccionados

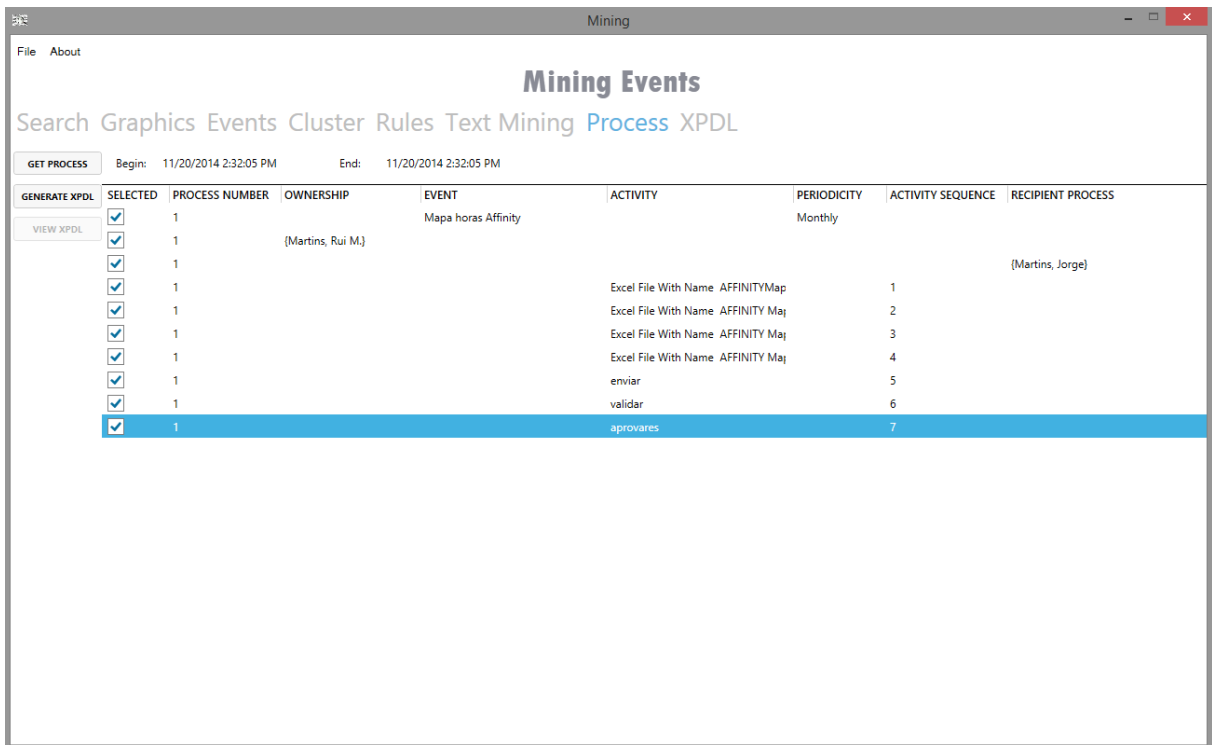


Figura 45 - Ecrã com a informação recolhida para a criação de um processo

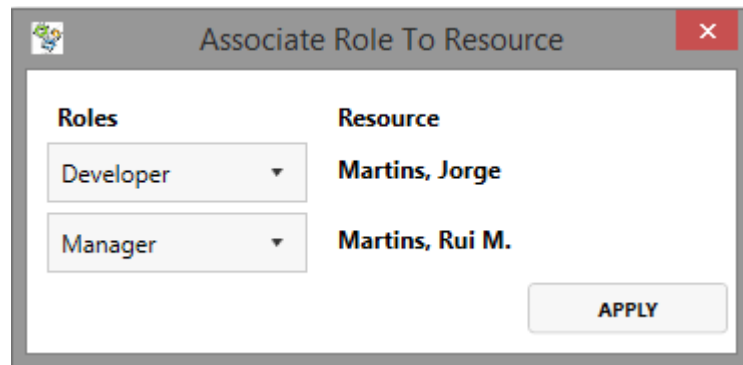


Figura 46 - Ecrã de validação de *roles*

Modelação

Após efetuadas todas as validações requeridas o utilizador poderá premir o botão “View XPDL” e dessa forma ser automaticamente enviado para o separador “XPDL”, onde poderá vislumbrar o conteúdo do ficheiro gerado.

Podemos observar essa situação através da figura 46.

Depois de criado o ficheiro, o utilizador poderá aceder à ferramenta Bizagi e importar o ficheiro previamente criado. Após a sua importação o utilizador terá o processo modelado sob a forma de BPMN. Podemos observar o resultado final, ou seja, a modelação através da figura 47

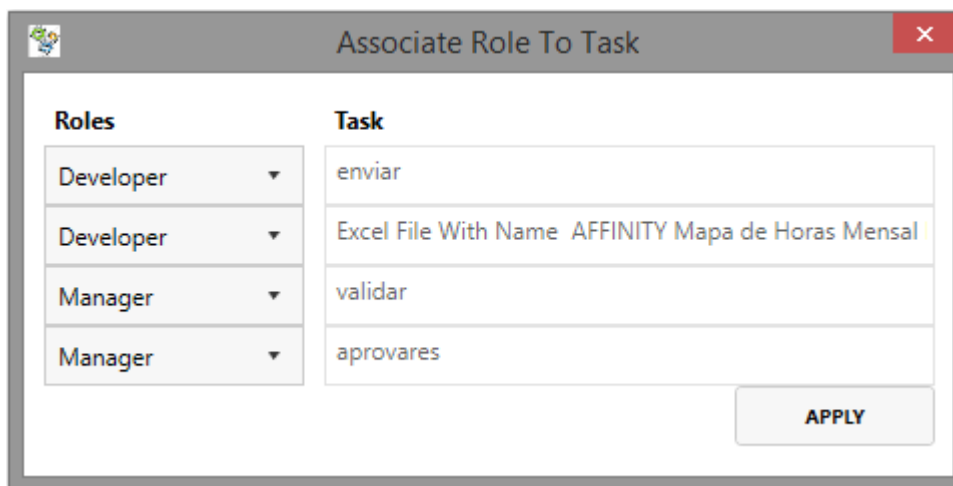


Figura 47 - Ecrã de validação de tarefas associadas a *roles*

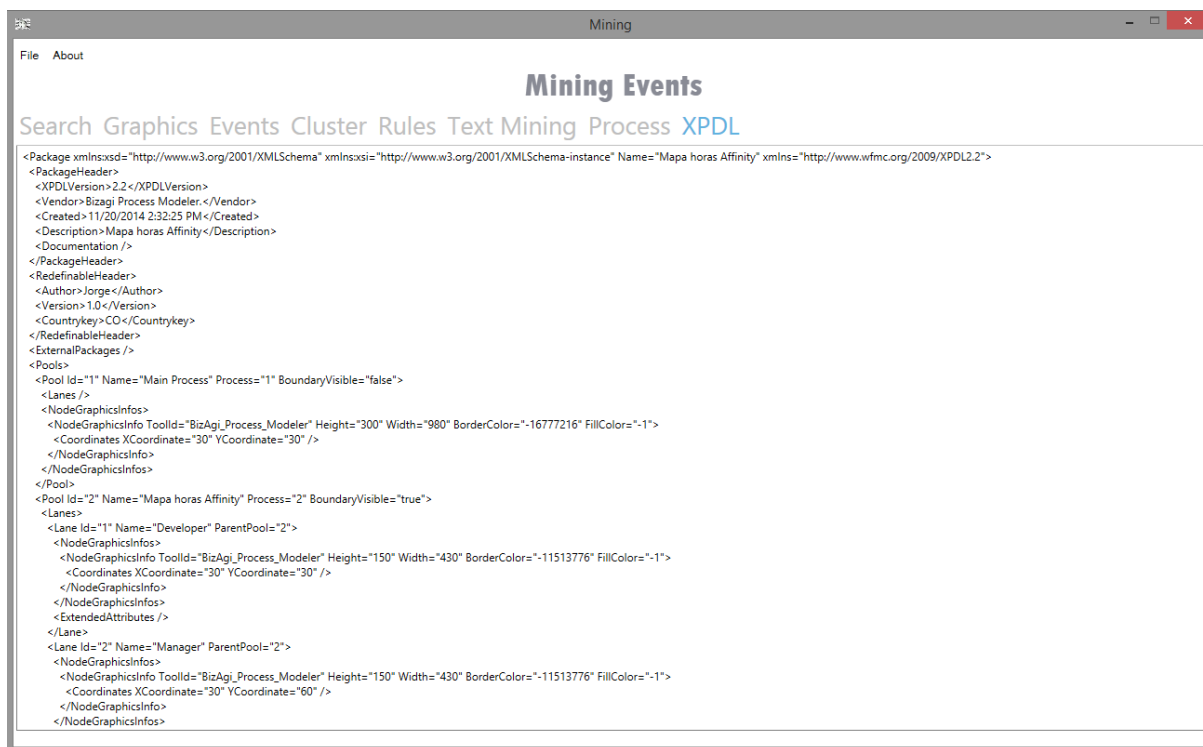


Figura 48 - Ecrã com o conteúdo do ficheiro *XPD*

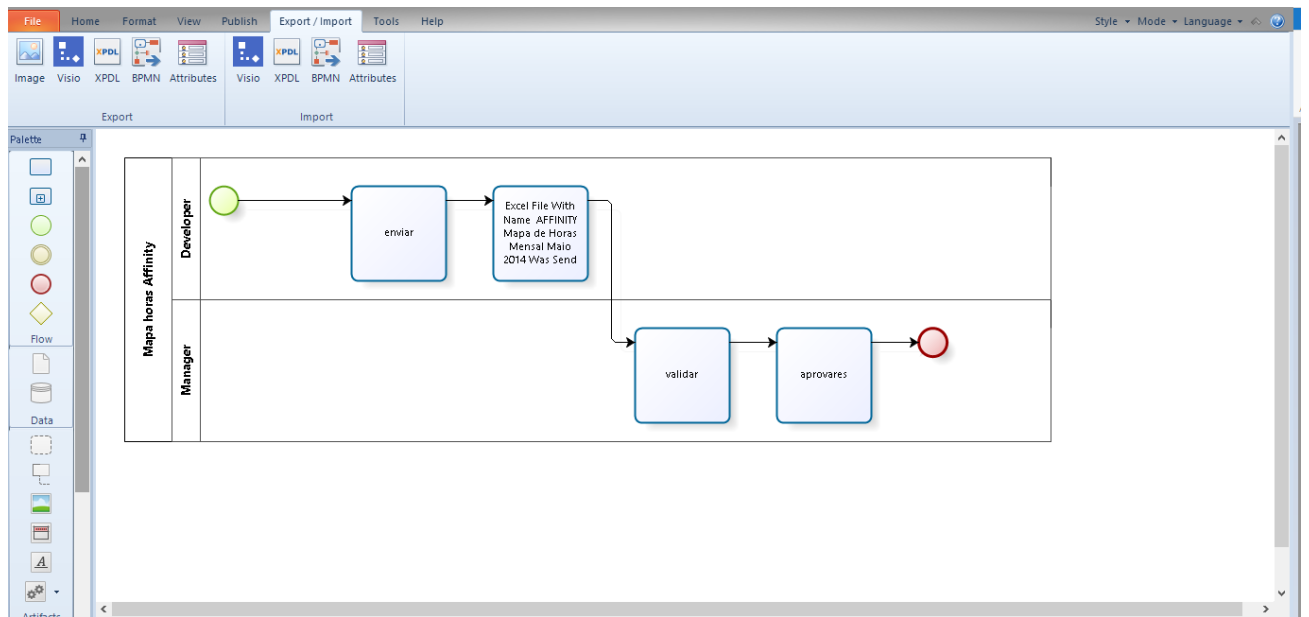


Figura 49 - Modelação do processo descontinuado

Conclusões do Capítulo

Com este capítulo pretendeu-se dar a conhecer a forma de funcionamento do protótipo desenvolvido com vista à demonstração da investigação previamente elaborada, bem como os requisitos fundamentais ao seu correto funcionamento.

Validação

Introdução de Capítulo

No decorrer deste capítulo serão apresentadas duas abordagens distintas com o objetivo de validar os resultados, ou seja, os processos modelados, bem como uma validação desenvolvida tendo por base o exemplo desenvolvido no capítulo da apresentação do protótipo.

Validações

O principal objetivo destas validações prende-se com a necessidade de demonstrar que é possível comprovar os conceitos teóricos apresentados e elaborados no decorrer desta investigação, mais precisamente aqueles que foram demonstrados no capítulo da metodologia, podem ser aplicados numa vertente prática e real, por forma a existir a possibilidade de ser implementada a nível organizacional. Para esta validação será então utilizado o protótipo desenvolvido como forma de evidenciar os resultados esperados.

O protótipo desenvolvido, que é capaz de discernir processos de negócio informais através do Outlook, apresenta uma complexa validação, pois esta tem que ser efetuada tanto por gestores da área de negócio específica de cada atividade como por recursos operacionais, para que os processos de negócio desenvolvidos sejam fiáveis e aplicáveis às realidades do negócio.

Por forma a facilitar as validações por parte dos intervenientes a que estas dizem respeito, existem dois ficheiros csv, um deles com todos os eventos existentes a nível aplicacional e o segundo respeitante a todos os recursos daí provenientes.

Para além da utilização deste mecanismo, o protótipo ainda possui também uma ferramenta que poderá ser útil aos utilizadores que tenham como função validar os processos que foram descortinados.

Este mecanismo encontra-se no separador “Search” apresentado previamente no capítulo do protótipo, onde é possível o utilizador ter acesso a todos os eventos existentes na aplicação e desta forma confrontar o produto final obtido, ou seja, o processo de negócio modelado com todos os eventos que alimentaram a sua criação.

Partindo dos exemplos previamente demonstrados nas figuras 21 e 22, relativas aos ficheiros gerados pela aplicação, podemos validar o processo modelado a título exemplificativo na apresentação do protótipo.

Tendo como ponto de partida o nome do processo gerado, é possível observar que em praticamente todos os campos relativos ao assunto dos eventos selecionados na figura 21 está a expressão selecionada como nome do processo.

No que diz respeito aos papéis atribuídos a cada um dos recursos essa validação pode ser concebida através do cruzamento entre os ficheiros existentes nas figuras 21 e 22, onde podemos confirmar que os dois recursos que interagem nos eventos da figura 21 existem no ficheiro da figura 22, fica então a faltar verificar se os seus papéis a nível organizacional estão de acordo com os que forma definidos na modelação do processo. Para tal, podemos utilizar o menu de configurações do protótipo onde existe a possibilidade de validar e alterar que papel está atribuído a cada recurso. Contudo, ainda existe a necessidade de validar se os dois recursos envolvidos nos eventos são únicos, ou se no caso de não o serem, estão em mais de

80% dos eventos. Podemos então validar que os recursos são os únicos intervenientes dos eventos selecionados, o que faz com que seja válida a modelação gerada.

Em termos das tarefas ou atividades definidas como essenciais para o processo, esta validação terá de ter como base o protótipo, pois é onde o utilizador pode observar a informação detalhada relativa a cada evento. Esta análise pode ser desenvolvida no separador “Events”, pois através deste método o utilizador poderá ter acesso a todas as informações relativas aos eventos.

Utilizando esta abordagem, é visível que as palavras que mais vezes são utilizadas e que correspondem aos critérios definidos como válidas no corpo dos assuntos, são aquelas que se encontram expostas como atividades ou tarefas a serem executadas.

Por fim, no que diz respeito às validações relativas aos papéis atribuídos a cada atividade ou tarefa, estas só podem ser desenvolvidas tendo por base a aplicação, pois antes de ser finalizada a modelação o utilizador tem a oportunidade de alterar ou confirmar o papel que ficará atribuído a cada tarefa. Contudo, e tendo em conta o exemplo utilizado na apresentação do protótipo, podemos observar que as tarefas que foram atribuídas pelo utilizador estão em conformidade com aquelas que estão atribuídas na modelação.

Como foi possível constatar podemos utilizar tanto o protótipo como os ficheiros carregados pelo protótipo para validar os dados que são obtidos como o produto final da análise dos eventos.

Contudo, e apesar de a validação ter sido apresentada através de um caso prático, como aquele que está demonstrado no capítulo do protótipo, esta validação poderia ser aplicada com outro caso prático que fosse selecionado pelo utilizador.

Por forma, a tornar esta abordagem mais consistente, no futuro pretende-se desenvolver esta validação para um maior conjunto de casos práticos, aumentando assim a confiança em relação a esta abordagem.

Com a finalidade de validar um maior número de casos práticos, tem-se como objetivo aplicar esta abordagem a um conjunto de eventos de grande porte, ou seja, utilizar esta abordagem numa secção de uma organização piloto, e validar se com um conjunto de dados de grande porte e de elevada complexidade é possível aplicar esta abordagem e obter resultados credíveis e aplicáveis tendo em conta a realidade vivida na organização em causa.

Para desenvolver este objetivo, torna-se necessário adquirir conhecimentos sólidos na área de negócio da organização em questão, tendo em vista um melhor enquadramento às necessidades da organização.

Depois de adquiridos, será executada a primeira fase desta abordagem, ou seja, a recolha e o tratamento dos dados existentes no correio eletrónico dos utilizadores.

Tendo finalizada a fase de recolha e tratamento destes dados será iniciada a análise aos dados previamente recolhidos, aplicando os algoritmos apresentados no capítulo da abordagem.

Com a finalização da análise aos dados, serão obtidos os resultados. Será então necessário validar os resultados obtidos tendo em conta a realidade vivida na organização.

Depois de validados os resultados obtidos haverá a necessidade de fazer uma análise no seu todo, isto é, contrabalançar os resultados obtidos e os resultados esperados.

A análise no seu todo será imprescindível para autenticar a consistência, a fiabilidade e a confiança da abordagem desenvolvida.

Conclusões do Capítulo

Com este capítulo pretendeu-se dar a conhecer as abordagens passíveis de serem utilizadas com a finalidade de validar os resultados obtidos, bem como apresentar uma demonstração prática da sua utilização.

Conclusões e Trabalho Futuro

Introdução de Capítulo

No decorrer deste capítulo será apresentado o trabalho futuro, bem como as conclusões retiradas de todo o trabalho desenvolvido.

Conclusões

Esta investigação assentou em três pilares fundamentais que foram os processos de negócio, meios e formas de comunicação e as organizações. Estes, foram explorados com o intuito de criar condições para que as organizações se tornem mais autónomas e mais eficientes, por forma a apresentarem resultados mais positivos, não só a nível financeiro, mas também a nível operacional.

Com esta dissertação de mestrado demonstra-se que é possível obter novos ou otimizar processos de negócio organizacionais tendo como base a informação informal que circula através do correio eletrónico interno da empresa.

Através da aplicação da abordagem desenvolvida nesta investigação é possível discernir regras ou padrões de interações entre os diversos utilizadores, podendo estas ser de diversos tipos, bem como estatísticas acerca dos emails que são trocados.

Tendo em conta as informações que são obtidas da análise desenvolvida aos emails que são recolhidos e tratados, torna-se possível gerar e modelar novos processos de negócio nas organizações.

A aplicação da abordagem apresentada nesta dissertação de mestrado permite então que seja possível criar, modelar e validar processos de negócio através de informação informal e não estruturada.

Tomando como ponto de partida o objetivo de capacitar as organizações de ferramentas capazes de melhorarem a sua forma de funcionamento tendo em conta o seu dia-a-dia e o seu conhecimento de causa, torna-se então possível aliar as novas tecnologias de informação aos recursos vivos das empresas.

Trabalho Futuro

No que diz respeito ao trabalho futuro, os principais pontos idealizados prendem-se com a extensão a outras plataformas de comunicação utilizadas pelas organizações, como por exemplos, o *Gmail*, o *Skype* ou ferramentas internas desenvolvidas à medida, mas que apresentem como finalidade a comunicação entre os demais colaboradores.

Outro ponto passível de evolução prende-se com a necessidade de incorporar este tipo de abordagem aos novos modelos de negócio, que se baseiam em grande parte em sistemas móveis, isto é, capacitar esta tecnologia por forma a tornar-se útil em dispositivos móveis, como *smartphones* ou *tablets*, independentemente do sistema operativo utilizado.

Por fim, o último ponto alvo de melhoria futura, seria a inclusão de mais algoritmos de *data mining*, de modo a tornar a análise dos dados mais robusta e mais variada.

Referências

- A.H.M. ter Hofstede, Aalst, W. M. P. van der, M. Adams, and N. Russell. (2010). *Modern Business Process Automation: YAWL and Its Support Environment*. Berlin: Springer.
- A.K.A de Medeiros, A. Guzzo, G. Greco, W.M.P. van der Aalst, A.J.M.M. Weijters, B. van Dongen, and D. Sacca. (2008). Process Mining Based on Clustering: A Quest for Precision. *A. ter Hofstede, B. Benatallah, and H.Y. Paik, editors, BPM 2007 International Workshops (BPI, BPD, CBP, ProHealth, RefMod, Semantics4ws)* (pp. 17–29). Berlin: volume 4928 of Lecture Notes in Computer Science.
- Aalst, W. M. P. van der, H.T. de Beer and B.F. van Dongen. (2005). Process Mining and Verification of Properties: An Approach Based on Temporal Logic. *On the Move to Meaningful Internet Systems 2005: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2005, volume 3760 of Lecture Notes in Computer Science* (pp. 130-147). Berlin: Springer.
- Aalst, W. M. P. van der. (2004). Business process management demystified: a tutorial on models, systems and standards for workflow management. Em J. Desel, W. Reisig, G. Rozenberg, & Eds., *vol. 3098 of Lecture Notes in Computer Science* (pp. 1–65). Berlin, Germany: Springer-Verlag.
- Aalst, W. M. P. van der. (2011). *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Berlin, Germany: Springer-Verlag.
- Aalst, W. M. P. van der. (2012). *Business Process Management: A Comprehensive Survey*. Department of Mathematics and Computer Science, Technische Universiteit Eindhoven, 5612 AZ Eindhoven, The Netherlands: Hindawi Publishing Corporation.
- Aalst, W. M. P. van der. (2012). *Challenges in Business Process Mining*. Eindhoven: Department of Mathematics and Computer Science, Eindhoven University of Technology .
- Aalst, W. M. P. van der. (February de 2012). Process Mining: Overview and Opportunities. *ACM Transactions on Management Information Systems, Vol. 99, No. 99, Article 99*, p. 16.
- Aalst, W. M. P. van der. (2013). Decomposing Petri nets for process mining: A generic Approach. *Science+Business Media*. New York: Springer.
- Aalst, W. M. P. van der. and Schahram Dustdar. (2012). *Process Mining Put Into Context*.

- Adam Smith. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*. Obtido em 10 de Novembro de 2011, de feedBooks: <http://www.feedbooks.com/book/210/an-inquiry-into-the-nature-and-causes-of-the-wealth-of-nations>.
- Alves de Medeiros, A. K., Weijters, A.J.M.M. and Aalst, W. M. P. (2005). Genetic process mining: A basic approach and its challenges. *Business Process Management Workshops (BPM)*. September 5, Nancy, France.
- Alvin Toffler. (1985). *The Adaptive Corporation*. McGraw-Hill.
- Ana Cruz. (2012). *SOCIAL BPM - THE ROLE OF SOCIAL NETWORKS IN BUSINESS*. Lisboa.
- Becker, J.; Kugeler, M.; Rosemann, M.; (2003). *Process Management. A Guide for the Design of Business Processes*. Berlin et al. : Spriger-Verlag.
- Bizagi. (02 de 01 de 2014). *Bizagi*. Obtido em 25 de 02 de 2014, de <http://elearning.bizagi.com>: http://elearning.bizagi.com/file.php/1/ModelingExecution/English/04_What%20is%20a%20process/process.mp4
- bizagi.com. (19 de 11 de 2014). *bizagi.com*. Obtido de www.bizagi.com: <http://www.bizagi.com/>
- Business Dictionary. (2014). *BusinessDictionary.com*. Obtido em 06 de 01 de 2014, de <http://www.businessdictionary.com>: <http://www.businessdictionary.com/definition/business-process-management-BPM.html>
- C. A. Ellis and G. J. Nutt. (1980). Office information systems and computer science. *Computing surveys, volume 1*, pp. 27–60.
- C.W. Günther and Aalst, W. M. P. van der. (2007). Fuzzy Mining: Adaptive Process Simplification Based on Multi-Perspective Metrics. *International Conference on Business Process Management (BPM 2007), volume 4714 of Lecture Notes in Computer Science* (pp. 328–343). Berlin: Springer.
- Chester I. Barnard. (1938). *The Functions of the Executive*. Cambridge: Harvard University Press.
- Christine Alvarado, Jaime Teevan, Mark S. Ackerman and David Karger. (2013). *Surviving the Information Explosion: How People Find Their Electronic Information*. Massachusetts : Massachusetts institute of technology — artificial intelligence laboratory.
- Colette Rolland. (1994). *A Multi-Model View of Process Modelling. Requirements Engineering. Vol 4, Nr 4*. Springer-Verlag.
- D. Harel and R. Marelly. (2003). *Come, Let's Play: Scenario-Based Programming Using LSCs and the Play-Engine*. Berlin: Springer.
- D.Hand, H. Mannila and P.Smyth. (2001). *Principles of Data Mining*. Cambridge, MA: MIT Press.

- Dantas, J. (2001). *Gestão da Inovação*. Vida Económica.
- Davenport, T. (1993). *Process Innovation: Reengineering work through information technology*. Harvard Business School Press: Boston.
- David Sifry's musings. (14 de 04 de 2014). *Sifry*. Obtido de Sifry.com: <http://www.sifry.com/alerts/archives/000493.html>
- E. Alpaydin. (2010). *Introduction to Machine Learning*. Cambridge, MA: MIT Press.
- E.M. Clarke, O. Grumberg, and D.A. Peled. (1999). Model Checking. *MIT Press, Cambridge, MA*.
- Ellis, C. A. (1979). Information control nets: a mathematical model of office information flow. *Proceedings of the Conference on Simulation, Measurement and Modeling of Computer Systems* (pp. 225–240). Colorado, USA: ACM Press, Boulder.
- Estivill-Castro, Vladimir. (20 de June de 2002). Why so many clustering algorithms — A Position Paper. *ACM SIGKDD Explorations Newsletter 4*, pp. 65-75.
- Ferreira, J.M.C.; Neves, J.; Abreu, P. e Caetano, A. (1996). *Psicossociologia das Organizações*. McGraw-Hill.
- G. Greco, A. Guzzo, L. Pontieri, and D. Saccà. (2006). Discovering Expressive Process Models by Clustering Log Traces. *IEEE Transaction on Knowledge and Data Engineering*, 18(8):1010–1027.
- Galbraith, J. (1995). *Designing Organizations*. San Francisco: Jossey-Bass Publishers.
- Gonçalves, J. (2000). As empresas são grandes coleções de processos. *Revista de Administração de Empresas (RAE)*, Sao Paulo, v. 40, n. 1, p. 6–19.
- Graham, M. & Lebaron, M. (1994). *The horizontal revolution*. E.U.A. San Francisco: Jossey-Bass .
- H. Mannila, H. Toivonen, and A.I. Verkamo. (1997). Discovery of Frequent Episodes in Event Sequences. *Data Mining and Knowledge Discovery*, (pp. 259–289).
- H.M.W. Verbeek, T. Basten, and Aalst, W. M. P. van der. (2001). Diagnosing Workflow Processes Using Woflan. *Computer Journal*, 44(4), 246–279.
- Hammer, M. & Champy, J. (1994). *Reengineering the corporation*. New York: HarperBusiness.
- Handry, C. (1990). *Inside Organisations*.
- Hannan, M.T. and J. Freeman. (1989). *Organizational Ecology* . Cambridge: MA: Harvard University Press.

- Harmon, P. (2007). *Business Process Change: A Guide for Business Managers and BPM and Six Sigma Professionals, second edition*. Morgan Kaufmann.
- Henri Fayol. (1930). *Industrial and General Administration*. London: Sir Isaac Pitman & Sons.
- Henri Fayol. (1949). *General and Industrial Management*. London: Sir Isaac Pitman & Sons.
- Henry J. Johansson et al. . (1993). *Business Process Reengineering: BreakPoint Strategies for Market Dominance*. John Wiley & Sons.
- Hezberg, F. (1959). *The Motivation to Work*. New York: John Wiley and Sons.
- Hverbeek. (12 de 11 de 2104). *ProcessMining*. Obtido de ProcessMining.org: <http://www.processmining.org/prom/start>
- I.H. Witten and E. Frank. (2005). *Data Mining: Practical Machine Learning Tools and Techniques (Second Edition)*. San Mateo, CA: Morgan Kaufmann.
- Infopedia. (2003-2014). Obtido em 24 de 02 de 2014, de <http://www.infopedia.pt>: [http://www.infopedia.pt/\\$setores-da-atividade-economica](http://www.infopedia.pt/$setores-da-atividade-economica)
- Jbuijs. (12 de 11 de 2014). *ProcessMining*. Obtido de ProcessMining.org: <http://www.processmining.org/xesame/start>
- José António Sena Pereira. (2013). *Processos de Negócio e Organizações*. Setúbal: Escola Superior de Ciências Empresariais de Setúbal, IPS.
- Kast, E. & Rosenzweig, E. & Johnson, A. (1967). *The Theory and Management of Systems, second edition*. New York: Mcgraw - Hill.
- Lorsch, J. and Lawrence, P. (1967). *Organization and Environment*.
- Ludwig von Bertalanffy. (1989). *A Pioneer of General Systems Theory*. T.E. Weckowicz.
- M. Bramer. (2007). *Principles of Data Mining*. Berlin: Springer.
- M. Dumas, Aalst, W. M. P. van der. and A.H.M. ter Hofstede. (2005). *Process-Aware Information Systems: Bridging People and Software through Process Technology*. New York: Wiley.
- M. Weske. (2007). *Business Process Management: Concepts, Languages, Architectures*. Berlim, Germany: Springer-Verlag.
- March, J. & Simon, H. (1958). *A Behavioral Theory of Decision Making*. Personnel Administration.
- Maslow, A. (1943). A Theory of Human Motivation. *Psychological Review*, Vol.50, pp. 370–396.
- McGregor, D. (1960). *The Human Side of Enterprise*.

- Mintzberg, H. (1994). *The Rise and Fall of Strategic Planning: Reconceiving the Roles for Planning, Plans, Planners*.
- Mole, J. (2003). *Mind Your Manners: Managing business cultures in Europe. third edition*. London: Nicholas Brealey Publishing.
- Munstermann, B., Eckhardt, A., & Weitzel, T. (2010). The performance impact of business process standardization: An empirical evaluation of the recruitment process. *Business Process Management Journal*, 29 - 56.
- N. Russell, W. M. P. van der Aalst, A. H. M. ter Hofstede and D. Edmond. (2005). Workflow resource patterns: identification, representation and tool support. *Proceedings of the 17th International Conference on Advanced Information Systems Engineering*, (pp. 216–232).
- NetCraft Ltd 2014. (14 de 04 de 2014). *netcraft*. Obtido de news.netcraft.com: <http://news.netcraft.com/archives/category/web-server-survey/>
- O'Connor, M.K. & Netting, F. E. (2009). *Organization Practice: A Guide to Understanding Human Service Organizations, second edition*. John Wiley & Sons, Inc.: New Jersey.
- P.H. Feiler and W.S. Humphrey. (1993). Software Process Development and Enactment: Concepts and Definitions. *Proc. 2nd Int. Conf. on "Software Process"*.
- Peter Drucker. (1942). *The Future of Industrial Man*.
- Peter Drucker. (2001). *Management in the Next Society*. Califórnia.
- Pfeffer, J. and G. R. Salancik . (1978). *The External Control of Organizations: A Resource Dependence Perspective*. New York: NY, Harper and Row.
- Porter, Michael E. (1991). *Estratégia Competitiva*. Rio de Janeiro: Campus.
- Porter, Michael E. (1996). "What is Strategy". Harvard: Harvard Business Review.
- R. Srikant and R. Agrawal. (1996). Mining Sequential Patterns: Generalization and Performance Improvements. *In Proceedings of the 5th International Conference on Extending Database Technology (EDBT '96)*, (pp. 3–17).
- R.J. van Glabbeek and W.P. Weijland. (1996). Branching Time and Abstraction in Bisimulation Semantics. *Journal of the ACM*, 43(3), 555–600.
- R.P.J.C. Bose and W.M.P. van der Aalst. (2009). Context Aware Trace Clustering: Towards Improving Process Mining Results. *H. Liu and Z. Obradovic, editors, Proceedings of the SIAM International Conference on Data Mining (SDM 2009)* (pp. 401–412). Philadelphia, PA: Society for Industrial and Applied Mathematics.

- Rego, A. e Braga, J. (2005). *Ética para engenheiros – Desafiando a Síndrome do Vaivém Challenger*. Lidel.
- Rolland, Colette and Pernici, C. Thanos. (1998). A Comprehensive View of Process Engineering. *Proceedings of the 10th International Conference CAISE'98. B. Lecture Notes in Computer Science 1413*. Springer.
- Sarkar, S. (2007). *Empreendedorismo e Inovação*. Escolar Editora.
- Shewhart, A. (1931). *Economic control of quality of manufactured product*. New York: D. Van Nostrand Company.
- Smith, A. (2006). *A Riqueza das Nações, 4ª edição*. Lisboa: Calouste Gulbenkian.
- Stefanou, C. J. , Sarmaniotis, C. & Stafyla, A. (2003). CRM and customer centric knowledge management: an empirical research. *Business Process Management*, pp. 617 – 634.
- T.M. Mitchell. (1997). *Machine Learning*. New York, NY: McGraw-Hill.
- Taylor, Frederick W. (1911). *Principles of Scientific Management*. New York: Harper & Row.
- Thomas J. Peters and Robert H. Waterman. (1982). *In Search of Excellence*.
- W. M. P. van der Aalst, A. H. M. terHofstede, B. Kiepuszewski & A. P. Barros. (2003). Workflow patterns. *Distributed and Parallel Databases, volume 14*, pp. 5-51.
- Weber, M. (2004). *A ética protestante e o “espírito” do capitalismo*. Antônio Flávio Pierucci (Ed.). São Paulo: Companhia das Letras.
- Weber, M. (2007). *Ciência e política: duas vocações*. São Paulo: Cultrix.
- Weick, K. (1979). *The Social Psychology of Organizing, second edition*. McGraw Hill.
- wfmc.org. (19 de 11 de 2014). *wfmc.org*. Obtido de www.wfmc.org: <http://www.wfmc.org/standards/xpdl>
- Yong Joseph Bakos. (06 de 08 de 2014). *mines.humanoriented*. Obtido de mines.humanoriented.com: http://mines.humanoriented.com/classes/2010/fall/csci568/portfolio_exports/mvoget/cluster/cluster.html
- Z. Manna and A. Pnueli. (1991). *The Temporal Logic of Reactive and Concurrent Systems: Specification*. New York: Springer.
- Zisman, M. D. (1977). *Representation, specification and automation of office procedures [Ph.D. thesis]*. Pennsylvania, USA: University of Pennsylvania, Warton School of Business.

- Zisman, M. D. (1978). Office automation: revolution or evolution. *Sloan Management Review*, volume 19, pp. 1–16.
- Zisman, M. D. (1978). Use of production systems for modeling asynchronous concurrent processes. *Pattern-Directed Inference Systems*, (pp. 53–68).
- ZurMuehlen, M. and Recker, J. (2008). Howmuch language is enough? Theoretical and practical use of the business process modeling notation. *Proceedings of the 20th International Conference on Advanced Information Systems Engineering* (pp. 465-479). Z. Bellahsene and M. L. Leonard.