

Análise da utilização de EHRs usando teoria de grafos em ficheiros de log

Luís Miguel Moreira de Sousa

2010

Mestrado de Informática Médica
Faculdade de Ciências | Faculdade de Medicina
Universidade do Porto

Orientador: Doutor Ricardo Correia

Co-orientador: Doutor Pedro Rodrigues

Agradecimentos

Os meus agradecimentos vão para:

- Os meus orientadores Dr. Ricardo Correia e Dr. Pedro Rodrigues por todo o seu acompanhamento e ajuda ao longo do desenvolvimento da tese;
- Cátia Pereira, Vítor Silva e Pedro Farinha por todas as contribuições com material de trabalho, informações e tempo ocupado a responder às minhas questões;
- Marco Ferraz, Adelaide Martins e Luís Silva que me ajudaram não só com ideias mas também com apoio moral, disponibilizaram os seus recursos e proporcionaram-me o tempo suficiente para completar a tese;
- E finalmente para os meus pais Jacinta de Sousa e Filomeno de Sousa, a quem também dedico esta tese, que apesar de todos os problemas sempre tiveram força para me ajudar em todos os sentidos e sempre que precisei.

Sem qualquer uma destas pessoas esta tese não teria sido possível de completar e por isso o meu muito obrigado a todos.

Sumário

Background: A avaliação de ferramentas informáticas na área da saúde é tão importante quanto a própria existência dessas ferramentas pois só assim se pode confirmar se de facto elas conseguem facilitar o trabalho dos profissionais. Nem sempre existe uma forma simples de fazer esta avaliação sendo muitas vezes complicado encontrar uma forma perceptível de apresentar os dados. No entanto, é habitual as aplicações informáticas gravarem ficheiros (chamados logs) que retratam, entre outros, dados sobre a utilização da aplicação.

Objectivos: O objectivo deste trabalho é desenvolver uma ferramenta que, através da análise dos ficheiros de log de aplicações informáticas usadas na saúde, apresente automaticamente uma representação gráfica e intuitiva dos fluxos de utilização destas aplicações e gere recomendações quanto às possíveis acções a tomar para o seu melhoramento.

Desenho do sistema: Um grafo é um diagrama composto por nós e arcos entre esses nós. Fazendo uma analogia entre os nós do grafo e estados na aplicação, e entre os arcos do grafo e transições na aplicação, podem gerar-se grafos ilustrativos da utilização das aplicações. Gerados esses grafos, eles podem ser estudados usando teoria de grafos de modo a permitir extrair informações sobre a forma como as aplicações são usadas. Para este efeito foi inicialmente definido um conjunto de requisitos que os ficheiros de logs devem ter para serem analisados e convertidos em grafos e posteriormente foi montada uma arquitectura de aplicação favorável ao objectivo definido. Foi desenvolvida uma aplicação de interpretação de logs e transformação para grafos com diversas funcionalidades estando entre as mais importantes a capacidade de apresentar uma representação gráfica do grafo com vários tipos de visualização e também a capacidade de aplicar algoritmos da teoria de grafos para gerar recomendações automáticas.

Avaliação: Foram recolhidos logs de diferentes aplicações informáticas activas na área da saúde e estes foram carregados na aplicação desenvolvida. Foram gerados

vários tipos de grafos ilustrativos da actividade dos utilizadores nas aplicações. Com vista a gerar recomendações usou-se o algoritmo “CLIQUEs” para obter grupos de estados fortemente ligados. Usou-se também um algoritmo “GRAPH MEDIAN” personalizado para encontrar os nós centrais das aplicações. Foi ainda aplicado o algoritmo “INVALID PATH” para determinar caminhos inválidos. Para obtermos uma base de comparação para os nossos resultados um perito no fluxo de informação e um responsável por uma das aplicações foram contactados. Depois de descrevermos ao perito os algoritmos acima apresentados ele deu a sua opinião sobre quais seriam os resultados esperados ao aplicarmos os algoritmos.

Discussão: A aplicação desenvolvida foi analisada em termos das suas funcionalidades e do seu funcionamento. Foi feita a interpretação dos resultados da avaliação comparando a opinião do perito com os resultados obtidos automaticamente pela aplicação desenvolvida de forma a comprovar a relação entre as recomendações emitidas e a realidade das aplicações. Quando esta comparação não foi possível devido a não termos dados de base foi tomada uma abordagem mais crítica, subjectiva e pessoal sobre os resultados emitidos pela aplicação. No final foi expressada a opinião do autor sobre a relevância da aplicação desenvolvida e da abordagem escolhida para responder ao passo da avaliação de aplicações relacionado com a forma como as pessoas utilizam os sistemas.

Abstract

Background: The assessment of tools used in the health area is as important as the very existence of these tools because only then one can confirm if they indeed can facilitate the work of the professionals. There isn't always an easy way to make this assessment and is often difficult to find a perceptible way to present data. However, it is usual for software applications to record files (called logs) that depict, among others, data on application usage.

Objectives: The objective of this work is to develop a tool that, by analyzing log files of software applications used in the health area, provides an automatic and intuitive graphical representation of the use patterns of these applications and generates recommendations for actions that might be taken to improve them.

System Design: A graph is a diagram composed by nodes and arcs between these nodes. By making an analogy between the nodes of the graph and states in an application, and between the arcs of the graph and transitions in an application one can generate graphs to illustrate the usage of applications. Once these graphs are generated, they can be studied using graph theory to extract information on how applications are used. For this purpose initially we defined a set of requirements that the log files must have in order to be analyzed and converted into graphs and later we mounted a favorable application architecture for implementing the defined objectives. We developed an application that is able to interpret logs files and transform them to graphs with various features, among the most important the ability to display a graphical representation of a graph with various types of visualization and also the ability to apply graph theory algorithms to generate automatic recommendations.

Assessment: Logs from different computer applications active in the health area were collected and loaded into the developed application. Various types of graphs were generated illustrating the activity of the users in the applications. In order to generate recommendations the "CLIQUEs" algorithm was used to obtain groups of strongly connected states. A customized "GRAPH MEDIAN" algorithm was used to find the

central nodes of the applications. It was also applied the "INVALID PATH" algorithm to determine invalid paths. To obtain a basis of comparison for our results an expert in the information flow and a responsible for one of the applications were contacted. After the algorithms presented above were described to the expert he gave his opinion on what the expected results would be once we applied them.

Discussion: The developed application was analyzed in terms of its features and its behavior. The results of the assessment were also interpreted by comparing the expert's opinion with the results obtained automatically by the application in order to find a relationship between the issued recommendations and the reality of the applications. When this comparison was not possible due to the absence of a baseline for the data, a more critical, subjective and personal approach was taken to interpret the results. In the end the author expressed his opinion not only on the application's relevance but also on the chosen approach to find how people use the systems.

Preâmbulo

O autor do documento é licenciado em Engenharia Informática pelo Instituto Superior de Engenharia do Porto.

No ano lectivo de 2008/2009 ingressou no Mestrado de Informática Médica leccionado na Faculdade de Medicina da Universidade do Porto e na Faculdade de Ciências da Universidade do Porto. No decorrer do primeiro ano de estudos escolheu um trabalho da cadeira “Sistemas Informáticos em Saúde I” relacionado com a análise de ficheiros de log e desenho de grafos. Este trabalho foi proposto pelo Prof. Ricardo Cruz-Correia no seguimento do projecto OPTIM financiado pela Fundação para a Ciência e Tecnologia e em desenvolvimento no Centro de Investigação em Tecnologias e Sistemas de Informação em Saúde da Faculdade de Medicina da Universidade do Porto. O objectivo do projecto é a optimização de sistemas de informação para a saúde pela melhoria da interface gráfica do utilizador e da gestão do armazenamento de dados usando para isso técnicas de aprendizagem automática em ficheiros de log com informações sobre os utilizadores. O autor demonstrou interesse em continuar o trabalho desenvolvido para o aplicar num caso prático. Daí surgiu a possibilidade de desenvolver, como trabalho do segundo ano de estudos, uma tese relacionada com esses temas.

Sendo assim foi feita uma proposta para o desenvolvimento de uma aplicação informática que analisasse ficheiros de log e, usando teoria de grafos, analisasse aplicações informáticas usadas na área da saúde. Esta proposta foi aceite pelo Prof. Ricardo Cruz-Correia recomendando o Prof. Pedro Rodrigues como co-orientador.

Índice

Agradecimentos	3
Sumário	4
Abstract	6
Preâmbulo	8
Índice.....	9
Acrónimos.....	11
Índice de figuras.....	12
Índice de tabelas.....	14
Organização da tese	15
1. Introdução	16
1.1. O panorama actual	16
1.2. Dificuldades	18
1.3. Os sistemas informáticos ajudam realmente?	19
2. Objectivo.....	21
3. Estado de arte	22
Métodos de avaliação.....	22
Ficheiros de log.....	24
Teoria de Grafos	28
Ferramentas existentes	30
4. Desenho do sistema.....	32
Definir pré-requisitos dos ficheiros de log.....	32
Conversão de logs para formato único.....	33
Geração de grafos a partir dos logs.....	38
Geração de recomendações a partir dos grafos.....	38
A aplicação – “Viajante”	39
A arquitectura da aplicação “Viajante”.....	39
Funcionalidades da aplicação “Viajante”	41
5. Avaliação	52
Avaliação pela aplicação “Viajante”	52
breast.care	52

intensive.care	62
Avaliações dos peritos e responsáveis	67
6. Discussão	72
A Aplicação “Viajante”	72
As análises	75
breast.care	75
intensive.care	83
Considerações finais	88
7. Trabalho futuro	90
Referências.....	92

Acrónimos

EHR – Electronic Health Record

IIS – Internet Information Services

IP – Internet Protocol

PHP – PHP Hypertext Preprocessor

SAM – Sistema de Apoio aos Médicos

SAPE – Sistema de Apoio à Prática de Enfermagem

SIGLIC – Sistema Informático de Gestão da Lista de Inscritos para Cirurgia

SONHO – Sistema de Gestão de Doentes Hospitalares

Índice de figuras

Ilustração 1 - Exemplo de um grafo direccionado	29
Ilustração 2 – Workflow de alto nível da aplicação “Viajante”	40
Ilustração 3 - Workflow completo da aplicação “Viajante”	41
Ilustração 4 - Exemplo de ficheiro de log com acções simplificadas	41
Ilustração 5 - Grafo com as designações simplificadas originais do log	42
Ilustração 6 - Exemplo de log com as acções simplificadas	42
Ilustração 7 – Exemplo de tabela de tradução	43
Ilustração 8 - Grafo com designações traduzidas pela tabela de tradução.....	43
Ilustração 9 - Upload e escolha de ficheiros de log	44
Ilustração 10 - Grafo conjunto com número de transições	45
Ilustração 11 - Grafo conjunto com demora nas transições	46
Ilustração 12 - Grafo conjunto com níveis de transição	47
Ilustração 13 - Grafo segmentado com número de transições (1º conjunto)	48
Ilustração 14 - Grafo segmentado com número de transições (2º conjunto)	48
Ilustração 15 - Grafo segmentado com demora nas transições (1º conjunto).....	49
Ilustração 16 - Grafo segmentado com demora nas transições (2º conjunto).....	49
Ilustração 17 - Grafo segmentado com níveis de transição (1º conjunto)	50
Ilustração 18 - Grafo segmentado com níveis de transição (2º conjunto)	50
Ilustração 19 - Escala de cor para os grafos cíclicos	51
Ilustração 20 - Escala de cinzentos para grafos em árvore	51
Ilustração 21 - Grafo cíclico completo da aplicação ICU+breast.care	53
Ilustração 22 – Grafo cíclico relativo apenas à aplicação ICU	54
Ilustração 23 - Grafo cíclico relativo apenas à aplicação breast.care	55
Ilustração 24 - Grafo cíclico segmentado da sessão com o id 2071506 da aplicação HSJ.ICU+breast.care	56
Ilustração 25 - Grafo cíclico segmentado da sessão com o id 2120433 da aplicação HSJ.ICU+breast.care	56
Ilustração 26 - Grafo cíclico segmentado da sessão com o id 2072478 da aplicação HSJ.ICU+breast.care	57

Ilustração 27 - Grafo cíclico segmentado da sessão com o id 2122535 da aplicação HSJ.ICU+breast.care	57
Ilustração 28 - Grafo em árvore da aplicação HSJ.ICU com corte no nível 5.....	58
Ilustração 29 - Grafo completo do intensive.care	63
Ilustração 30 – Grafo da aplicação breast.care relativo ao diagnóstico e associado à resposta para o algoritmo dos CLIQUES (Silva, 2010).....	68
Ilustração 31 – Grafo da aplicação breast.care relativo ao tratamento e associado à resposta para o algoritmo dos CLIQUES (Silva, 2010).....	69
Ilustração 33 - Grafo da aplicação breast.care relativo ao diagnóstico e associado à resposta para o algoritmo “CUSTOM GRAPH MEDIAN” (Silva, 2010)	70
Ilustração 32 – Grafo da aplicação breast.care relativo ao tratamento e associado à resposta para o algoritmo “CUSTOM GRAPH MEDIAN” (Silva, 2010)	70
Ilustração 34 - Screenshot da aplicação “Viajante”	72
Ilustração 35 - Grafo do módulo breast.care filtrado a 4 nós.....	78
Ilustração 36 - Vizinhança directa dos nós "Enfermeiros(inicio)" e "Enfermeiros(grava=false)"	86

Índice de tabelas

Tabela 1 - Ferramentas conhecidas de análise de logs existentes e suas funcionalidades	30
Tabela 2 - Tabela descritiva do formato de log do IIS	34
Tabela 3 - Tabela descritiva do formato de log do Apache WebServer (Common Log Format).....	35
Tabela 4 - Tabela descritiva do formato de log do Apache WebServer (Combined Log Format).....	36
Tabela 5 - Tabela descritiva da exportação do formato de log do HSJ.ICU	36
Tabela 6 - Tabela descritiva das operações contidas no log do HSJ.ICU.....	37
Tabela 7 - Tabela descritiva das operações contidas no log do Intensive.care.....	37
Tabela 8 - Algoritmos da teoria de grafos utilizados e a sua utilidade	38
Tabela 9 - Tipos de grafos gerados pelo “Viajante”	44
Tabela 10 - Comparação entre os nós relativos ao breast.care encontrados automaticamente pelo algoritmo “CUSTOM GRAPH MEDIAN” e os referenciados pelo perito	80

Organização da tese

Para dar início a este trabalho são abordados os temas da utilização de sistemas informáticos na área da saúde e da sua avaliação, as dificuldades que os utilizadores podem sentir ao usar estes sistemas e é questionada a utilidade dos sistemas informáticos usados actualmente. De seguida iremos definir especificamente os objectivos do trabalho, estando estes relacionados com o desenvolvimento de uma ferramenta capaz de analisar aplicações informáticas usadas na saúde. Para suportar a metodologia idealizada apresentamos o estado da arte. Este está dividido em quatro partes: métodos de avaliação, ficheiros de log, teoria dos grafos e aplicações de análise existentes, onde em cada uma apresentamos uma definição do que são e estudos que têm vindo a ser feitos utilizando estes métodos. No desenho do sistema vamos definir o processo pelo qual a ferramenta de análise se deverá reger, a sua arquitectura e funcionalidades. Estando este processo completo são apresentadas as avaliações feitas automaticamente pela ferramenta desenvolvida às aplicações escolhidas bem como a visão de peritos e responsáveis por essas aplicações quanto ao que deveriam ser os resultados das análises. Na discussão a própria ferramenta será analisada em termos de funcionalidades e funcionamento e vai ser feita uma interpretação dos grafos e análises por ela geradas. As opiniões dos peritos serão contrapostas a esses resultados de forma a estabelecer ou quebrar a linha de confiança da nossa abordagem. No final é apresentada uma lista de referências utilizadas na tese.

1. Introdução

Para se tomar uma decisão consciente é necessário obter a maior quantidade de informação relevante possível que possamos encontrar sobre o assunto em causa. Por vezes é difícil encontrar esta informação, de outras vezes a informação encontrada é de difícil compreensão e em áreas onde a tomada de decisão que pode salvar uma vida tem que ser tomada rapidamente é imprescindível que a informação que suporta essa tomada de decisão esteja acessível rapidamente e de uma forma simples de compreender.

Os computadores são uma ferramenta de indiscutível auxílio em áreas onde a informação muda rapidamente ou onde a quantidade de dados disponíveis é tão grande que é necessário ser trabalhada para ser apresentada informação útil.

1.1. O panorama actual

Na área da saúde está a ser feita uma transição e conseqüente adaptação às novas tecnologias. Os Electronic Health Record (EHR) são muito conhecidos nesta área. EHR é um conceito em evolução definido como um conjunto longitudinal de informações de saúde em formato electrónico sobre os pacientes individuais e populações (Gunter T, 2005). As vantagens de usar registos clínicos electrónicos em relação ao papel são claras no que toca à acessibilidade, pesquisa e perceptibilidade da informação guardada. No entanto para utilizar um sistema informático é necessário haver um período de adaptação tanto ao sistema em si como muitas vezes ao próprio mundo electrónico. Mesmo após este período ter sido ultrapassado ainda é preciso lidar com possíveis erros da aplicação, indisponibilidades e lentidão do sistema.

Em Portugal o uso de EHRs está generalizado pelos hospitais (Ministério.da.Saúde, 2009). Alguns dos sistemas mais utilizados são o Sistema Integrado de Informação Hospitalar (SONHO), o Sistema de Apoio ao Médico (SAM), o Sistema Informático de Gestão da Lista de Inscritos para Cirurgia (SIGLIC) e o Sistema de Apoio à Prática de Enfermagem (SAPE).

O SONHO (Administração.Central.do.Sistema.de.Saúde, 2010) é um sistema integrado de informação hospitalar que se baseia na filosofia de “um doente – um número único de identificação”. É constituído por oito módulos (integrador, urgência, consulta externa, internamento, bloco operatório, hospital de dia, arquivo e facturação) que visam dar corpo e consistência aos objectivos de carácter estrutural e funcional deste sistema de informação. Do ponto de vista estrutural, o objectivo principal deste sistema é o de criar as infra-estruturas mínimas necessárias para um *Sistema Integrador de Informação Hospitalar*, que permita englobar progressivamente, e à medida das necessidades de cada hospital, novos módulos/aplicações interligadas com as existentes. Pretende-se, ainda, assegurar que todos os critérios de normalização definidos e implementados são assumidos de forma natural pelas novas aplicações. Do ponto de vista funcional, o objectivo principal deste sistema é o de controlar o fluxo de doentes hospitalares. O doente é o centro do modelo funcional. O SONHO começou a ser instalado em 1995 e está actualmente instalado em aproximadamente 90% dos hospitais nacionais do Sistema Nacional de Saúde.

O SAM (Administração.Central.do.Sistema.de.Saúde, 2010) é uma aplicação informática em tecnologia Web que tem como base a Base de Dados do SONHO. Este sistema está orientado para a actividade do médico e permite integrar aplicações clínicas de acordo com as necessidades e prioridades de cada hospital. SAM tem como principal objectivo auxiliar os médicos em algumas das tarefas decorrentes da prática diária nos hospitais.

O SIGLIC (Administração.Central.do.Sistema.de.Saúde, 2010) é um sistema informático que pretende orientar os serviços de prestação de cuidados cirúrgicos para que, de forma articulada, regulada e sustentada, mantenham o equilíbrio entre a procura e a resposta, atendendo às necessidades de acesso dos cidadãos. SIGLIC tem como principais objectivos melhorar a resposta e adequação dos cuidados às necessidades em saúde, tendo em consideração a equidade, qualidade, oportunidade, transparência, eficácia, efectividade e eficiência, garantir o cumprimento dos tempos médios de espera estabelecidos em Portaria, com a garantia de prestação de cuidados adequados e garantir a qualidade da informação através de uma estrutura de

informação nacional homogénea baseada num sistema de recolha de dados que decorra do processo de produção. O SIGLIC começou a ser implementado em 2004.

O SAPE (Administração.Central.do.Sistema.de.Saúde, 2010) é um software aplicativo que usa, como referencial de linguagem, a International Classification for Nursing Practice do International Council of Nurses – CIPE (versão BETA 2), que permite efectuar o planeamento e o registo da actividade decorrente da prestação de cuidados de enfermagem nas instituições de saúde. O SAPE é orientado para a actividade diária do enfermeiro e visa a organização e o tratamento de informação, processada na documentação de enfermagem, sobre a situação clínica do doente. Tem como objectivos específicos suportar a actividade diária de enfermagem e normalizar o sistema de registos de enfermagem. O SAPE começou a ser implementado em 2002.

1.2. Dificuldades

As três principais partes de um sistema informático são o seu software, o seu hardware e os seus utilizadores mas normalmente esta última parte é atirada para segundo plano construindo assim um sistema que por mais potente e rápido que possa ser pode ter problemas tão simples como não fazer o que o utilizador necessita (Carroll and Daughtrey, 2007). Como já foi referido acima, existe sempre um período de adaptação quando se passa a utilizar um sistema novo, e nesse período até é possível sentir-se uma quebra na produtividade. O que tem que haver é um envolvimento dos utilizadores desde o processo de análise e levantamento de requisitos, passando pelo desenvolvimento, formação para a utilização e após isso, manter contactos frequentes para obter não só o grau de satisfação dos utilizadores mas também feedback sobre possíveis erros e formas de melhorar o sistema que os utilizadores possam detectar. Este processo acarreta um elevado grau de complexidade. É necessário saber que perguntas fazer, como interpretar os resultados e que acções tomar para colmatar as lacunas apontadas. Além disso traz um problema acrescido. Mesmo que as perguntas sejam as certas, as respostas dos utilizadores podem não reflectir a realidade (Carroll and Rosson, 1987).

Existem métodos que se baseiam nas acções dos utilizadores e não nas suas palavras. Um dos mais conhecidos funciona alocando uma pessoa a cada utilizador, pessoa esta que vai registando as acções do utilizador sem interferir no processo, ou pelo menos, sem interferir propositadamente. Existem vários problemas com esta abordagem, pois para o utilizador se sentir à vontade e até para o teste ser realista, o utilizador tem que estar no seu ambiente natural onde utiliza o sistema. Tendo isto em mente é necessário agora fazer entrar em cena um visualizador que vai seguir de perto os passos do utilizador tirando notas sobre o seu comportamento. Este visualizador pode ter um efeito no utilizador do sistema pois o utilizador não está habituado a ser monitorizado e poderá sentir-se vigiado ou mesmo pressionado. O ideal seria utilizar um método que não acrescentasse ruído ao processo.

1.3. Os sistemas informáticos ajudam realmente?

As aplicações informáticas apresentam-se como ferramentas que permitem ajudar os profissionais a cumprir o seu trabalho de forma mais simples, organizada e rápida. Para conseguirmos determinar se estão realmente a fazer o seu trabalho temos de ser capazes de medir os resultados (McNamara, 2010). Existem diversas formas de testar aplicações informáticas desde testes de capacidade até entrevistas com os utilizadores. No fundo, para fazer uma avaliação simples de um sistema primeiro é necessário delinear um conjunto de operações e resultados esperados e segundo é preciso testar o sistema para ver se os resultados obtidos estão em concordância com os esperados. Os testes podem ser levados a cabo por especialistas ou mesmo pelos próprios utilizadores das aplicações mas ambas as hipóteses têm desvantagens. Um conjunto de especialistas suficientemente grande para testar apropriadamente um sistema é demasiado caro enquanto que os utilizadores, por mais baratos que fiquem, atribuem sempre uma avaliação tendenciosa, por mais inadvertidamente que o façam.

A EuroRec (European Institute for Health Records) (EuroRec, 2010) é uma organização europeia independente sem fins lucrativos que certifica e promove a alta qualidade dos EHRs europeus. A organização está a guiar a padronização nos EHR que irá permitir medir e comparar a performance destas aplicações existindo já diversos sistemas certificados (Healthcare.Informatics.Society.of.Ireland, 2009;

CSC.Healthcare.EMEA, 2010; SRC.Infonet, 2010). As certificações são atribuídas às aplicações que passarem com sucesso num conjunto de critérios de performance e funcionalidade.

Entre esses critérios (EuroRec, 2010) estão, por exemplo, a verificação de se o log de auditoria tem registo dos logins e logouts dos utilizadores ou ainda se o log de auditoria tem registo dos eventos de administração de segurança, critérios estes que implicitamente obrigam à gravação desses ficheiros de log. Estas iniciativas revelam que existe um interesse em recorrer a este tipo de fonte de informação para auditoria e melhoramento dos sistemas.

O que propomos neste documento é uma forma de medição que aproveita as informações internas dos sistemas para extrair informações acerca da utilização do próprio sistema. Mas tal como a falta de informação pode ser prejudicial, também o seu excesso é por vezes origem de erros. Sendo assim, para além de extrair informações imparciais, propomos uma forma de apresentação da informação visualmente atractiva e fácil de entender que ela sozinha pode ser objecto de estudo através de um ramo da matemática e da ciência de computadores já bastante estudado chamado Teoria de Grafos.

2. Objectivo

O objectivo deste trabalho é criar uma ferramenta que, através da análise dos ficheiros de log de aplicações informáticas usadas na área da saúde, apresente representações gráfica e intuitiva dos fluxos de utilização dessas aplicações e gere recomendações automáticas quanto às possíveis acções a tomar para melhorar a sua usabilidade.

Este objectivo foi subdividido em:

- Desenho e implementação da aplicação de análise “O Viajante”;
- Avaliação de aplicações informáticas usadas na área da saúde.

3. Estado de arte

Para suportar a metodologia da nossa abordagem, neste capítulo vamos apresentar o que tem vindo a ser feito usando os conceitos propostos. Para isso vamos falar dos métodos de avaliação actuais e quais são as suas vantagens e desvantagens, passando depois para o uso de ficheiros de log como fonte de informação para análises. Abordamos também o uso de teoria de grafos e como ela tem ajudado os investigadores em diversos tipos de estudos. Por fim apresentamos algumas ferramentas usadas actualmente que fazem uso dos conceitos explorados.

Métodos de avaliação

Na saúde, como em qualquer outra área, há um ciclo contínuo de planeamento-avaliação (Trochim, 2006) que permite o avanço do conhecimento num determinado tema. A fase de avaliação geralmente consiste em cinco passos:

1. Formulação das questões de avaliação e hipóteses;
2. Conceptualização da execução das medições;
3. Desenho da execução da avaliação;
4. Análise dos dados da avaliação;
5. Uso dos resultados para tomadas de decisão;

Existem diversos modelos de avaliação (Learning.Technologies.at.Virginia.Tech., 2009; McNamara, 2010), como por exemplo orientada a objectivos ou orientada a processos, que podem ser adaptadas ao objecto de estudo, mas para fazer uma avaliação é necessário primeiro reunir informações sobre o que se pretende avaliar. Existem vários métodos de recolha de informação (APPEAL, 2001; Learning.Technologies.at.Virginia.Tech., 2009; McNamara, 2010) que podem ser usados para o efeito, cada um com vantagens e desvantagens associadas à sua utilização:

A revisão de documentos consiste na recolha de informação escrita pré-existente como minutas, relatórios ou artigos. A vantagem evidente é que esta informação já existe e está muitas vezes disponível facilmente. Em contra-partida pode demorar muito tempo a fazer a revisão desses documentos e pode estar-se a lidar com dados antigos.

Outro método existente para recolher informação é por observação. Neste caso a recolha de informação é feita em tempo real, observando o funcionamento da aplicação em ambiente natural ou num laboratório preparado para o efeito. Mas a vantagem de se conseguir obter uma impressão imediata sobre como a aplicação funciona e como os utilizadores interagem com ela é rapidamente sobreposta por vários problemas. Alguns deles são por exemplo a má ou difícil interpretação dos dados recolhidos, problemas relacionados com eventos que são dependentes da altura em que são feitos e que podem não ser detectados no momento da análise ou até mesmo problemas relacionados com a forma como as pessoas se comportam quando estão a ser observadas, alterando a sua forma natural de agir.

Uma outra forma de recolha de informação é efectuando entrevistas. O objectivo deste método é obter informação directamente perguntando aos utilizadores acerca das suas experiências. Tem a vantagem de se conseguir informação bastante específica e como as entrevistas são individuais há sempre mais margem para se colocar questões mais sensíveis sem que o inquirido se tenha de preocupar com os seus pares ficarem a saber a sua resposta. Mais uma vez a dificuldade de análise dos resultados é uma questão a ter em conta quando se usa este método pois os dados adquiridos com este método podem variar largamente. Para além disso, o resultado da interacção directa entre entrevistador e entrevistado, o primeiro pode influenciar as respostas. As entrevistas são também um método bastante dispendioso em termos de tempo.

Os Focus Groups reúnem grupos de pessoas, tipicamente entre 10 e 12, para se perceber qual é a sua opinião quanto à aplicação. Este método consegue ser relativamente rápido e pode identificar bastantes problemas e efeitos da aplicação de uma só vez. Sendo que os utilizadores dizem de sua justiça, os avaliadores conseguem adquirir um melhor conhecimento da aplicação a partir dos relatos dos próprios

participantes. Para este método também existem dificuldades de análise e interpretação das discussões em termos de tempo dispendido. A necessidade de existência de um bom facilitador do grupo capaz de coordenar convenientemente a discussão é imprescindível. Neste método é também possível que tenham que existir incentivos para os participantes.

A recolha de informação por Case Studies pretende descrever em detalhe os aspectos relacionados com a aplicação, muitas vezes usando vários outros métodos de recolha de forma a construir um caso de estudo. Esta abordagem é muito poderosa e pode resultar numa análise bastante detalhada se compilada e estruturada convenientemente. A maior desvantagem deste método é a sua complexidade na recolha, organização e descrição de informação.

Os questionários oferecem outra abordagem à questão da recolha de informação permitindo inquirir pessoas individualmente sobre um conjunto de tópicos previamente definidos. É uma excelente forma de recolher informação de muitas pessoas e pode ser feita de forma confidencial ou mesmo anónima. Oferece a consistência necessária para efectuar estudos pré e pós possibilitando assim uma boa forma de comparação dos efeitos da aplicação. No entanto os questionários são mais eficientes quando se usam perguntas do tipo Sim/Não ou Verdadeiro/Falso sendo complicado possibilitar a recolha de informações por palavras do próprio participante que depois seja fácil de analisar. Isto pode levar a que o questionário se torne mesmo demasiado impessoal. Também tem que haver algum tempo gasto inicialmente a planear e desenvolver o questionário pois como não existe a possibilidade de interacção em tempo real com os participantes, as perguntas às quais se desejam obter respostas têm que ser todas explanadas anteriormente.

Ficheiros de log

Logs são registos discretos das acções dos utilizadores durante o uso de uma aplicação informática. Estes ficheiros oferecem a possibilidade de recolher informação sobre a interacção computador-humano de um grande número de utilizadores, ao longo do tempo enquanto os utilizadores estão envolvidos em tarefas

habituais e em ambientes habituais (Guzdial, 1993). Apesar de poderem ter diversas apresentações e estruturas, os logs são tipicamente ficheiros contendo texto legível que as aplicações geram automaticamente e que contêm informações acerca das diversas acções que vão sendo efectuadas na aplicação. Existem inúmeros tipos de ficheiros de log:

- 1) Logs de instalação;
- 2) Logs de web counters;
- 3) Logs de servidor;
- 4) ...

Os logs de instalação são gerados quando o programa está a ser instalado. O ficheiro de log vai guardando que acções estão a ser tomadas para o programa ficar correctamente instalado. Se ocorrer algum erro na instalação o ficheiro de log deve conter informação mais específica acerca do sucedido.

Web counters são programas que contam o número de visitantes de uma determinada página de Internet. Normalmente não necessitam de base de dados para guardar estes valores pelo que usam um ficheiro de texto onde escrevem e lê os valores para apresentar aos utilizadores. Estes programas são habitualmente muito simples e escrevem no ficheiro de log apenas o número de visitas e a referência para a página em questão.

Server logs são ficheiros criados e mantidos automaticamente por um servidor que contém informação acerca da actividade realizada pelo mesmo. Um dos exemplos mais conhecidos é o web server access log que contém diversas informações acerca das páginas carregadas por um web server, incluindo o endereço do Internet Protocol (IP) do utilizador cliente, um selo temporal, informação de que página foi requisitada, qual foi o tamanho da página, quanto tempo demorou a carregar, etc.

Ex.

```
...
127.0.0.1 -- [26/Dec/2008:14:33:18 +0000] "GET /homepage.php HTTP/1.1" 200 4068
127.0.0.1 -- [26/Dec/2008:14:33:23 +0000] "GET /index.php?img=gifLogo HTTP/1.1" 200 5172
127.0.0.1 -- [26/Dec/2008:14:34:10 +0000] "GET /favicon.ico HTTP/1.1" 404 209
...
```

As vantagens em usar logs para análise são várias. Ao usar estes ficheiros reduz-se à partida a carga de trabalho necessária para obter os dados e é possível fazê-lo, ao contrário do que acontece por via de outros métodos, de forma não intrusiva (M. Guzdial, 1993; Warren, 2002). Estes ficheiros são criados automaticamente pelas aplicações sem restrições e tecnologicamente de forma fácil (Mark Guzdial, 1994; Gerald Stermsek, 2007), logo apresentam-se não só como uma forma fácil e barata de obter dados (Mark Guzdial, 1994) mas também segura em termos de redução ou mesmo eliminação do efeito Hawthorne (Draper, 2009), visto que a sua recolha não requer uma interacção directa com o utilizador (Mark Guzdial, 1994). Os dados podem portanto ser recolhidos fora de um laboratório e enquanto os utilizadores estão a realizar tarefas reais, não requerendo um ambiente diferente do usual que poderia confundir a análise (Guzdial, 1993), obtendo assim resultados com uma maior validade ecológica (Mark Guzdial, 1994).

Por outro lado as aplicações conseguem gerar uma grande quantidade de dados sobre a sua utilização e um dos maiores problemas tem vindo a ser o que fazer com tão grande quantidade de dados gerado. Os registos de log podem facilmente atingir as centenas ou milhares de megabytes dependendo da utilização da aplicação. Com este volume de dados e sem um método automatizado de análise dos mesmos é difícil determinar o que é útil e como apresentar essa porção de informação útil de uma forma significativa (Guzdial, 1993; Mark Guzdial, 1994; Taha, 2004). Uma forma de fazer sentido desta informação é efectuar um pré-processamento sobre os ficheiros de log que queremos analisar (Gerald Stermsek, 2007) mas este processamento extra vai ser responsável por uma demora mais longa na análise dos resultados e está sujeito a erros no pré-processamento que podem levar a conclusões erradas na análise. Mas mesmo que se interprete os dados de raiz, sem qualquer pré-processamento, os resultados têm que ser interpretados e contextualizados consistentemente (Bertot, 1997). Algumas desvantagens intrínsecas dos logs são por exemplo a incapacidade de fornecer dados sobre quem não está a utilizar a aplicação ou que partes da aplicação não estão a ser utilizadas ou se os utilizadores encontram as respostas aos seus pedidos (Zawitz, 1998). Para agravar estes pontos existem também desvantagens técnicas como por exemplo endereços de IP dinâmicos que dificultam a capacidade de isolamento de utilizadores ou ainda em alguns tipos de aplicações a utilização de

Cache e de separadores no browser que podem fazer com que as estatísticas não sejam uma medida tão válida para determinar a utilização da aplicação quanto poderiam ser (Warren, 2002; Gerald Stermsek, 2007).

Em suma, é necessário algum esforço para extrapolar conclusões a partir de logs mas o valor desse esforço é a capacidade de comparar padrões ao longo do tempo (Nicholas, 1999; Mariner, 2002) e como referiu Nicholas et al. “The data raises questions, it does not in itself provide answers. But just knowing what questions to ask is progress in itself” (1999).

Os ficheiros de log têm vindo a ser usados para um grande número de análises em áreas tão distintas como a gestão, o marketing ou até no próprio desenvolvimento de aplicações alertando o programador para possíveis problemas (M. Guzdial, 1993; Warren, 2002). Foram usados para tarefas tão diversas como para detectar medidas de usabilidade, padrões de utilização, inferir conhecimento ou especialidades, reconhecer oportunidades para optimização de hiperligações, extrair os interesses dos utilizadores e fazer análises etnográficas (Mark Guzdial, 1994; Gerald Stermsek, 2007; Ramadhan, 2009).

Foram encontrados benefícios no uso de ficheiros de log para fazer avaliações efectivas e melhoramento de serviços electrónicos, identificar comunidades de utilizadores, identificar padrões de uso de recursos electrónicos, informar o redesenho e desenvolvimento de portais e avaliar a satisfação do utilizador final (Taha, 2004). Usados em junção com outras técnicas como por exemplo estudos de usabilidade, questionários e entrevistas, a análise de ficheiros de log permitiu formular questões mais informadas (Warren, 2002).

Há ainda diversos estudos em que o uso de ficheiros de log foi fundamental:

Os logs de transacção do sistema de E-journal ScienceDirect OnSite de Taiwan foram analisados e os resultados foram interpretados para se ficar a par dos padrões de utilização e dos comportamentos de busca de informação por parte dos utilizadores finais (Ke, Kwakkelaar et al., 2002). Foram estudados os hábitos de busca de informação académica através da análise de server logs: O caso do Teachers College

Library Website (Asunka, Chae et al., 2009). Foi determinado o uso e a utilidade de funções em E-journals recorrendo aos ficheiros de log: A experiência do SuperJournal Project (Eason, Yu et al., 2000);

Também directamente na área da saúde estes ficheiros já estão a ser utilizados em alguns estudos:

A longevidade e frequência do uso dos dados clínicos de um sistema de informação hospitalar foram calculadas recorrendo a ficheiros de log (Cruz-Correia, Wyatt et al., 2010). Uma análise da precisão clínica do sistema de monitorização de tumores respiratórios CyberKnife foi feita usando ficheiros de log (Hoogeman, Prévost et al., 2009). Uma análise da precisão clínica do rastreamento do movimento respiratório em tempo-real foi feita através da análise de ficheiros de log (Heijmen, Nuytens et al., 2008).

Teoria de Grafos

Desde há muito tempo que se tem dado importância à forma de representação de dados. Os gráficos foram distinguidos de outros sistemas de sinais como, por exemplo representações ilustradas, por serem monossémicos (Bertin, 1983), deixando uma reduzida margem de dúvida para outras interpretações. Várias revisões concluíram que pesquisas na área da aprendizagem feitas antes dos anos 90, apesar de falharem em apresentar uma base teórica que explique como é que os gráficos auxiliam os processos, documentaram que há claramente benefícios em usar apresentações visuais nos mesmos (Winn, 1987; Hegarty, Carpenter et al., 1991; Kozma, 1991). Mais tarde estudos revelaram que as representações gráficas podem ser computacionalmente eficientes quando são desenhadas de forma a que a informação que representam se torne saliente para quem as está a visualizar. Para além disso essas representações são eficazes quando a sua interpretação se baseia mais em processos cognitivos que sejam efectuados automaticamente pelo nosso sistema de percepção visual e não tanto em processos computacionais complexos (Vekiri, 2002).

A teoria de grafos é um ramo conjunto da matemática e da ciência de computadores que estuda conjuntos de objectos e as relações entre eles. O objecto básico de estudo da teoria de grafos é o grafo. Um grafo é representado por um conjunto de pontos (vértices ou nós) que indicam objectos do grupo e rectas (arestas ou arcos) (Caldwell, 1995) que representam ligações entre os objectos. Tanto os arcos como os nós podem ter informações extra internas, por exemplo, um arco pode ter um valor associado (chamado peso) que pode representar um atributo da relação entre os dois objectos que interliga (distância, tempo, número de transições, etc).

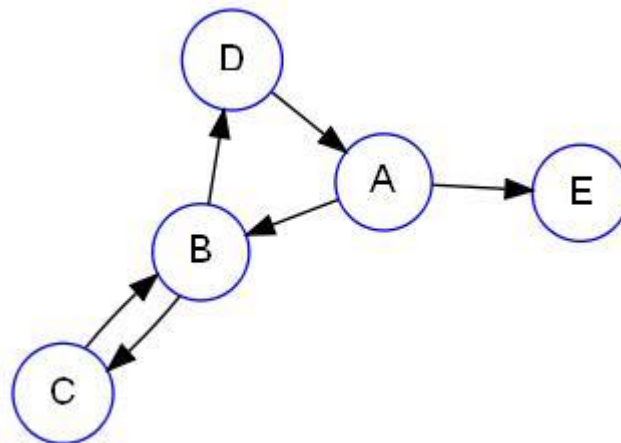


Ilustração 1 - Exemplo de um grafo direccionado

As suas características tornam fácil uma analogia com um website (Zhou, Chen et al., 2001) onde os nós ou vértices são equivalentes às páginas e as arestas ou arcos pode corresponder ao número de transições ou, até, à demora nas transições.

A teoria de grafos contempla vários algoritmos (Wikipedia, 2010) que aplicados aos grafos conseguem revelar variadas informações. Entre os mais conhecidos existe o algoritmo do caminho mais curto que, para um determinado grafo com informação dos pesos nos arcos, dados dois vértices, indica o caminho mais curto entre eles. Assim como este, existem outros algoritmos que podem ser interessantes para retirar informações dos grafos. Nos próximos capítulos serão descritos alguns algoritmos considerados interessantes e qual o seu efeito em grafos resultantes da leitura de ficheiros de log reais.

Os grafos já foram usados em diversos estudos com diferentes objectivos, desde a representação do processo de navegação dos utilizadores (Ramadhan, 2009) até ao

uso dos seus conceitos e definições para a análise da usabilidade de web sites (Thelwall and Wilkinson, 2003; Blazewicz, Pesch et al., 2005; Yan, Qin et al., 2005).

Aproveitando o modelo de grafo existem várias medidas que ajudam à sua interpretação. Estas medidas foram extensamente aplicadas na análise de redes sociais (Scott, 2000) e mesmo sugeridas para serem usadas na análise de ficheiros de log (Winne, Gupta et al., 1994).

Recentemente um estudo (Abedin and Sohrabi, 2009) indica que o uso de teoria de grafos em conjunto com ficheiros de log para avaliar a usabilidade e estrutura de um web site traz alguns benefícios em comparação com outros métodos. O estudo indica que a teoria de grafos possibilita um modo eficaz de ilustração da navegação e actividades dos utilizadores num web site. Para além disso, avalia as associações entre as páginas baseando-se no que realmente aconteceu enquanto que outros métodos focam-se principalmente nas opiniões e modelos mentais que os utilizadores têm da estrutura do web site. Finalmente, visto que o método de input de dados (ficheiros de log) está disponíveis em todos os web servers, concluem que este método é bastante prático e usável até para outros web sites.

Ferramentas existentes

Existem actualmente diversas ferramentas, muito direccionadas, que conseguem extrair informação relevante acerca das páginas disponibilizadas por um servidor, como por exemplo estatísticas sobre as páginas mais visitadas, detecção automática de situações de erro do servidor ou informação sobre o computador do utilizador cliente (browser utilizado, sistema operativo, etc). O quadro abaixo contém informação acerca de algumas das ferramentas mais conhecidas:

Tabela 1 - Ferramentas conhecidas de análise de logs existentes e suas funcionalidades

Aplicação	Usa grafos?	É grátis?	Processa vários tipos de logs?	Tem noção de caminho frequente ou sessão?
AWStats	Não	Sim	Sim	?
aiSee Apache2GDL	Sim	Não	Não	Sim

Aplicação	Usa grafos?	É grátis?	Processa vários tipos de logs?	Tem noção de caminho frequente ou sessão?
FlashStats	Não	Não	Sim	Sim
LogSite	Não	Sim	Não	Não
WebSuxess	Não	Não	Sim	Sim
Webtrends Analytics	Não	Não	Sim	?

Como podemos ver existem várias ferramentas de análise de logs mas da pesquisa realizada apenas uma delas, de formato comercial, usa o conceito de grafo e unicamente como forma de apresentação visual. Em duas das aplicações não foi possível perceber se existia a noção de caminho frequente ou sessão.

4. Desenho do sistema

De modo a criar uma aplicação que respondesse aos desafios identificados os seguintes passos gerais foram seguidos:

- Definir pré-requisitos dos ficheiros de log;
- Conversão de logs para formato único;
- Geração de grafos a partir dos logs;
- Geração de recomendações a partir dos grafos;

De seguida podemos ver mais em detalhe em que consiste cada um dos passos.

Definir pré-requisitos dos ficheiros de log

De forma a poder trabalhar com logs eles terão que obedecer a um conjunto de requisitos mínimos para que seja possível extrair dados dos ficheiros.

Requisitos de estrutura dos ficheiros de log

Os ficheiros de log recolhidos deverão ser os que registam a actividade dos utilizadores e devem obedecer aos seguintes requisitos:

- Os ficheiros devem encontrar-se escritos em texto legível e não em formato binário. Isto permitirá a interpretação e transcrição directa para os grafos e recomendações;
- Os ficheiros deverão conter informação da data em que cada evento ocorre, sendo esta data tão específica quanto pelo menos o segundo. Isto servirá para obter um cálculo mais correcto e relevante em alguns dos algoritmos usados para processamento da informação;
- Os ficheiros deverão conter informação que identifique o utilizador ou a sessão onde decorreu determinada acção. Isto permitirá identificar grupos de utilizadores;

- Os ficheiros deverão conter informação que identifique as acções ou pontos de passagem relevantes para o estabelecimento de percursos dos utilizadores nas aplicações. Sem esta informação não é possível recriar as utilizações das aplicações feitas pelos utilizadores;
- A informação contida nos ficheiros de log deverá conter um separador de informações conhecido e consistente e estas informações deverão manter a estrutura e o formato ao longo de todo o ficheiro (ex: idêntica aos modelos de log das aplicações Apache e IIS).

Requisitos de quantidade de logs

Deverão ser recolhidos ficheiros de log suficientes que contenham informação de utilização de vários meses. Estas informações podem estar contidas em apenas um ficheiro de log ou em vários.

Requisitos de informação nos logs

Para que alguns dos estudos possam ser levados a cabo (ex: possível agrupamento de utilizadores/clustering) é necessários que existam vários utilizadores no sistema em causa e que informação relativa aos mesmos e às acções que eles executam apareça retratada nos ficheiros de log.

Conversão de logs para formato único

Após a recolha de ficheiros de log que obedeçam aos pré-requisitos definidos anteriormente estes deverão passar por um processo de conversão para um formato global interno à aplicação a desenvolver. Esta conversão será feita automaticamente através de um filtro que será incluído na própria aplicação. O resultado desta acção será finalmente a informação usada para transformar os dados contidos nos logs em grafos. Este formato global interno tenciona normalizar os dados que são utilizados na aplicação. Originalmente os logs a que tivemos acesso foram de dois sistemas no activo no Hospital de São João do Porto mas devido à geração habitual de logs em

dois dos web servers mais utilizados a aplicação terá filtros para os seguintes tipos de log:

- IIS WebServer;
- Apache WebServer;
- HSJ.ICU + breast.care;
- intensive.care.

IIS WebServer

Os logs do IIS WebServer têm como separador o carácter “,” (virgula) e o formato mais habitual é o seguinte (Microsoft, 2003):

Client IP address, User name, Date, Time, Service and instance, Server name, Server IP, Time taken, Client bytes sent, Server bytes sent, Service status code, Windows status code, Request type, Target of operation, Parameters
Ex: 192.168.114.201, -, 03/20/01, 7:55:20, W3SVC2, SERVER, 172.21.13.45, 4502, 163, 3223, 200, 0, GET, /DeptLogo.gif, -,

Tabela 2 - Tabela descritiva do formato de log do IIS

Campo	Aparece como	Descrição
Client IP address	192.168.114.201	O endereço IP do cliente.
User name	-	O utilizador é anónimo.
Date	03/20/01	Esta entrada do ficheiro de log foi feita em 20 Março 2001.
Time	7:55:20	Esta entrada no ficheiro de log foi feita às 7:55 AM.
Service and instance	W3SVC2	Isto é um site Web e a instância do site é a 2.
Server name	SERVER	O nome do servidor.
Server IP	172.21.13.45	O endereço IP do servidor.
Time taken	4502	Esta acção demorou 4,502 milissegundos.
Client bytes sent	163	O número de bytes enviados do cliente para o servidor.
Server bytes sent	3223	O número de bytes enviados do servidor para o cliente.
Service status code	200	O pedido foi satisfeito com sucesso.
Windows status code	0	O pedido foi satisfeito com sucesso.
Request type	GET	O utilizador enviou um comando GET ou download.
Target of operation	/DeptLogo.gif	O utilizador queria fazer download do ficheiro DeptLogo.gif.
Parameters	-	Não foram passados parâmetros.

Apache WebServer

Os logs do Apache WebServer têm como separador o carácter “ ” (espaço) e os dois formatos mais habituais são os seguintes (Apache, 2000):

Common Log Format:

Client IP address, Client identity, UserID, Date time and timezone, Client request, HTTP status code, Return object size
Ex: 127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif HTTP/1.0" 200 2326

Tabela 3 - Tabela descritiva do formato de log do Apache WebServer (Common Log Format)

Campo	Aparece como	Descrição
Client IP address	127.0.0.1	O endereço IP do cliente.
Client identity	-	A identidade RFC 1413 do cliente (indisponível).
UserID	Frank	O userid do utilizador.
Date time and timezone	[10/Oct/2000:13:55:36 -0700]	A data, hora e fuso horário de quando o servidor acabou de processar o pedido.
Client request	"GET /apache_pb.gif HTTP/1.0"	O pedido do cliente. O método GET, o recurso pedido apache_pb.gif e o protocolo HTTP HTTP/1.0.
HTTP status code	200	O código do estado HTTP enviado ao cliente. 2xx é uma resposta com sucesso, 3xx um redireccionamento, 4xx um erro no cliente, e 5xx um erro no servidor.
Return object size	2326	O tamanho do objecto enviado ao cliente.

Combined Log Format:

O Combined Log Format é muito parecido com o Common Log Format acrescentando ainda alguns campos:

Client IP address, Client identity, UserID, Date time and timezone, Client request, HTTP status code, Return object size, Referer, User-Agent

```
Ex: 127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif HTTP/1.0" 200 2326
"http://www.example.com/start.html" "Mozilla/4.08 [en] (Win98; I ;Nav)"
```

Os campos adicionais são:

Tabela 4 - Tabela descritiva do formato de log do Apache WebServer (Combined Log Format)

Campo	Aparece como	Descrição
Referer	"http://www.example.com/start.html"	O site que referenciou o pedido actual.
User-Agent	"Mozilla/4.08 [en] (Win98; I ;Nav)"	Informação de identificação do browser do cliente.

HSJ.ICU – breast.care

Por uma questão de protecção dos dados dos utilizadores os logs obtidos desta aplicação sofreram algumas alterações. Os dados utilizados para a análise feita neste documento são na realidade uma exportação pré-processada do ficheiro de log original. Sendo assim, o formato de log utilizado após exportação e codificação de alguns dados tem como separador o carácter “;” (ponto e virgula) e os campos de informação são os seguintes:

```
Sessão, Data, Código da operação, Descrição do código, Número sequencial, Identificador do relatório, Observação, Formulário, Descrição completa  
Ex.: 2072478;2-8-10 8:47;11;breastcare;150107;;434;Antecedentes;breastcare | Antecedentes
```

Tabela 5 - Tabela descritiva da exportação do formato de log do HSJ.ICU

Campo	Aparece como	Descrição
Sessão	2072478	Identificador da sessão
Data	2-8-10 8:47	Data em que a operação ocorreu
Código da operação	11	Número identificador do tipo de operação
Descrição do código	breastcare	Número identificador do estado da operação
Número sequencial	150107	Número sequencial identificador da operação
Identificador do relatório	(nulo)	Número identificador do relatório envolvido
Observação	434	Observações gerais

Campo	Aparece como	Descrição
Formulário	Antecedentes	Formulário visualizado
Descrição completa	breastcare Antecedentes	Descrição do código + Formulários

De modo a dar algum sentido a esta informação, e visto que o campo que identifica a operação é um código, foi compilado outro ficheiro com as equivalências correspondentes. Este ficheiro tem apenas duas colunas:

Id, Descrição
Ex.: 3, Pesquisa de utente

Tabela 6 - Tabela descritiva das operações contidas no log do HSJ.ICU

Campo	Aparece como	Descrição
Id	3	Identificador da operação
Descrição	Pesquisa de utente	Descrição da operação

Intensive.care

O formato de log utilizado pela aplicação intensive.care tem como separador o carácter “;” (ponto e vírgula) e os campos de informação são os seguintes:

Id_Sequencial, Data, Formulário, Sessão
Ex.: 267;08-05-2004 12:43;Terapeutica;14280

Tabela 7 - Tabela descritiva das operações contidas no log do Intensive.care

Campo	Aparece como	Descrição
Id_Sequencial	267	Identificador sequencial
Data	08-05-2004 12:43	Data em que a operação ocorreu
Formulário	Terapeutica	Nome do formulário acedido
Sessão	14280	Identificador da sessão

Geração de grafos a partir dos logs

Uma vez aplicados os filtros aos ficheiros de log recolhidos e obtida a informação necessária, a aplicação a desenvolver deverá gerar vários tipos de grafos que serão explicados mais à frente na secção da aplicação. Neste passo a aplicação deverá também gerar automaticamente uma representação gráfica do grafo ilustrando a utilização da aplicação de que o log originou.

Geração de recomendações a partir dos grafos

A partir dos grafos gerados no passo anterior são geradas as recomendações de utilização usando para isso algoritmos de processamento de grafos.

As recomendações estão divididas em dois tópicos gerais:

- **Melhorias:** Mudança significativa na navegação da aplicação ou reformulação do código da aplicação que fará com que esta apresente funcionalidades estendidas ou um aumento na performance da sua utilização;
- **Erros:** Detecção de problemas existentes na aplicação que traduzem situações indesejáveis em termos de funcionamento ou impossibilidades de acesso na aplicação;

Os algoritmos de processamento de grafos considerados relevantes para a geração de recomendações foram os seguintes:

Tabela 8 - Algoritmos da teoria de grafos utilizados e a sua utilidade

Algoritmo	Descrição e Interesse	Tipo de recomendação
Cliques	<p>Este algoritmo encontra conjuntos de n nós em que quaisquer dois nós do conjunto têm uma ligação bidireccional entre eles. O número de elementos n é maximizado até se encontrar o conjunto máximo no grafo com estas características, denominado clique máximo.</p> <p>Serve para encontrar blocos de acções interligadas definidos pelos próprios utilizadores ao trabalharem com a aplicação.</p>	Melhoria

Algoritmo	Descrição e Interesse	Tipo de recomendação
Custom Graph Median	<p>Este algoritmo ordena os nós do grafo por ordem decrescente de número de transições de saída.</p> <p>Serve para encontrar o conjunto de nós centrais, ou seja, os nós a partir dos quais saem mais transições. Esta ordenação alerta para que nós podem ter um fluxo maior e que devem ser otimizados.</p>	Melhoria
Invalid Path	<p>Este algoritmo encontra transições directas entre nós.</p> <p>Serve para encontrar nós que não deveriam ter qualquer transição o que permite revelar possíveis erros.</p>	Erro

A aplicação – “Viajante”

A aplicação criada será uma web-application programada em PHP que interpreta ficheiros de log de diversos formatos e apresenta e analisa grafos constituídos a partir dos logs. Não fazendo uso de bases de dados no sentido convencional do termo dispõe em alternativa de um repositório de ficheiros de log. De agora em diante, para uma mais fácil distinção entre a aplicação desenvolvida e as aplicações estudadas, as referências à aplicação desenvolvida serão feitas usando o nome “Viajante”.

A arquitectura da aplicação “Viajante”

O “Viajante” necessita de um servidor web que interprete ficheiros PHP. Existe um directório “tmp” na raiz da aplicação que age como o repositório de ficheiros de log que lhe servem de input. Este directório precisa de ter permissões de leitura e escrita para o servidor web devido à funcionalidade de upload de logs que será mais tarde especificada. O servidor que disponibilize a aplicação necessita ainda de ter instalado a biblioteca GraphViz (Ellson, Gansner et al., 2010) que vai permitir desenhar grafos em tempo de execução.

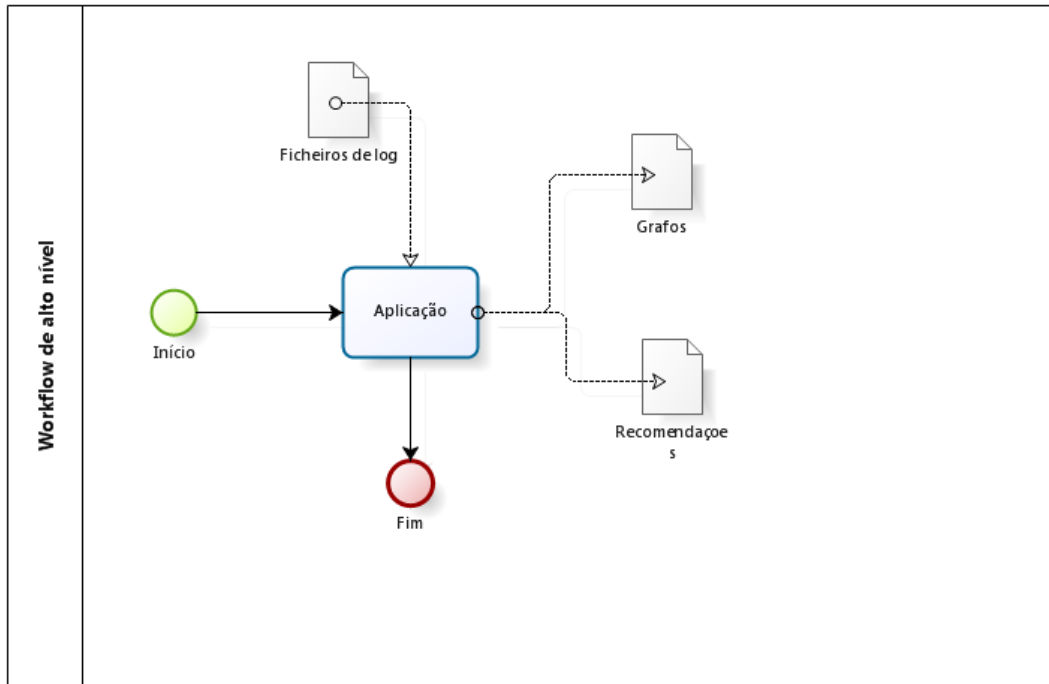
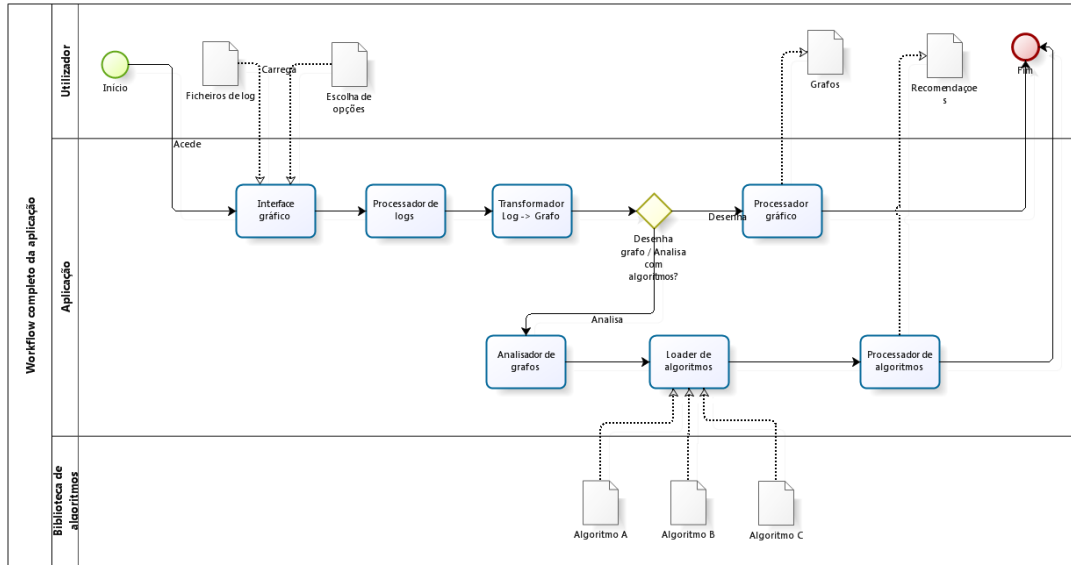


Ilustração 2 – Workflow de alto nível da aplicação “Viajante”

O workflow interno da aplicação é mais complexo e pode ser retratado pelo seguinte modelo:



powered by

 BizAgil
 Process Modeler

Ilustração 3 - Workflow completo da aplicação "Viajante"

Funcionalidades da aplicação "Viajante"

Muitas vezes, para tentar poupar espaço, uma aplicação não grava no ficheiro de log as informações num formato perceptível. É muito mais rápido e ocupa muito menos espaço gravar num log a acção "01" do que gravar a acção "Pesquisa de pacientes". As implicações práticas deste artifício no "Viajante" fazem-se notar no desenho do grafo da aplicação estudada pois os nós (ou vértices) vão estar identificados com números e não com a designação da acção.

```

1 1,101,2010-07-26
2 1,102,2010-07-26
3 1,104,2010-07-26
4 2,101,2010-07-26
5 2,102,2010-07-26
6 2,104,2010-07-26
7 3,102,2010-07-26
8 3,104,2010-07-26
9 ...
  
```

Ilustração 4 - Exemplo de ficheiro de log com acções simplificadas

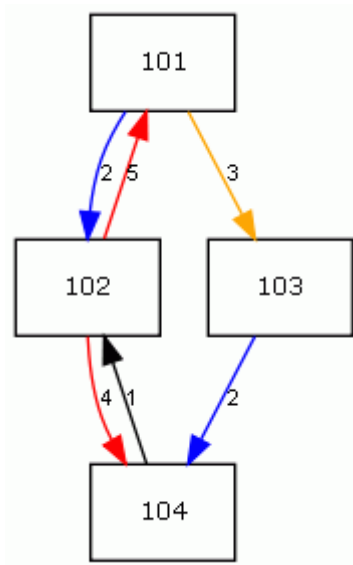


Ilustração 5 - Grafo com as designações simplificadas originais do log

Para contornar este problema foi criada uma tabela de tradução que, em tempo de execução e desenho do grafo, interpreta os nós e renomeia-os para uma designação mais amigável. Esta tabela é configurada manualmente pois cada aplicação pode usar números ou identificadores diferentes para as acções que seja possível executar ou estados em que a aplicação possa estar. Sendo assim, com a tabela preenchida, para o mesmo log vamos obter o seguinte resultado:

```

1 1,101,2010-07-26
2 1,102,2010-07-26
3 1,104,2010-07-26
4 2,101,2010-07-26
5 2,102,2010-07-26
6 2,104,2010-07-26
7 3,102,2010-07-26
8 3,104,2010-07-26
9 ...
  
```

Ilustração 6 - Exemplo de log com as acções simplificadas

Tabela de tradução:

101	:	Login
102	:	Pesquisar
103	:	Ver relatório
104	:	Editar
	:	
	:	
	:	

Ilustração 7 – Exemplo de tabela de tradução

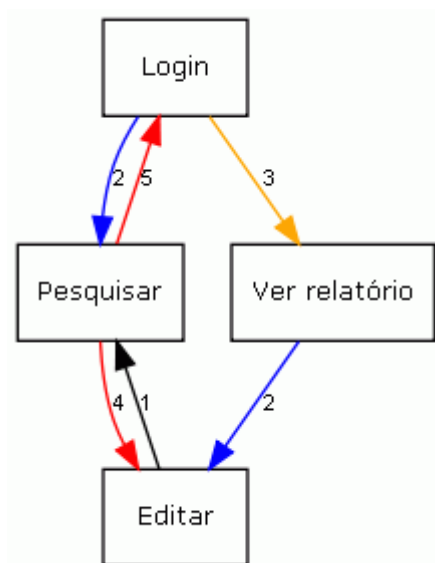


Ilustração 8 - Grafo com designações traduzidas pela tabela de tradução

Para efeitos de prova de conceito esta tabela permite actualmente definir 10 designações distintas para 10 acções inscritas no log. Caso necessário o número de campos é facilmente expandido de forma programática no “Viajante”. Ainda que tenham sido adicionados templates de base para a leitura de tipos de logs habituais (IIS, Apache) existe a opção de personalizar o tipo de log, permitindo fazer leituras a tipos de logs menos convencionais. Esta funcionalidade foi conseguida deixando em aberto a configuração dos campos necessários para interpretar o log: o índice do campo de sessão, o índice do campo de acção efectuada e o carácter separador de dados no ficheiro. Estas três informações são as únicas necessárias para se conseguir usar as capacidades básicas do “Viajante”. Alguns campos extra poderão ser necessários para, por exemplo, análises de durações entre transições.

Devido à habitual grande dimensão dos ficheiros de log uma funcionalidade de upload de ficheiros foi adicionada ao “Viajante”. Este upload permite enviar ficheiros para o servidor para que o processamento seja feito de uma forma mais rápida. Mais uma vez esta funcionalidade é uma prova de conceito, sendo mais um facilitador do que propriamente uma funcionalidade obrigatória para o núcleo da aplicação. Sendo assim não há qualquer organização ou tratamento na forma como os ficheiros de log chegam ao servidor. Ao fazer upload de ficheiros eles são criados numa pasta da aplicação juntamente com todos os outros para que possam mais tarde ser escolhidos através do interface da aplicação para serem interpretados.

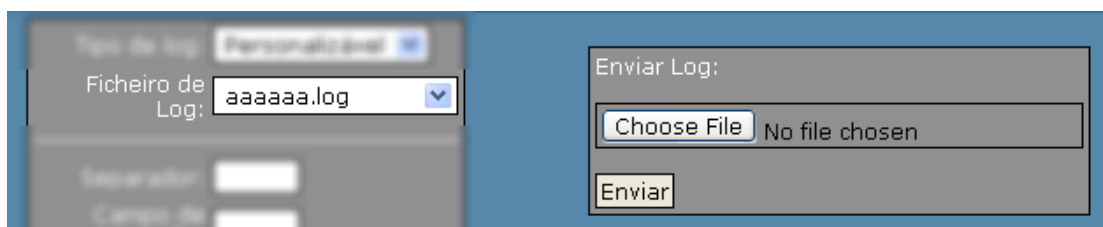


Ilustração 9 - Upload e escolha de ficheiros de log

O “Viajante” permite desenhar vários tipos de grafos. O quadro abaixo apresenta as possibilidades. É também possível filtrar quais os nós e quais as sessões a ter em conta para o desenho do grafo. Isto permite isolar parte do grafo geral para que seja mais fácil estudar subconjuntos do mesmo e isolar utilizadores individuais ou grupos de utilizadores para auxílio, por exemplo, a estudos de clustering.

Tabela 9 - Tipos de grafos gerados pelo “Viajante”

	Grafo conjunto	Grafo segmentado
Número de transições (cíclico)	Obtemos informação agrupada sobre a aplicação em termos de fluxos.	Obtemos informação individual ou segmentada sobre a aplicação em termos de fluxos.
Demora nas transições (cíclico)	Obtemos informação agrupada sobre a aplicação em termos de tempos.	Obtemos informação individual ou segmentada sobre a aplicação em termos de tempos.
Níveis de transição (árvore)	Obtemos informação detalhada sobre os passos que os utilizadores tomaram.	Obtemos informação detalhada individual ou segmentada sobre os passos que os utilizadores tomaram.

Ao gerarmos um grafo conjunto com o número de transições conseguimos obter um fluxo global da aplicação estudada. Conseguimos determinar a partir daí o caminho mais percorrido e detectar erros de utilização inválida. No grafo exemplificativo abaixo podemos ver o caminho mais percorrido “Login” -> ”Pesquisar” -> “Ver relatório” -> “Logout” com 18 transições no geral.

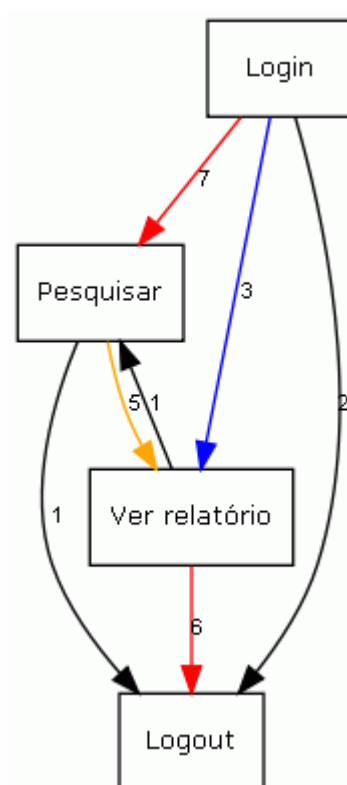


Ilustração 10 - Grafo conjunto com número de transições

Ao gerarmos um grafo conjunto com a demora nas transições conseguimos encontrar os nós (páginas, formulários, etc) que demoram mais tempo a processar a informação. O valor dos arcos deste tipo de grafo é a média de tempo que levou a fazer essa transição tendo em conta todas as vezes que ela ocorreu. Podemos interpretar o grafo abaixo e descobrir que a transição que demorou mais tempo foi a “Login” -> “Pesquisar”. Recomendações de optimização na aplicação podem surgir a partir daqui.

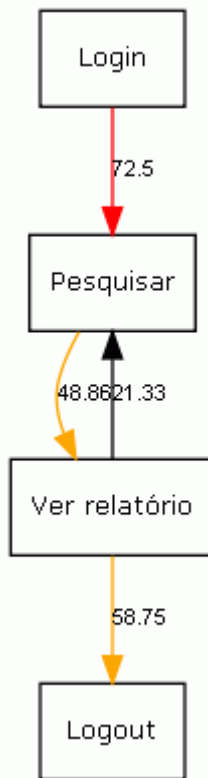


Ilustração 11 - Grafo conjunto com demora nas transições

Ao gerarmos um grafo conjunto com os níveis de transição conseguimos encontrar os caminhos percorridos de uma forma não cíclica. No grafo abaixo podemos ver em cada passo de utilização qual a probabilidade de escolha do passo seguinte. Se assumirmos o Login como o nível 1, podemos ver no nível 3 que se um utilizador está no nó “VerRelatório” existe 50% de hipótese de fazer uma nova pesquisa antes de fazer logout.

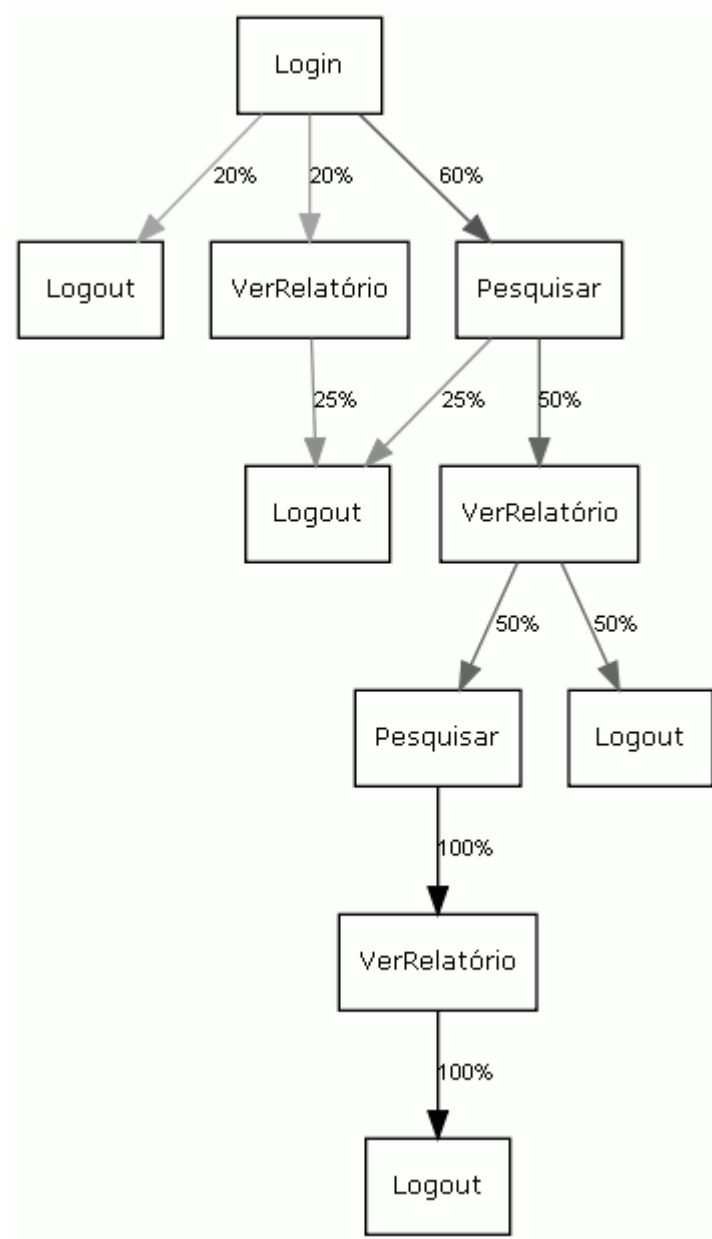


Ilustração 12 - Grafo conjunto com níveis de transição

Ao gerarmos um grafo segmentado com o número de transições conseguimos obter fluxos de utilização individuais da aplicação estudada. Efectuando uma comparação a diversos destes grafos podemos conseguir encontrar clusters de utilizadores. Este caso tem interesse para aplicações com vários “roles” ou perfis de utilização.

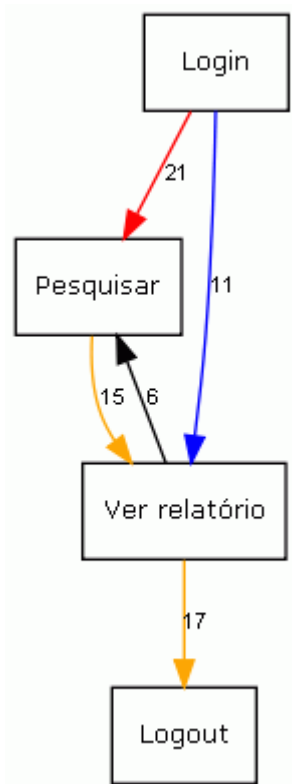


Ilustração 13 - Grafo segmentado com número de transições (1º conjunto)

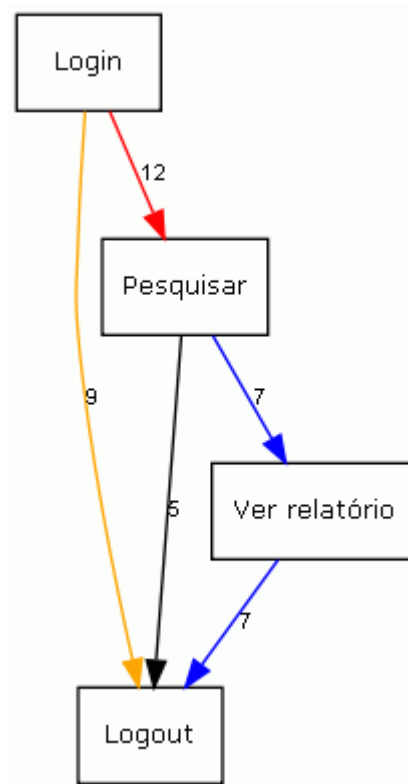


Ilustração 14 - Grafo segmentado com número de transições (2º conjunto)

Ao gerarmos um grafo segmentado com a demora nas transições conseguimos obter tempos de demora de utilização individuais ou agrupados na aplicação. Uma comparação de diversos destes grafos pode revelar uma relação entre clusters de utilizadores (ou perfis de utilizadores) e demora de acções na aplicação. Mais uma vez este caso tem interesse para aplicações com vários “roles” ou perfis de utilização.

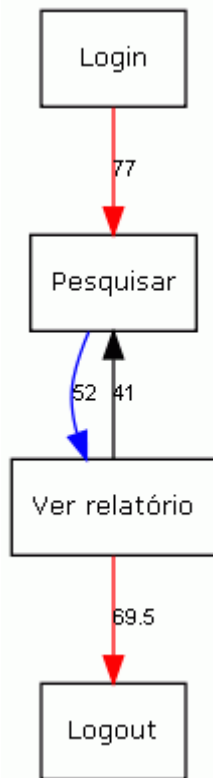


Ilustração 15 - Grafo segmentado com demora nas transições (1º conjunto)

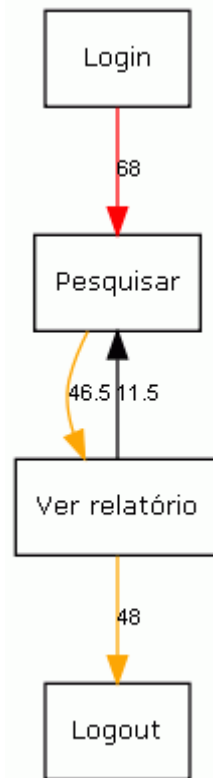


Ilustração 16 - Grafo segmentado com demora nas transições (2º conjunto)

Ao gerarmos um grafo segmentado com os níveis de transição conseguimos obter os caminhos não cíclicos individuais ou segmentados na aplicação. Uma comparação de diversos destes grafos pode revelar uma relação entre clusters de utilizadores e a quantidade de páginas ou formulários que estes vêm por sessão.

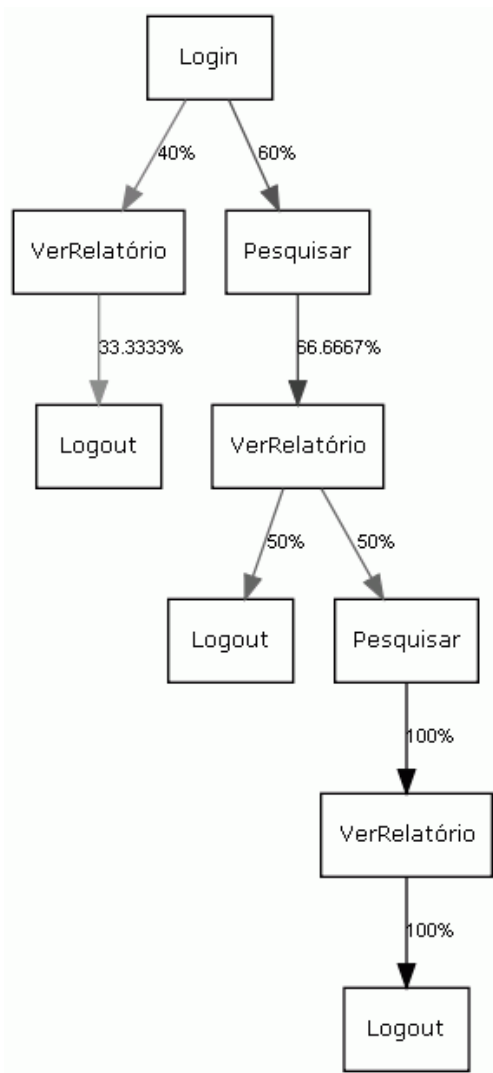


Ilustração 17 - Grafo segmentado com níveis de transição (1º conjunto)

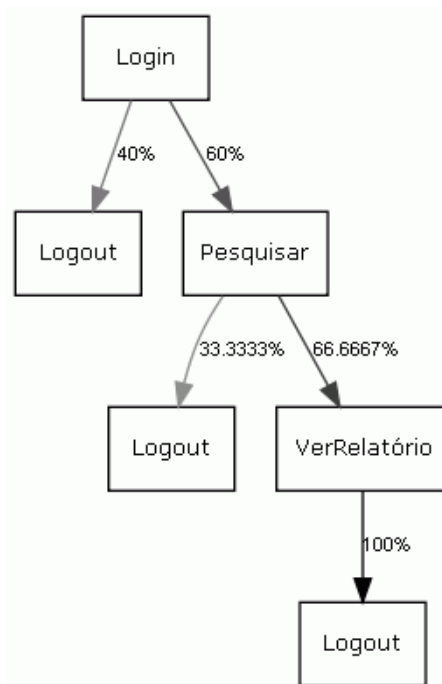


Ilustração 18 - Grafo segmentado com níveis de transição (2º conjunto)

Pode existir a impossibilidade de criar alguns destes tipos de grafos dependendo da informação existente no ficheiro de log. Nomeadamente os grafos com a demora nas transições exigem que o ficheiro de log contenha esse tipo de informação.

Como pode ser visto nos grafos exemplificativos acima foram associadas cores distintas às transições tanto dos grafos cíclicos como dos grafos em árvore de modo a fornecer informação visual mais pronta aquando da visualização de grafos gerados. Para os grafos cíclicos existe uma escala de cor que indica quais foram os caminhos com maior número de transições.



Ilustração 19 - Escala de cor para os grafos cíclicos

Mais à esquerda na escala significa que existe um maior número de transições. Mais à direita indica um menor número de transições. É possível haver mais do que uma transição com a mesma cor, não representando estas cores portanto as 4 transições com maior número mas sim dando uma visão global sobre o grafo indicando que conjuntos de transições mais frequentes existem. Qualquer uma das cores pode ser aplicada a várias transições. Para a escolha de cores é encontrado o intervalo entre as transições com maior e menor peso. De seguida encontram-se os valores equivalentes a 25%, 50% e 75% desse intervalo. Às transições são finalmente associadas as cores dependendo do peso da transição e do intervalo de pesos encontrado anteriormente.

Para os grafos em árvore foi adoptada uma escala de cinzentos para revelar as transições mais efectuadas.



Ilustração 20 - Escala de cinzentos para grafos em árvore

Os dez níveis da escala de cinzentos permitem fazer um cálculo idêntico ao dos grafos cíclicos com a diferença que nos grafos em árvore o que é usado como peso nas transições é a percentagem de transições efectuadas e não o número de transições. Isto permite associar directamente a escala à percentagem de transições, sendo que o cinzento mais carregado equivale a 100% das transições de um determinado nó para outro e o cinzento mais claro significa que a transição ocorreu em menos de 10% dos casos.

5. Avaliação

Após os logs terem sido recolhidos eles foram enviados para o “Viajante” que os analisou gerando representações gráficas de grafos que ilustram os logs e aplicando os algoritmos de teoria de grafos anteriormente indicados. Essas informações, juntamente com as expectativas de peritos e detalhes sobre as aplicações, são apresentadas neste capítulo.

Avaliação pela aplicação “Viajante”

De seguida apresentamos a avaliação efectuada pelo “Viajante” às duas aplicações estudadas. Foi feita uma primeira avaliação à aplicação “breast.care” desenhando diversos grafos e aplicando os três algoritmos de análise. Seguidamente fazemos o mesmo processo para a aplicação “intensive.care”.

breast.care

O breast.care é um módulo da plataforma de base ICU. O log que contém informações sobre a utilização do breast.care contém também informação sobre a utilização do ICU. Os utilizadores podem movimentar-se livremente entre a aplicação base e o módulo sendo que os registos de log ficam gravados numa linha temporal contínua. O ficheiro de log utilizado é referente ao período de 01-08-2010 a 19-10-2010 e contém 538470 registos individuais de informações relativas às acções realizadas nesse período. O ficheiro contém, entre outras, informações relativas à sessão, acção e hora associadas a cada acção executada.

Ex.:

```
Session_ID;Date;OP_Code;Descr_Code;Seq_Number;Report_ID;Obs;Form;Complete_Descr  
2127524;9-9-10 16:43;59;Listagem de relatórios;;;51.-1 [n=0];;Listagem de relatórios
```

Foram gerados diversos grafos relativos a toda a aplicação, apenas às acções de base do ICU e apenas às acções do breast.care.

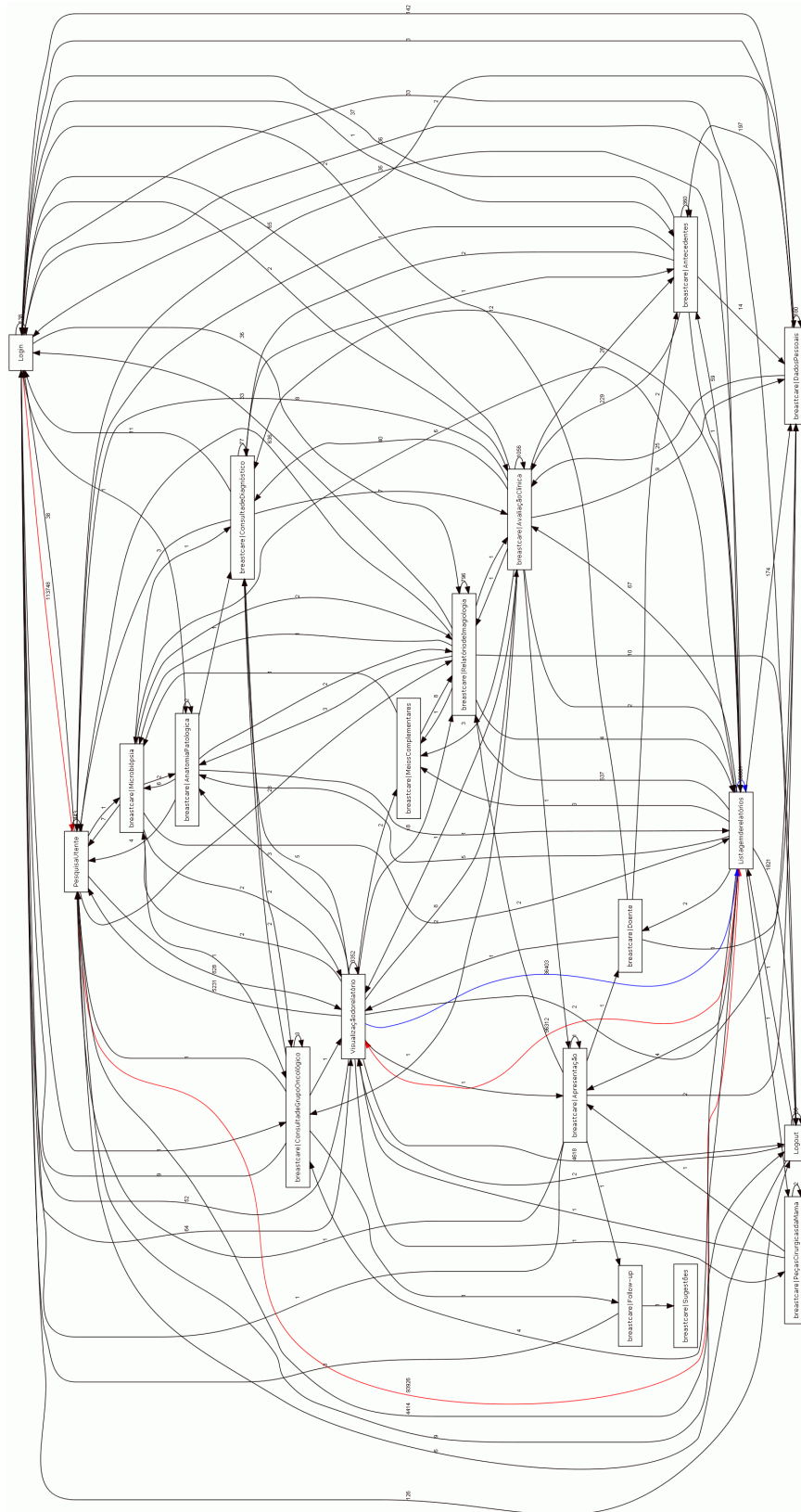


Ilustração 21 - Grafo cíclico completo da aplicação ICU+breast.care

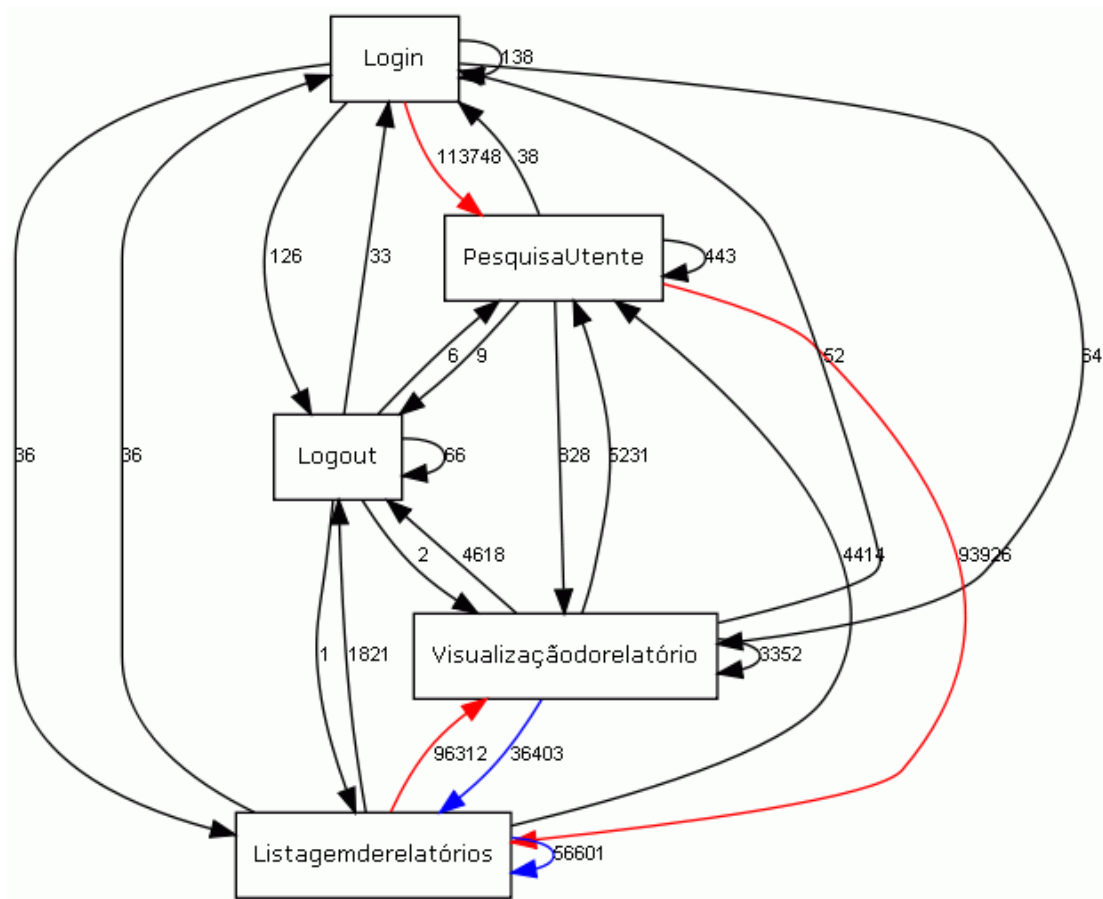


Ilustração 22 – Grafo cíclico relativo apenas à aplicação ICU

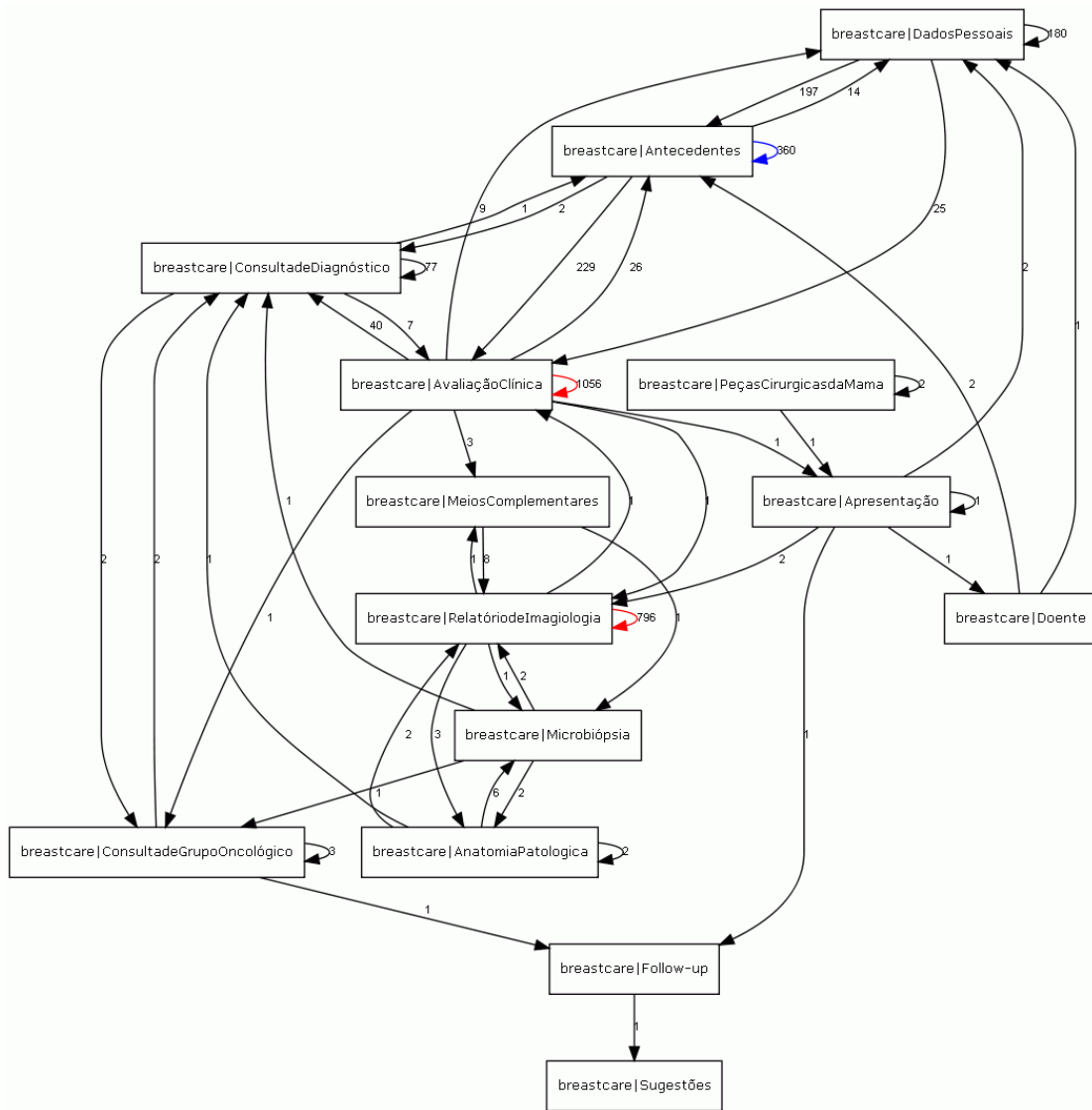


Ilustração 23 - Grafo cíclico relativo apenas à aplicação breast.care

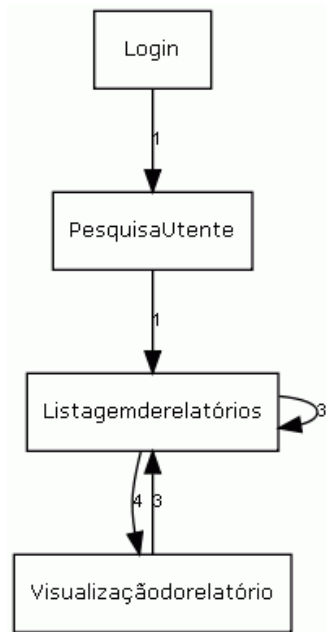


Ilustração 24 - Grafo cíclico segmentado da sessão com o id 2071506 da aplicação HSJ.ICU+breast.care

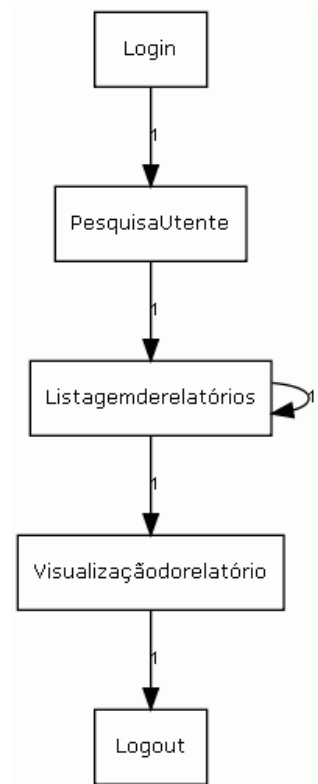


Ilustração 25 - Grafo cíclico segmentado da sessão com o id 2120433 da aplicação HSJ.ICU+breast.care

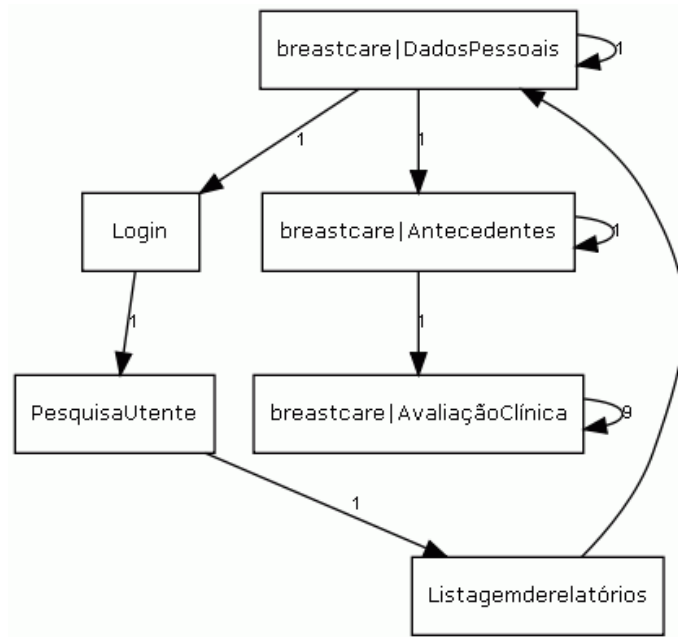


Ilustração 26 - Grafo cíclico segmentado da sessão com o id 2072478 da aplicação HSJ.ICU+breast.care

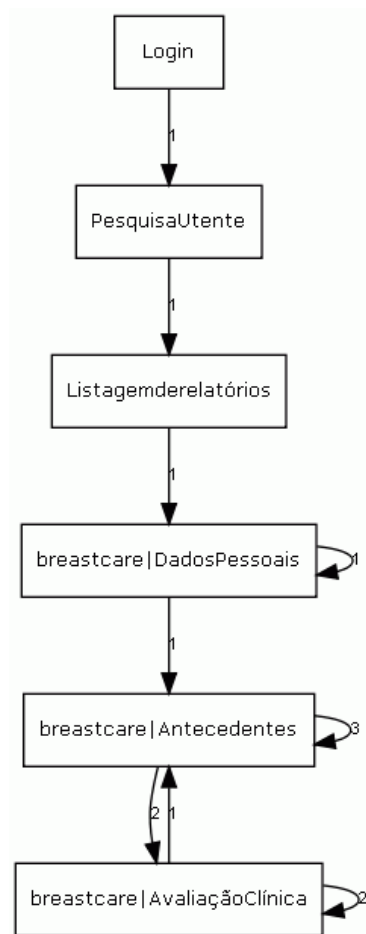


Ilustração 27 - Grafo cíclico segmentado da sessão com o id 2122535 da aplicação HSJ.ICU+breast.care

O algoritmo “CLIQUES” foi aplicado sobre o grafo geral da aplicação (ICU+breast.care) acima apresentado. Foram detectados 19 subconjuntos de clique 1 e 36 subconjuntos de clique 2. Estes subconjuntos não são apresentados devido à sua irrelevância. Foram no entanto também detectados os seguintes cliques:

Subconjuntos de Clique 3: 33 resultados

breastcare | Antecedentes + breastcare | AvaliaçãoClínica + breastcare | DadosPessoais
breastcare | Antecedentes + breastcare | AvaliaçãoClínica + breastcare | ConsultadeDiagnóstico
breastcare | AnatomiaPatologica + breastcare | Microbiópsia + breastcare | Relatóriodelmagiologia
PesquisaUtente + breastcare | Microbiópsia + breastcare | Relatóriodelmagiologia
PesquisaUtente + Visualizaçãodorelatório + breastcare | Microbiópsia
Logout + PesquisaUtente + Visualizaçãodorelatório
Login + breastcare | AvaliaçãoClínica + breastcare | Relatóriodelmagiologia
Login + breastcare | AvaliaçãoClínica + breastcare | DadosPessoais
Login + breastcare | Antecedentes + breastcare | DadosPessoais
Login + breastcare | Antecedentes + breastcare | AvaliaçãoClínica
Login + Visualizaçãodorelatório + breastcare | AvaliaçãoClínica
Login + PesquisaUtente + breastcare | Relatóriodelmagiologia
Login + PesquisaUtente + Visualizaçãodorelatório
Login + Logout + Visualizaçãodorelatório
Login + Logout + PesquisaUtente
Listagemderelatórios + breastcare | Microbiópsia + breastcare | Relatóriodelmagiologia
Listagemderelatórios + breastcare | AvaliaçãoClínica + breastcare | Relatóriodelmagiologia
Listagemderelatórios + breastcare | Antecedentes + breastcare | AvaliaçãoClínica
Listagemderelatórios + breastcare | AnatomiaPatologica + breastcare | Relatóriodelmagiologia
Listagemderelatórios + breastcare | AnatomiaPatologica + breastcare | Microbiópsia
Listagemderelatórios + Visualizaçãodorelatório + breastcare | Microbiópsia
Listagemderelatórios + Visualizaçãodorelatório + breastcare | AvaliaçãoClínica
Listagemderelatórios + PesquisaUtente + breastcare | Relatóriodelmagiologia
Listagemderelatórios + PesquisaUtente + breastcare | Microbiópsia
Listagemderelatórios + PesquisaUtente + Visualizaçãodorelatório
Listagemderelatórios + Logout + Visualizaçãodorelatório
Listagemderelatórios + Logout + PesquisaUtente
Listagemderelatórios + Login + breastcare | Relatóriodelmagiologia
Listagemderelatórios + Login + breastcare | AvaliaçãoClínica
Listagemderelatórios + Login + breastcare | Antecedentes

Listagemderelatórios + Login + Visualizaçãodorelatório

Listagemderelatórios + Login + PesquisaUtente

Listagemderelatórios + Login + Logout

Subconjuntos de Clique 4: 13 resultados

Login + breastcare | Antecedentes + breastcare | AvaliaçãoClínica + breastcare | DadosPessoais

Login + Logout + PesquisaUtente + Visualizaçãodorelatório

Listagemderelatórios + breastcare | AnatomiaPatologica + breastcare | Microbiópsia + breastcare | Relatóriodelmagiologia

Listagemderelatórios + PesquisaUtente + breastcare | Microbiópsia + breastcare | Relatóriodelmagiologia

Listagemderelatórios + PesquisaUtente + Visualizaçãodorelatório + breastcare | Microbiópsia

Listagemderelatórios + Logout + PesquisaUtente + Visualizaçãodorelatório

Listagemderelatórios + Login + breastcare | AvaliaçãoClínica + breastcare | Relatóriodelmagiologia

Listagemderelatórios + Login + breastcare | Antecedentes + breastcare | AvaliaçãoClínica

Listagemderelatórios + Login + Visualizaçãodorelatório + breastcare | AvaliaçãoClínica

Listagemderelatórios + Login + PesquisaUtente + breastcare | Relatóriodelmagiologia

Listagemderelatórios + Login + PesquisaUtente + Visualizaçãodorelatório

Listagemderelatórios + Login + Logout + Visualizaçãodorelatório

Listagemderelatórios + Login + Logout + PesquisaUtente

Subconjuntos de Clique 5: 1 resultado

Listagemderelatórios + Login + Logout + PesquisaUtente + Visualizaçãodorelatório

Recomendação :: Melhoria

Cada nó contido no(s) clique(s) de nível 5 deve conter links de fácil acesso para todos os outro nós do clique.

* As recomendações emitidas não levam em conta a lógica da aplicação e por isso devem sempre ser interpretadas tendo em atenção essas informações.

Aplicando o algoritmo “CUSTOM GRAPH MEDIAN” o resultado é o seguinte:

Nº vezes origem	Nome do nó
114018	Login
103557	Listagemderelatórios

94825	PesquisaUtente
46346	Visualizaçãodereatório
689	breastcare Relatóriodelmagiologia
366	breastcare DadosPessoais
284	breastcare Antecedentes
157	breastcare AvaliaçãoClínica
42	Logout
24	breastcare ConsultadeDiagnóstico
17	breastcare Microbiópsia
14	breastcare AnatomiaPatologica
14	breastcare ConsultadeGrupoOncológico
9	breastcare MeiosComplementares
8	breastcare Apresentação
6	breastcare Doente
4	breastcare Follow-up
2	breastcare PeçasCirurgicasdaMama
0	breastcare Sugestões

Recomendação :: Melhoria

Os nós “Login” e “Listagemderelatórios” representam mais de 50% das transições e devem ter acesso fácil a partir de qualquer nó.

* As recomendações emitidas não levam em conta a lógica da aplicação e por isso devem sempre ser interpretadas tendo em atenção essas informações.

Para a aplicação do algoritmo “INVALID PATH” foram identificados como nós chave “Login” e “Logout” correspondendo ao nó de entrada e ao nó de saída. O algoritmo foi aplicado sobre o grafo geral da aplicação (ICU+breast.care) acima apresentado e o resultado foi o seguinte:

Há uma falha. Depois de acederem ao nó Logout, 66 utilizador(es) fizeram a transição Logout_Logout

Há uma falha. Depois de acederem ao nó Logout, 6 utilizador(es) fizeram a transição Logout_PesquisaUtente

Há uma falha. Depois de acederem ao nó Logout, 1 utilizador(es) fizeram a transição Logout_Listagemderelatórios

Há uma falha. Depois de acederem ao nó Logout, 2 utilizador(es) fizeram a transição Logout_Visualizaçãodorelatório

* As recomendações emitidas não levam em conta a lógica da aplicação e por isso devem sempre ser interpretadas tendo em atenção essas informações.

intensive.care

Para o intensive.care foi usado um ficheiro de log referente à sua utilização no período entre 04-05-2004 e 30-09-2010. O ficheiro contém 133344 registos individuais de informações relativas às acções realizadas nesse período. O ficheiro contém, entre outras, informações relativas à sessão, acção e hora associadas a cada acção executada.

Ex.:

ID;Date;Action;Session_ID

352;11-05-2004 16:22;Terapeutica;14303

Foi gerado um grafo completo da aplicação.

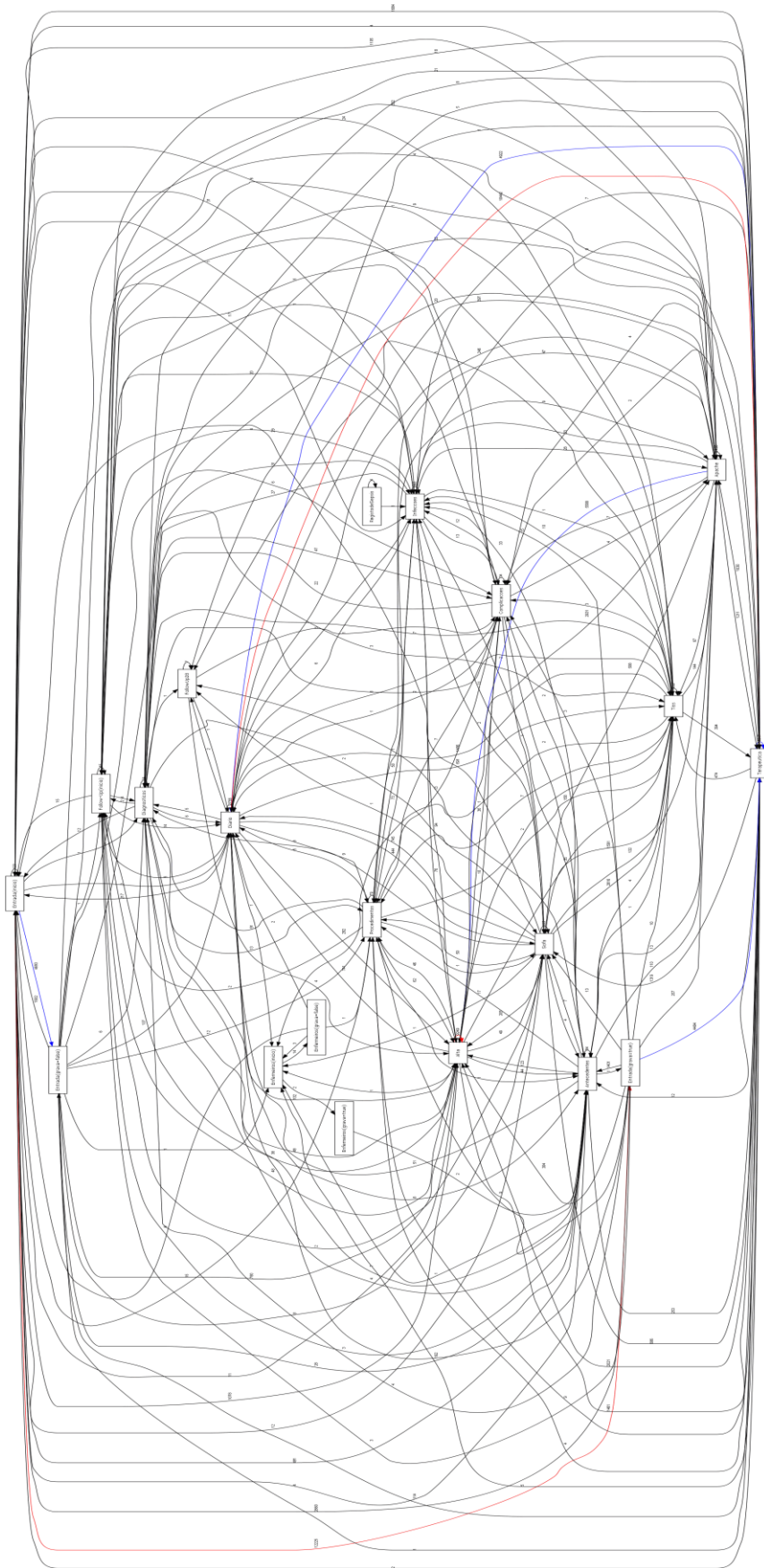


Ilustração 29 - Grafo completo do intensive.care

O algoritmo “CLIQUES” foi aplicado ao grafo acima obtendo todos os subconjuntos contendo desde 1 até 10 elementos. Por questões de relevância apenas os últimos 3 subconjuntos de 8, 9 e 10 elementos são mostrados:

Subconjuntos de clique 8: 63 resultados

Apache + Diagnosticos + Diario + Infeccoes + Procedimentos + Sofa + Terapeutica + Tiss
Antecedentes + Diagnosticos + Diario + Infeccoes + Procedimentos + Sofa + Terapeutica + Tiss
Antecedentes + Apache + Diario + Infeccoes + Procedimentos + Sofa + Terapeutica + Tiss
Antecedentes + Apache + Diagnosticos + Infeccoes + Procedimentos + Sofa + Terapeutica + Tiss
Antecedentes + Apache + Diagnosticos + Diario + Procedimentos + Sofa + Terapeutica + Tiss
Antecedentes + Apache + Diagnosticos + Diario + Infeccoes + Sofa + Terapeutica + Tiss
Antecedentes + Apache + Diagnosticos + Diario + Infeccoes + Procedimentos + Terapeutica + Tiss
Antecedentes + Apache + Diagnosticos + Diario + Infeccoes + Procedimentos + Sofa + Tiss
Antecedentes + Apache + Diagnosticos + Diario + Infeccoes + Procedimentos + Sofa + Terapeutica
Antecedentes + Apache + Diagnosticos + Diario + Entrada(inicio) + Infeccoes + Sofa + Terapeutica
Antecedentes + Apache + Complicacoes + Diagnosticos + Diario + Infeccoes + Procedimentos + Terapeutica
Alta + Diagnosticos + Diario + Infeccoes + Procedimentos + Sofa + Terapeutica + Tiss
Alta + Apache + Diario + Infeccoes + Procedimentos + Sofa + Terapeutica + Tiss
Alta + Apache + Diagnosticos + Infeccoes + Procedimentos + Sofa + Terapeutica + Tiss
Alta + Apache + Diagnosticos + Diario + Procedimentos + Sofa + Terapeutica + Tiss
Alta + Apache + Diagnosticos + Diario + Infeccoes + Sofa + Terapeutica + Tiss
Alta + Apache + Diagnosticos + Diario + Infeccoes + Procedimentos + Terapeutica + Tiss
Alta + Apache + Diagnosticos + Diario + Infeccoes + Procedimentos + Sofa + Tiss
Alta + Apache + Diagnosticos + Diario + Infeccoes + Procedimentos + Sofa + Terapeutica
Alta + Apache + Diagnosticos + Diario + Follow-Up(inicio) + Infeccoes + Terapeutica + Tiss
Alta + Apache + Diagnosticos + Diario + Entrada(inicio) + Infeccoes + Sofa + Terapeutica
Alta + Apache + Complicacoes + Diagnosticos + Diario + Infeccoes + Procedimentos + Terapeutica
Alta + Apache + Complicacoes + Diagnosticos + Diario + Follow-Up(inicio) + Infeccoes + Terapeutica
Alta + Antecedentes + Diario + Infeccoes + Procedimentos + Sofa + Terapeutica + Tiss
Alta + Antecedentes + Diagnosticos + Infeccoes + Procedimentos + Sofa + Terapeutica + Tiss
Alta + Antecedentes + Diagnosticos + Diario + Procedimentos + Sofa + Terapeutica + Tiss
Alta + Antecedentes + Diagnosticos + Diario + Infeccoes + Sofa + Terapeutica + Tiss
Alta + Antecedentes + Diagnosticos + Diario + Infeccoes + Procedimentos + Terapeutica + Tiss
Alta + Antecedentes + Diagnosticos + Diario + Infeccoes + Procedimentos + Sofa + Tiss
Alta + Antecedentes + Diagnosticos + Diario + Infeccoes + Procedimentos + Sofa + Terapeutica
Alta + Antecedentes + Diagnosticos + Diario + Entrada(inicio) + Infeccoes + Sofa + Terapeutica

Alta + Antecedentes + Complicacoes + Diagnosticos + Diario + Infeccoes + Procedimentos + Terapeutica

Alta + Antecedentes + Apache + Infeccoes + Procedimentos + Sofa + Terapeutica + Tiss

Alta + Antecedentes + Apache + Diario + Procedimentos + Sofa + Terapeutica + Tiss

Alta + Antecedentes + Apache + Diario + Infeccoes + Sofa + Terapeutica + Tiss

Alta + Antecedentes + Apache + Diario + Infeccoes + Procedimentos + Terapeutica + Tiss

Alta + Antecedentes + Apache + Diario + Infeccoes + Procedimentos + Sofa + Tiss

Alta + Antecedentes + Apache + Diario + Infeccoes + Procedimentos + Sofa + Terapeutica

Alta + Antecedentes + Apache + Diario + Entrada(inicio) + Infeccoes + Sofa + Terapeutica

Alta + Antecedentes + Apache + Diagnosticos + Procedimentos + Sofa + Terapeutica + Tiss

Alta + Antecedentes + Apache + Diagnosticos + Infeccoes + Sofa + Terapeutica + Tiss

Alta + Antecedentes + Apache + Diagnosticos + Infeccoes + Procedimentos + Terapeutica + Tiss

Alta + Antecedentes + Apache + Diagnosticos + Infeccoes + Procedimentos + Sofa + Tiss

Alta + Antecedentes + Apache + Diagnosticos + Infeccoes + Procedimentos + Sofa + Terapeutica

Alta + Antecedentes + Apache + Diagnosticos + Entrada(inicio) + Infeccoes + Sofa + Terapeutica

Alta + Antecedentes + Apache + Diagnosticos + Diario + Sofa + Terapeutica + Tiss

Alta + Antecedentes + Apache + Diagnosticos + Diario + Procedimentos + Terapeutica + Tiss

Alta + Antecedentes + Apache + Diagnosticos + Diario + Procedimentos + Sofa + Tiss

Alta + Antecedentes + Apache + Diagnosticos + Diario + Procedimentos + Sofa + Terapeutica

Alta + Antecedentes + Apache + Diagnosticos + Diario + Infeccoes + Terapeutica + Tiss

Alta + Antecedentes + Apache + Diagnosticos + Diario + Infeccoes + Sofa + Terapeutica

Alta + Antecedentes + Apache + Diagnosticos + Diario + Infeccoes + Procedimentos + Tiss

Alta + Antecedentes + Apache + Diagnosticos + Diario + Infeccoes + Procedimentos + Terapeutica

Alta + Antecedentes + Apache + Diagnosticos + Diario + Infeccoes + Procedimentos + Sofa

Alta + Antecedentes + Apache + Diagnosticos + Diario + Entrada(inicio) + Sofa + Terapeutica

Alta + Antecedentes + Apache + Diagnosticos + Diario + Entrada(inicio) + Infeccoes + Terapeutica

Alta + Antecedentes + Apache + Diagnosticos + Diario + Entrada(inicio) + Infeccoes + Sofa

Alta + Antecedentes + Apache + Complicacoes + Diario + Infeccoes + Procedimentos + Terapeutica

Alta + Antecedentes + Apache + Complicacoes + Diagnosticos + Infeccoes + Procedimentos + Terapeutica

Alta + Antecedentes + Apache + Complicacoes + Diagnosticos + Diario + Infeccoes + Procedimentos

Subconjuntos de clique 9: 12 resultados

Antecedentes + Apache + Diagnosticos + Diario + Infeccoes + Procedimentos + Sofa + Terapeutica +

Tiss

Alta + Apache + Diagnosticos + Diario + Infeccoes + Procedimentos + Sofa + Terapeutica + Tiss

Alta + Antecedentes + Diagnosticos + Diario + Infeccoes + Procedimentos + Sofa + Terapeutica + Tiss

Alta + Antecedentes + Apache + Diario + Infeccoes + Procedimentos + Sofa + Terapeutica + Tiss

Alta + Antecedentes + Apache + Diagnosticos + Infeccoes + Procedimentos + Sofa + Terapeutica + Tiss

Alta + Antecedentes + Apache + Diagnosticos + Diario + Procedimentos + Sofa + Terapeutica + Tiss

Alta + Antecedentes + Apache + Diagnosticos + Diario + Infeccoes + Sofa + Terapeutica + Tiss

Alta + Antecedentes + Apache + Diagnosticos + Diario + Infeccoes + Procedimentos + Terapeutica + Tiss

Alta + Antecedentes + Apache + Diagnosticos + Diario + Infeccoes + Procedimentos + Sofa + Tiss

Alta + Antecedentes + Apache + Diagnosticos + Diario + Infeccoes + Procedimentos + Sofa + Terapeutica

Alta + Antecedentes + Apache + Diagnosticos + Diario + Entrada(inicio) + Infeccoes + Sofa + Terapeutica

Alta + Antecedentes + Apache + Complicacoes + Diagnosticos + Diario + Infeccoes + Procedimentos + Terapeutica

Subconjuntos de clique 10: 1 resultado

Alta + Antecedentes + Apache + Diagnosticos + Diario + Infeccoes + Procedimentos + Sofa + Terapeutica + Tiss

Recomendação :: Melhoria

Cada nó contido no(s) clique(s) de nível 10 deve conter links de fácil acesso para todos os outro nós do clique.

* As recomendações emitidas não levam em conta a lógica da aplicação e por isso devem sempre ser interpretadas tendo em atenção essas informações.

Aplicando o algoritmo “CUSTOM GRAPH MEDIAN” ao grafo obtemos o seguinte resultado:

Nº vezes origem	Nome do nó
16955	Entrada(inicio)
12370	Diario
11513	Terapeutica
9365	Entrada(grava=true)

9082	Apache
6577	Alta
4489	Sofa
3831	Tiss
3527	Entrada(grava=false)
1098	Antecedentes
278	Procedimentos
265	Infeccoes
247	Diagnosticos
122	Follow-Up(inicio)
107	Complicacoes
16	Enfermeiros(inicio)
12	Enfermeiros(grava=false)
6	FollowUp28
2	Enfermeiros(grava=true)
1	RegistodeSepsis

Recomendação :: Melhoria

Os nós “Entrada(inicio)”, “Diario” e “Terapeutica” representam mais de 50% das transições e devem ter acesso fácil a partir de qualquer nó.

* As recomendações emitidas não levam em conta a lógica da aplicação e por isso devem sempre ser interpretadas tendo em atenção essas informações.

Devido à inexistência de informação acerca de qual o nó de saída para esta aplicação não foi possível aplicar o algoritmo “INVALID PATH” para este grafo.

Avaliações dos peritos e responsáveis

Para obtermos uma base de comparação para os nossos resultados foi contactado um perito nos fluxos de informação da aplicação breast.care e o responsável pela aplicação HSJ.ICU. Relativamente à aplicação intensive.care não foi possível entrar em contacto com os responsáveis.

Para a discussão dos resultados relativos à aplicação breast.care entrámos em contacto com um perito nos fluxos de informação da aplicação breast.care utilizada no Hospital de São João. De modo a fazer a ponte entre os conhecimentos do perito e os algoritmos matemáticos que implementámos foram-lhe descritos esses algoritmos com exemplos de resultados abstractos e foi-lhe perguntado, tendo em conta os resultados dos seus estudos, qual seria o resultado esperado se os algoritmos fossem aplicados em logs do breast.care. Em resposta ao algoritmo dos “CLIQUES” o perito dividiu os resultados em dois grupos devido a ter identificado também dois tipos de utilizadores:

Para o grupo de utilizadores ligados mais à parte de diagnóstico as páginas mais acedidas seriam: a página da ecografia, página da mamografia e página do relatório imagiológico.

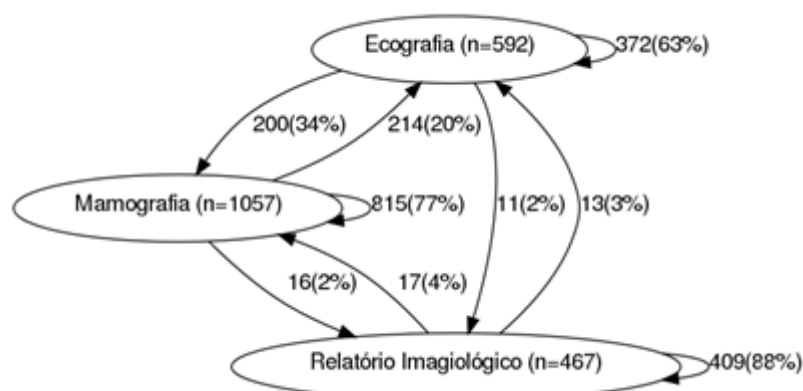


Ilustração 30 – Grafo da aplicação breast.care relativo ao diagnóstico e associado à resposta para o algoritmo dos CLIQUES (Silva, 2010)

Para o grupo de utilizadores ligados mais à parte do tratamento as páginas mais acedidas seriam: a página da consulta de diagnóstico, a página dos antecedentes das pacientes, a página de dados pessoais e a página da avaliação clínica.

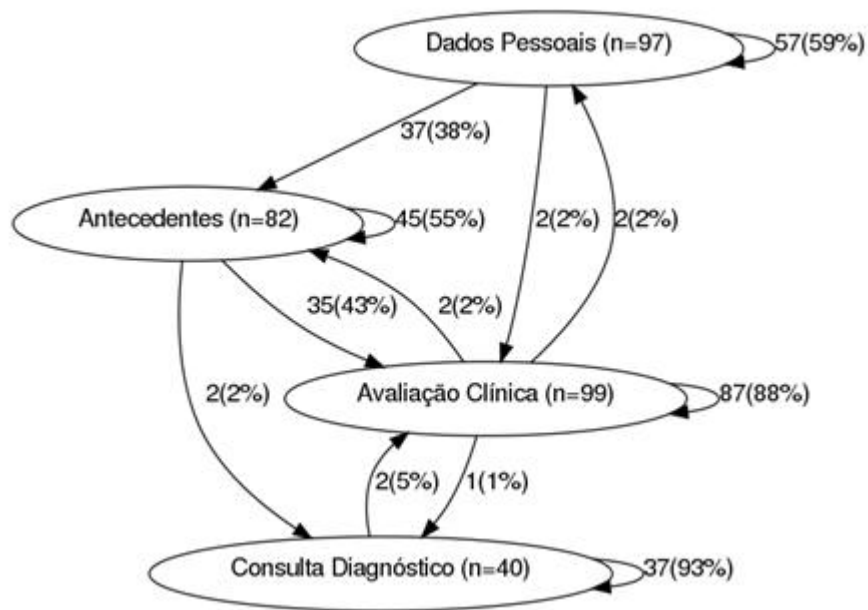


Ilustração 31 – Grafo da aplicação breast.care relativo ao tratamento e associado à resposta para o algoritmo dos CLIQUES (Silva, 2010)

Relativamente ao algoritmo “CUSTOM GRAPH MEDIAN” a resposta do perito foi bastante parecida com a do algoritmo “CLIQUES” com as seguintes diferenças:

- A página de pesquisa do paciente foi referida com uma das principais nos resultados esperados para este algoritmo;
- As páginas foram descritas por ordem decrescente de transições de saída;

Sendo assim, os resultados referidos pelo perito foram novamente divididos em dois grupos. Para o grupo de utilizadores ligados mais à parte de diagnóstico as páginas ordenadas seriam: página da ecografia e página da mamografia.

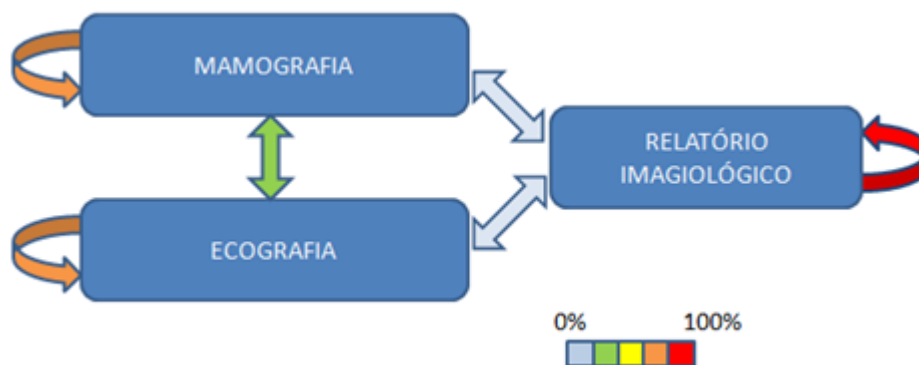


Ilustração 32 - Grafo da aplicação breast.care relativo ao diagnóstico e associado à resposta para o algoritmo “CUSTOM GRAPH MEDIAN” (Silva, 2010)

Para o grupo de utilizadores ligados mais à parte do tratamento as páginas ordenadas seriam: a página de dados pessoais, a página dos antecedentes das pacientes, a página da avaliação clínica e a página da consulta de diagnóstico.

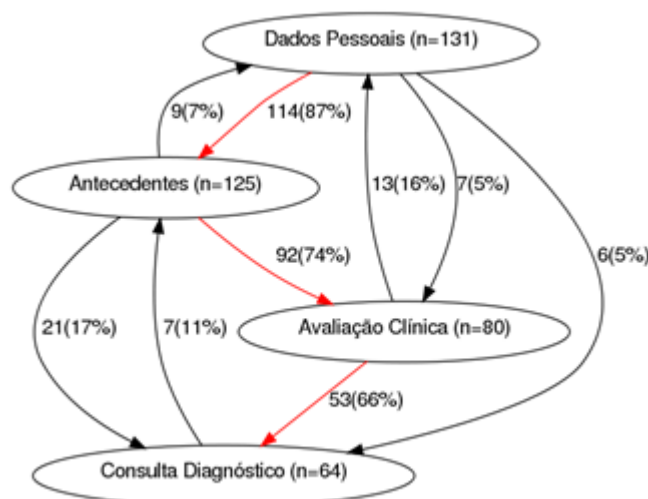


Ilustração 33 – Grafo da aplicação breast.care relativo ao tratamento e associado à resposta para o algoritmo “CUSTOM GRAPH MEDIAN” (Silva, 2010)

A aplicação HSJ.ICU é a plataforma que suporta o módulo breast.care pelo que é importante perceber o seu funcionamento de modo a tirar algumas conclusões quanto aos resultados obtidos. Foi contactado o responsável por esta aplicação e foram-lhe colocadas algumas questões relativas à arquitectura e funcionamento da aplicação. Desta interacção foram retiradas várias informações:

- Existem 20 acções que guardam informação nos logs, entre elas acessos a páginas, cliques em menus, acesso a webservices e vários tipos de visualização de relatórios;
- Existe apenas uma entrada no sistema que inicia a sessão e propaga as permissões do utilizador pelos módulos existentes;
- Esta sessão pode terminar explicitamente por pedido do utilizador ao aceder à página de saída, se fechar o browser ou se não houver actividade na aplicação durante 10 minutos;
- Não existe paginação na página de relatórios, existindo no entanto uma filtragem por serviço. Esta filtragem é uma acção do servidor, gravando novo log de acesso cada vez que isto acontece;
- Os relatórios são visualizados numa “*iframe*”;
- A página de saída redirecciona o utilizador automaticamente para a página de entrada;
- O registo do log é efectuado durante o processamento da página;

6. Discussão

Tal como vimos no capítulo anterior, o “Viajante” foi capaz de gerar vários grafos a partir dos logs das aplicações. A estes grafos foram aplicados algoritmos da teoria de grafos que levaram à emissão de recomendações automáticas. Foi recolhida também a opinião de um perito e de um responsável por uma das aplicações.

Neste capítulo vamos comparar os resultados obtidos pelo “Viajante” com as opiniões do perito resultantes de estudos passados. Os vários grafos vão também ser interpretados tentando encontrar neles características determinantes para a compreensão do fluxo de informação nas aplicações estudadas. Antes de tudo isso vamos começar por analisar o “Viajante” em termos das suas funcionalidades e do seu funcionamento.

A Aplicação “Viajante”

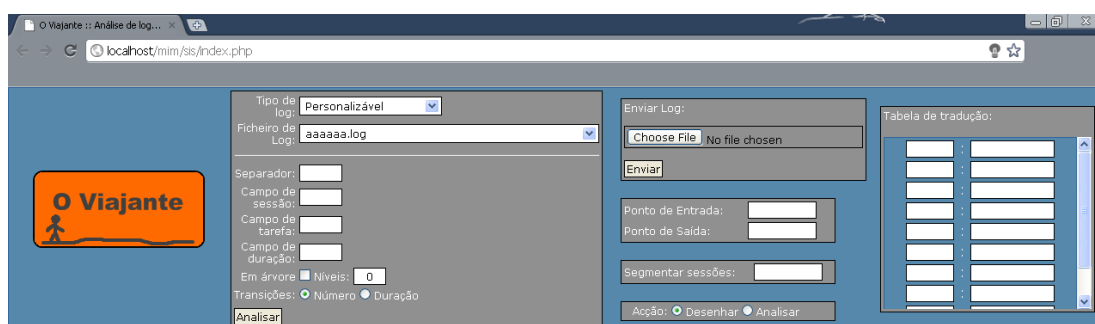


Ilustração 34 - Screenshot da aplicação “Viajante”

Da recolha de métodos de avaliação feita no estado da arte conseguimos depreender que as desvantagens nos métodos actuais prendem-se com dificuldades de recolha, organização, análise ou interpretação dos dados recolhidos e com factores humanos, onde o facto de envolver pessoas pode influenciar o resultado. Ao mesmo tempo existem vantagens claras em termos de riqueza dos dados ao envolver as pessoas no processo, visto que são elas quem mais interage com o objecto de estudo, e mesmo quem ditam se a aplicação tem sucesso. Outra vantagem é que ao recolher muitos

dados de forma rápida respeitantes a várias pessoas é possível trabalhá-los em conjunto ou combiná-los para obter informações. Com o objectivo de reunir estas vantagens não deixando que as desvantagens dos métodos actuais interfiram, o “Viajante” analisou ficheiros de log usando assim informações de utilização vindas directamente das aplicações em estudo não envolvendo no entanto pessoas no processo. Estes ficheiros continham informações sobre a utilização das aplicações, que eram gravadas automaticamente, segmentadas por sessão de utilização e com diversas informações acerca das acções tomadas pelos utilizadores. Adoptando esta abordagem conseguimos envolver os utilizadores sem adulterar os resultados e, embora tendo em conta as desvantagens associadas aos ficheiros de log, conseguimos obter muitos dados de uma fonte fiável e barata.

O “Viajante” combinou os dados recolhidos e fez análises utilizando teoria de grafos que permitiram uma interpretação dos dados de uma forma mais simples e intuitiva, gerando ele próprio recomendações automáticas para os casos em que se detectaram anomalias ou possíveis melhorias nas aplicações. É de lembrar que estas recomendações têm que ser contextualizadas à aplicação estudada pois são recomendações genéricas que não têm qualquer noção do workflow necessário para a aplicação funcionar correctamente.

Quanto ao funcionamento e arquitectura, a opinião do autor é que o “Viajante” cumpre os requisitos que foram propostos. Ao adoptar uma arquitectura baseada na web o “Viajante” fica disponível de qualquer sítio e na maioria das plataformas e sistemas operativos. A aplicação desenvolvida também não necessita de requisitos mínimos elevados, tanto devido à arquitectura adoptada como à optimização de código que foi uma preocupação presente desde o início do seu desenvolvimento. Usar software open-source foi também uma vantagem pois deste modo a aplicação pode ser facilmente auditada e constantemente melhorada sendo possível adicionar novas funcionalidades de forma simples. Foi ainda utilizada uma biblioteca personalizável de desenho de grafos (GraphViz) que contribuiu bastante para as capacidades do “Viajante”.

Em termos de funcionalidades a aplicação possui várias características, sendo uma das principais a capacidade de gerar grafos de vários tipos. Estes diferentes tipos de

grafos permitem obter uma visão de vários ângulos acerca da aplicação estudada. Ao passo que os grafos cíclicos com número de transições ilustram o uso geral da aplicação em estudo, evidenciando as relações entre os seus nós e apontando as transições mais efectuadas, os grafos em árvore ajudam-nos a perceber os percursos dos utilizadores à medida que a usam, mostrando informações sobre probabilidades de transições. Apesar do “Viajante” o permitir, não foram desenhados grafos cíclicos com demora nas transições pois o tipo de logs a que tivemos acesso não continha informação sobre esses valores. De qualquer forma este tipo de grafos pode ser útil para se perceber quais as transições mais demoradas, equivalendo possivelmente às mais exigentes em termos de processamento.

Ao implementar a coloração automática de transições o “Viajante” oferece uma possibilidade de interpretação dos grafos mais intuitiva direccionando o observador para as transições mais significativas. Existe também a possibilidade de aplicar filtros de leitura para que sejam desenhados apenas os nós desejados e filtros de sessão para que sejam consideradas apenas transições de determinadas sessões de utilização.

Para optimização do processo existe a possibilidade de enviar os ficheiros de log para o servidor da aplicação. Isto vai permitir ganhar tempo nas análises visto que o ficheiro, uma vez no servidor, é tratado localmente de forma mais rápida, fazendo ainda com que o tráfego de dados entre a aplicação e o computador cliente seja reduzido apenas ao necessário.

Desde que sejam respeitados os pré-requisitos identificados no capítulo do desenho do sistema, quaisquer ficheiros de log podem ser enviados para análise. A aplicação foi construída de forma flexível e configurável de modo a permitir especificar o formato do ficheiro de log a analisar.

Ainda que tenha sido identificada como uma funcionalidade relevante para a aplicação, a tabela de tradução não foi utilizada no decorrer das análises visto que os logs a que tivemos acesso continham informação textual descritiva das acções efectuadas. Sendo assim, embora isto não invalide a relevância da funcionalidade, nestes ficheiros de log não foi necessária a utilização desta tabela.

Outra característica muito importante é a capacidade de fazer análises teóricas que, como veremos abaixo, se revelou imprescindível para tirarmos algumas das conclusões. Estas análises são definidas pelos algoritmos da teoria de grafos, algoritmos estes que, desde que codificados na mesma linguagem que o “Viajante”, podem ser adicionados facilmente ao conjunto de algoritmos de análise, ficando completamente integrados na aplicação. Os resultados destes algoritmos são automaticamente interpretados pela aplicação com o objectivo de gerar recomendações de melhoria ou avisos de erro. Ainda que devam ser sempre contextualizadas, elas deverão ajudar o utilizador na interpretação dos resultados da sua análise.

Revista a aplicação desenvolvida, passamos à discussão das análises efectuadas.

As análises

Alguns dos grafos gerados são de interpretação simples pois contêm poucos nós e embora possam ter muitas ligações entre eles a distinção de cores ajuda a identificar transições e caminhos mais habituais. No entanto para grafos mais complexos a análise usando teoria de grafos foi crucial para que se pudesse fazer sentido dos dados. Nesta secção é apresentada uma possível interpretação e validação dos resultados obtidos e descritos na secção anterior através da comparação dos mesmos com resultados obtidos num estudo anterior para uma das aplicações.

breast.care

O “Viajante” gerou vários grafos relativos ao HSJ.ICU+breast.care. O primeiro grafo “Ilustração 21 - Grafo cíclico completo da aplicação ICU+breast.care” pretende ilustrar o funcionamento da aplicação completa. Como podemos ver o grafo não é muito intuitivo devido a ter muitas transições que a partir de certo ponto se tornam impossíveis de seguir ou até de perceber qual o seu valor associado. Visualizada num ecrã de computador esta questão é atenuada mas se redimensionarmos e imprimirmos a imagem para efeitos de relatório ou análise em suporte físico torna-se difícil de a

interpretarmos. De qualquer forma uma das claras vantagens é a associação de cores chamando claramente a atenção para as transições mais relevantes. Visto que, devido à sua complexidade, este grafo não é de fácil interpretação e como existia conhecimento de base sobre quais os nós da aplicação base e quais os nós do módulo breast.care foram aplicados filtros para obter dois grafos distintos: um apenas dos nós de base do HSJ.ICU (Ilustração 22 – Grafo cíclico relativo apenas à aplicação ICU) e outro apenas dos nós do módulo breast.care (Ilustração 23 - Grafo cíclico relativo apenas à aplicação breast.care). Desta feita é possível fazer a interpretação dos grafos, ainda que manualmente, de forma bastante mais simples.

De seguida foram gerados quatro grafos de sessões individuais. Os primeiros dois são relativos a utilizadores da aplicação base pois vemos que nenhum nó relativo ao breast.care foi acedido. Os outros dois são relativos a utilizadores do módulo breast.care. Enquanto que nos primeiros dois grafos os utilizadores fazem uso da aplicação para pesquisar utentes, listar e visualizar relatórios, nos outros dois grafos os utilizadores usam nós de base para pesquisar utentes e listar relatórios, saltando depois para os nós do breast.care dados pessoais, antecedentes e avaliação clínica dos utentes. É interessante como a utilização da aplicação é distinta entre os dois tipos de sessão mas idêntica nos dois primeiros casos e nos dois últimos. Com a ajuda de um algoritmo de comparação de grafos seria possível encontrar casos de utilização idênticos automaticamente e talvez acrescentar alguma capacidade de clustering ao “Viajante”. Esta preocupação é referida no trabalho futuro como uma das possíveis melhorias para a aplicação desenvolvida.

O último grafo desenhado (Ilustração 28 - Grafo em árvore da aplicação HSJ.ICU com corte no nível 5) é um grafo em árvore referente apenas à aplicação de base com corte no nível 5. Este tipo de grafos permite ao observador perceber especificamente quais são os percursos dos utilizadores nas aplicações. Observando o grafo podemos ver que, imediatamente após ter encontrado o utente e estando agora na listagem de relatórios, a probabilidade de a seguir visualizar um relatório é aproximadamente 29% enquanto a probabilidade de listar novamente relatórios é aproximadamente 67%. Podemos concluir então que após encontrarem o paciente, a maioria dos utilizadores da aplicação passa duas vezes pela listagem de relatórios antes de visualizarem um relatório. Este tipo de conclusão nunca pode ser tirado a partir de um grafo cíclico

pois nesse tipo de grafos não existe a noção de sequência e portanto o tipo de grafo em árvore foi considerado interessante e objecto de futuros desenvolvimentos.

Passando aos algoritmos vamos tomar em conta a resposta do perito para o algoritmo dos “CLIQUES”. Podemos deprender que existem dois cliques importantes na aplicação:

- Clique de nível 3 composto pelos nós “ecografia”, “mamografia” e “relatório imagiológico”;
- Clique de nível 4 composto pelos nós “dados pessoais”, “antecedentes”, “avaliação clínica” e “consulta diagnóstico”;

Dos resultado obtidos automaticamente pelo “Viajante” e apresentados na secção anterior destacamos os seguintes cliques:

Subconjuntos de clique 3

breastcare|Antecedentes + breastcare|AvaliaçãoClínica + breastcare|DadosPessoais

breastcare|Antecedentes + breastcare|AvaliaçãoClínica + breastcare|ConsultadeDiagnóstico

Subconjunto de clique 4

Login + breastcare|Antecedentes + breastcare|AvaliaçãoClínica + breastcare|DadosPessoais

Apesar de não ter sido encontrado um clique que confirmasse a 100% a opinião do perito existem dois cliques de nível 3 que contêm 3 das 4 acções apontadas e qualquer acção contida nesses cliques é precisamente uma das acções indicadas. É de lembrar que o algoritmo dos cliques obriga a que exista uma ligação bidireccional entre quaisquer dois nós do subgrupo. A inexistência de um clique que contenha os 4 nós indicados no estudo pode ser justificada pela inexistência de apenas uma das ligações. Neste caso a ligação inexistente é entre o nó “DadosPessoais” e o nó “ConsultadeDiagnóstico” pois ainda que ambos tenham ligações com os outros nós, sempre que um está presente o outro não está. Desenhando o grafo apenas desses 4 nós obtemos o seguinte:

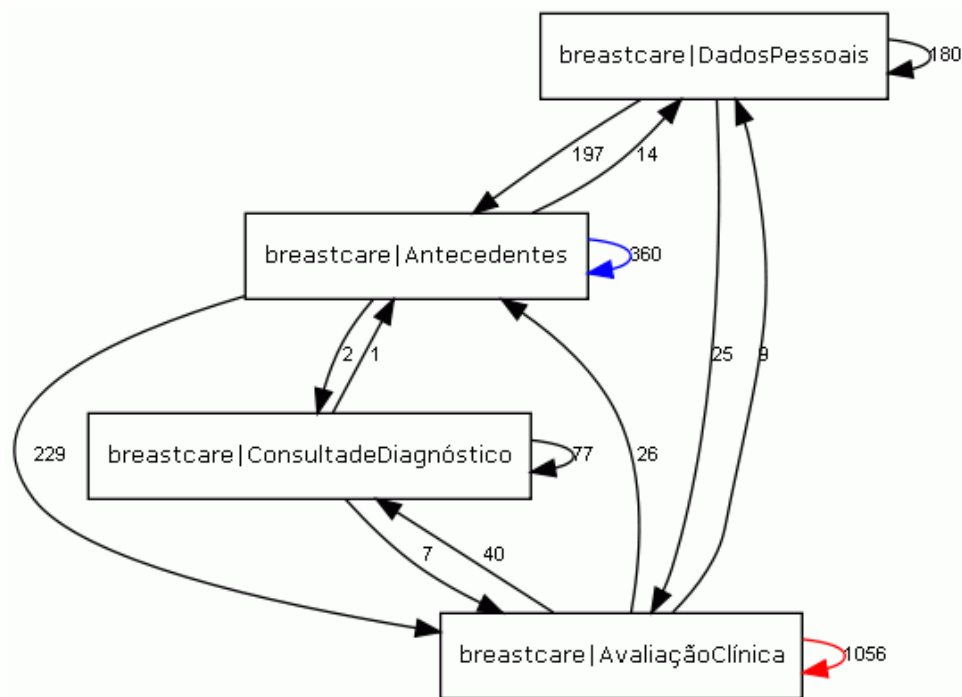


Ilustração 35 - Grafo do módulo breast.care filtrado a 4 nós

De facto não existem transições directas entre “ConsultadeDiagnóstico” e “DadosPessoais”. De qualquer forma, e apesar de não ter sido encontrado o clique entre os quatro nós indicados, é possível verificar que existem percursos definidos entre eles assim como apresentados no estudo anterior. É também de lembrar que o log analisado que deu origem a este grafo é limitado no tempo e é possível que algumas transições possam não ter ocorrido no decorrer desse período.

Quanto ao clique de nível 3 indicado pelo perito e composto pelos nós “ecografia”, “mamografia” e “relatório imagiológico” não foi identificado automaticamente nenhum clique idêntico. O único nó que é apresentado relacionado com esse clique é o “breastcare|RelatórioImagiologia”. Após verificação directamente no ficheiro de log, constatámos que não existem acções relativas aos nós “ecografia” e “mamografia”, pelo que não podem aparecer também nos grafos desenhados a partir do ficheiro. Esta particularidade podia dever-se, por exemplo, ao facto de o período em que foram recolhidos dados para o log não ter abrangido estas acções mas visto que essas acções foram apontadas como das mais executadas seria pouco provável não as ter encontrado na amostra do log utilizada. De modo a tentar perceber esta situação o responsável pela aplicação foi novamente contactado e questionado sobre

estes dois nós. Foi então esclarecido que esses dois nós foram fundidos com o “breastcare|Relatório de Imagiologia”, deixando portanto de existir isoladamente.

As recomendações resultantes deste algoritmo para a aplicação completa foram:

Cada nó contido no(s) clique(s) de nível 5 deve conter links de fácil acesso para todos os outros nós do clique.

Sendo que o clique de nível 5 é este:

Listagem de relatórios + Login + Logout + Pesquisa Utente + Visualização de relatório

A recomendação vem no sentido de que se estes 5 nós têm ligações entre todos eles então essas opções de transição que devem existir na aplicação devem também estar organizada de forma a que seja fácil aceder a elas, sempre com a salvaguarda, no entanto, que a recomendação deve ser contextualizada na aplicação.

Quanto ao algoritmo “CUSTOM GRAPH MEDIAN” o perito referenciou dois blocos de nós como resposta:

- Dados Pessoais -> Antecedentes -> Avaliação Clínica -> Consulta Diagnóstico;
- Ecografia e Mamografia.

Como vimos nos nossos resultados, o Login foi o nó que originou mais ligações para outros nós (114018) e como podemos ver pelo grafo ilustrado na “Ilustração 22 – Grafo cíclico relativo apenas à aplicação ICU” a maior parte delas é para a Pesquisa Utente. Sendo o “Login” o nó de entrada da aplicação, o facto de ter mais transições a partir dele do que qualquer outro nó indica que os utilizadores fazem uso da aplicação de uma forma relativamente fluida sem alternar muito nos nós internos pois, caso isso acontecesse, cada sessão de utilizador teria apenas uma passagem do “Login” para outro nó mas diversas passagens de nós dentro da aplicação de uns para

os outros. Isto faria com que provavelmente muitos desses nós internos tivessem um valor no “CUSTOM GRAPH MEDIAN” acima do valor do “Login”.

Ao contrário do que poderia ser de esperar, o segundo nó com valor mais alto na lista de “CUSTOM GRAPH MEDIAN” não é o nó para onde o “Login” mais faz transições (PesquisaUtente) mas é sim o “Listagemderelatórios” com mais do que 8700 transições de saída do que o “PesquisaUtente”. Uma justificação lógica para isto é que um utente tem vários relatórios associados mas um relatório é apenas de um utente. Isto faz com que quando o utente seja encontrado o utilizador da aplicação salte para o “Listagemderelatórios” mas por cada uma destas acções seja depois possível fazer vários saltos do “Listagemderelatórios” para o “Visualizaçãodorelatório” à medida que o utilizador visualiza os diversos relatórios do utente encontrado.

Quanto aos nós relativos ao breast.care estes começam um pouco abaixo do início da lista mas há uma grande similaridade entre nós encontrados e os nós que o perito referenciou. Os nós encontrados foram:

Tabela 10 - Comparação entre os nós relativos ao breast.care encontrados automaticamente pelo algoritmo “CUSTOM GRAPH MEDIAN” e os referenciados pelo perito

Nós encontrados automaticamente	Nós referenciados pelo perito
...	Dados Pessoais
689 - breastcare RelatóriodeImagiologia	Antecedentes
366 - breastcare DadosPessoais	Avaliação Clínica
284 - breastcare Antecedentes	Consulta Diagnóstico
157 - breastcare AvaliaçãoClínica	
42 - Logout	
24 - breastcare ConsultadeDiagnóstico	
...	

De facto o nó relativo ao breast.care de onde saem mais ligações é o RelatóriodeImagiologia, nó que não foi referido pelo perito para esta análise. Uma possível razão para que o perito não tenha referido este nó é que a maioria das transições deste nó vão para o nó “PesquisaUtente”, e esse é um nó do ICU não do

módulo breast.care. O nó de “Logout” aparece nesta lista devido às suas transições para o nó de “Login”, uma transição relativamente comum e legítima na aplicação. Tomando em conta estas considerações a lista de nós encontrados automaticamente é uma correspondência perfeita com a lista referenciada pelo perito, tanto em termos de nós como em termos de ordem de cada nó.

Mais uma vez quanto aos nós “ecografia” e “mamografia” indicados pelo perito estes foram fundidos com o nó “breastcare|RelatórioImagiologia” e portanto nenhuma conclusão pode ser retirada nesta questão.

A recomendação final resultante deste algoritmo foi:

Os nós “Login” e “Listagemderelatórios” representam mais de 50% das transições e devem ter acesso fácil a partir de qualquer nó.

Esta recomendação vem no sentido de que se estes nós são responsáveis por mais de 50% das transições de toda a aplicação isso significa que as suas acções são bastante requisitadas e, sendo assim, qualquer nó deveria ter um acesso fácil a estes nós. Mais uma vez a informação deve ser contextualizada porque pode, por exemplo, haver nós que não permitam fazer a transição directa para estes nós. O limite dos 50% é um valor configurável e neste caso poderia ter sido aumentado visto que o “Login” é o nó de entrada e pode não fazer muito sentido todos os outros nós terem uma ligação para ele mesmo que este represente uma grande parte das transições da aplicação.

O algoritmo de verificação alertou para alguns problemas relacionados com o nó “Logout”. Existem transições do “Logout” para todos os nós da aplicação base. Sendo que o log está a ser interpretado no seguimento temporal das acções, que é orientado à sessão e que as sessões são únicas, é pouco provável isto ser um problema da análise efectuada como, por exemplo, se a interpretação tivesse sido feita orientada aos endereços IP que podem ser dinâmicos. Isto significa que no decorrer das sessões, em várias instâncias diferentes, utilizadores fizeram logout e de seguida visitaram outra página dentro da aplicação que não é a página de entrada (Login). Sem outro tipo de análise, nomeadamente do próprio workflow interno da aplicação, não é possível

identificar a causa deste comportamento mas algumas possibilidades para isto podem ser:

- Um problema de segurança na aplicação que permite ver páginas sem uma sessão autenticada;
- O logout pode ter algum problema que evita que este seja efectivamente feito;
- O sistema de gravação de logs está programado de forma a que primeiro grava a acção chamada e só depois é verificada a condição que permite ou proíbe a execução da acção;

Em qualquer um dos casos serve como aviso e é aconselhável uma verificação da aplicação de modo a validar ou excluir estas e identificar outras possíveis causas para o acontecido. O clique de nível 5 anteriormente encontrado confirma o resultado do algoritmo de verificação pois a existência deste clique significa não só que de todos os nós do clique houve transições para o “Logout” mas também que do “Logout” houve directamente transições para todos os outros nós do clique. O grafo “Ilustração 22 – Grafo cíclico relativo apenas à aplicação ICU” ilustra esse clique.

As recomendações resultantes da aplicação deste algoritmo foram:

Há uma falha. Depois de acederem ao nó Logout, 66 utilizador(es) fizeram a transição Logout_Logout

Há uma falha. Depois de acederem ao nó Logout, 6 utilizador(es) fizeram a transição Logout_PesquisaUtente

Há uma falha. Depois de acederem ao nó Logout, 1 utilizador(es) fizeram a transição Logout_Listagemderelatórios

Há uma falha. Depois de acederem ao nó Logout, 2 utilizador(es) fizeram a transição Logout_Visualizaçãodorelatório

O “INVALID PATH” é um algoritmo simples que apenas testa se existem ligações entre conjuntos de dois nós. Sendo que o “Login” foi marcado como nó de entrada na

aplicação e o “Logout” como nó de saída não deveria haver transições com origem no nó “Logout” para qualquer outro nó que não fosse o “Login”.

intensive.care

Mais uma vez o desenho do grafo completo da aplicação revelou-se pouco útil devido à sua complexidade, embora a distinção de cores nas transições tenha sido novamente uma característica importante. Nesta aplicação o foco não foi tanto nas representações gráficas dos grafos mas sim na aplicação de algoritmos, visto que não existiam dados anteriores para confirmação. Ainda foi desenhado um grafo em árvore com corte no nível 5 tal como na aplicação HSJ.ICU mas verificou-se que a quantidade de nós de entrada era tão grande que a árvore se tornava ilegível. Dezasseis dos vinte nós totais da aplicação serviram como ponto de partida. Esta questão também pode estar relacionada com a forma como os nós foram gravados no ficheiro de log pois, por exemplo, existem três nós chamados “Entrada” mas seguidos do que parecem ser parâmetros da aplicação (“grava” e “inicio”). Sem mais informações quanto ao funcionamento e arquitectura da aplicação não é possível retirar nenhuma conclusão quanto a esta situação.

Passando aos algoritmos, ao correr o algoritmo dos “CLIQUEs” nesta aplicação obtivemos um clique máximo de nível 10. Esse clique contém os seguintes nós:

Alta + Antecedentes + Apache + Diagnosticos + Diario + Infeccoes + Procedimentos + Sofa +
Terapeutica + Tiss

Visto que não temos forma de confirmar este resultado por uma fonte externa à semelhança do que aconteceu na aplicação breast.care, foi tomada uma abordagem mais crítica, subjectiva e pessoal para tentar interpretar este resultado. Ao identificar todos os nós que não constam do clique podemos tentar encontrar alguma semelhança entre eles. Como os únicos dados que temos sobre esse nós são o nome, a semelhança pesquisada foi em termos de objectivo esperado. Os nós fora do clique máximo são:

Complicacoes + Enfermeiros(grava=false) + Enfermeiros(grava=true) + Enfermeiros(inicio) +

Entrada(grava=false) + Entrada(grava=true) + Entrada(inicio) + Follow-Up(inicio) + FollowUp28 +
RegistodeSepsis

Desde já podemos observar que existem vários nós com nomes idênticos (Enfermeiros, Entrada, Follow-up) e com informações que parecem ser parâmetros (grava, inicio). Para tentarmos perceber onde se inserem estes nós, calculámos o clique máximo onde o nó se encontra obtendo assim o conjunto de nós que se relaciona bidireccionalmente com esse nó:

Complicacoes: clique máximo de nível 9
Enfermeiros(grava=false): clique máximo de nível 2
Enfermeiros(grava=true): clique máximo inexistente
Enfermeiros(inicio): clique máximo de nível 2
Entrada(grava=false): clique máximo de nível 3
Entrada(grava=true): clique máximo de nível 3
Entrada(inicio): clique máximo de nível 9
Follow-Up(inicio): clique máximo de nível 8
FollowUp28: clique máximo de nível 2
RegistodeSepsis: clique máximo inexistente

Se tivermos em conta o clique máximo de cada um podemos definir dois grupos bastante distintos neste conjunto de nós:

- Complicacoes, Entrada(inicio) e Follow-Up(inicio) com 8 e 9 ligações bidireccionais;
- Todos os outros nós com apenas 2, 3 ou nenhuma ligação bidireccional;

Visto que o nível do clique máximo dos nós do primeiro grupo se aproxima bastante do nível do clique máximo do grafo completo podemos tentar perceber quais são as relações entre eles.

O clique máximo envolvendo o nó “Complicacoes” (nível 9) é constituído também pelos nós “Alta“, “Antecedentes“, “Apache“, “Diagnosticos“, “Diario“, “Infeccoes“, “Procedimentos“ e “Terapeutica“. Para o clique máximo do grafo faltam ligações bidireccionais para os nós “Sofa” e “Tiss”.

O clique máximo envolvendo o nó “Entrada(inicio)” (nível 9) é constituído também pelos nós “Alta”, “Antecedentes”, “Apache”, “Diagnosticos”, “Diario”, “Infecoes”, “Sofa”, “Terapeutica”. Para o nó máximo do clique faltam ligações bidireccionais para os nós “Procedimentos” e “Tiss”.

O nó “Follow-Up(inicio)” (nível 8) aparece em dois cliques máximos. O primeiro é constituído também pelos nós “Alta”, “Apache”, “Diagnosticos”, “Diario”, “Infecoes”, “Terapeutica”, “Tiss” faltando os nós “Antecedentes”, “Procedimentos” e “Sofa” para o clique máximo do grafo completo. O segundo é constituído também pelos nós “Alta”, “Apache”, “Complicacoes”, “Diagnosticos”, “Diario”, “Infecoes”, “Terapeutica” faltando os nós “Antecedentes”, “Procedimentos”, “Sofa” e “Tiss” para o clique máximo do grafo completo, embora este último tenha ligação com o nó “Complicacoes”, um dos nós que não consta do clique máximo do grafo.

Ao agregar estes quatro cliques obtemos todos os nós do clique máximo do grafo completo e ainda o nó “Complicacoes”. Analogamente à situação encontrada no breast.care onde o clique indicado pelo perito não foi confirmado na totalidade, basta a não existência de uma das ligações em cada um destes cliques para fazer com que os nós não sejam abrangidos no clique máximo. De qualquer forma, estes nós têm um conjunto de ligações muito próximas às do clique máximo do grafo completo. Ao contrário destes nós, os nós do segundo conjunto acima encontrado que têm apenas 2, 3 ou nenhuma ligação bidireccional estão mais isolados:

Os nós “Enfermeiros(grava=true)” e “RegistodeSepsis” não têm qualquer ligação bidireccional podendo indicar que são nós de sistema (inacessíveis ao utilizador) ou que são nós de passagem em acções que são efectuadas sequencialmente.

Os nós cujo clique máximo onde aparecem é de nível 2 são o “Enfermeiros(grava=false)”, o “Enfermeiros(inicio)” e o “FollowUp28”.

Sendo de nível 2 o clique contém apenas 2 nós e uma ligação bidireccional entre eles. Os nós “Enfermeiros(inicio)” e “Enfermeiros(grava=false)” têm uma transição bidireccional entre eles sendo precisamente esse o clique máximo para ambos os nós.

Se desenharmos o grafo com os nós que se relacionam directamente com estes dois obtemos o seguinte grafo:

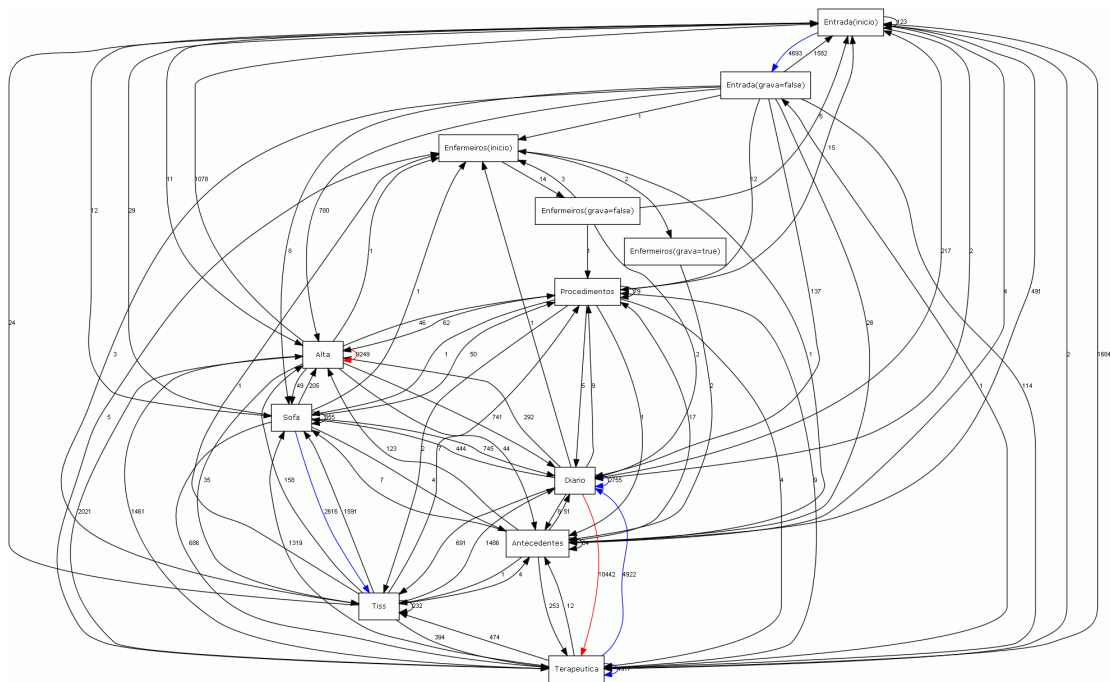


Ilustração 36 - Vizinhaça directa dos nós "Enfermeiros(inicio)" e "Enfermeiros(grav=false)"

O nó “Enfermeiros(inicio)” é um nó centralizador das transições dos nós relacionados. A este nó chegam 8 ligações e saem apenas 2. Em contra-partida, ao nó “Enfermeiros(grav=false)” chega apenas 1 ligação enquanto saem 4. De facto estes nós acusam um comportamento diferente dos nós que constituem os cliques de nível mais alto. Não são incluídos no conjunto de nós em que o utilizador navega livremente saltando de qualquer um para qualquer outro pois nestes dois existe um sentido bem definido: O “Enfermeiros(inicio)” é um nó de chegada e o “Enfermeiros(grav=false)” um nó de saída.

Do conjunto de nós que não pertencem ao clique máximo encontrado restam apenas os nós “Entrada(grav=false)” e “Entrada(grav=true)” ambos de nível 3. Os cliques correspondentes são:

- Entrada(grav=false) + Entrada(inicio) + Terapeutica
- Antecedentes + Entrada(grav=true) + Entrada(inicio)

Devido ao seu nome podemos ser levados a crer que eles correspondem à mesma página da aplicação embora ela receba um parâmetro diferente. Se partirmos deste pressuposto, então existe ainda o nó “Entrada(inicio)” que também pode equivaler à mesma página desta vez sem receber parâmetro nenhum. Nesse caso teríamos apenas um clique de nível 3 Antecedentes-Entrada(*)-Terapeutica. Sem mais informações no entanto não nos é possível perceber por que razão a página de “Entrada(*)” está apenas relacionadas bidireccionalmente com os nós “Terapeutica” e “Antecedentes”. Seria interessante perceber o porquê desta relação através de entrevistas com utilizadores, peritos ou responsáveis deste sistema.

A recomendação obtida por aplicação deste algoritmo foi:

Cada nó contido no(s) clique(s) de nível 10 deve conter links de fácil acesso para todos os outros nós do clique.

Sendo que o clique de nível 10 é o seguinte:

Alta + Antecedentes + Apache + Diagnosticos + Diario + Infecoes + Procedimentos + Sofa + Terapeutica + Tiss

Analogamente à explicação desta recomendação para o breast.care qualquer destes nós deve ter um fácil acesso à opção de transição para todos os outros nós do clique.

O algoritmo “CUSTOM GRAPH MEDIAN” revelou, tal como no estudo da aplicação breast.care, que o nó mais central é o nó de entrada, neste caso o “Entrada(inicio)”, indicando mais uma vez que os utilizadores fazem um percurso mais rectilíneo do que circular na aplicação. Um facto curioso é que os únicos nós que estão entre as posições iniciais da lista e que não fazem parte do clique máximo encontrado anteriormente são os relativos à entrada (“Entrada(inicio)”, “Entrada(grava=true)” e “Entrada(grava=false)”). De resto todos os nós que não pertencem ao clique máximo estão nas posições finais da tabela. Isto indica que os nós do clique para além de serem os mais interligados são também os mais utilizados.

A recomendação resultante da aplicação deste algoritmo foi:

Os nós “Entrada(inicio)”, “Diario” e “Terapeutica” representam mais de 50% das transições e devem ter acesso fácil a partir de qualquer nó.

Mais uma vez verificamos que um dos nós de entrada desta aplicação é o nó que origina mais transições. Os nós “Diário” e “Terapeutica” são os seguintes na lista e por esta razão devem, caso a lógica da aplicação permita, ter acesso directo a partir de qualquer outro nó.

Como referido na secção dos resultados o algoritmo de verificação “INVALID PATH” não pode ser aplicado porque não foi identificado um nó de saída da aplicação. Este facto pode ser relevante pois, tendo em conta o período temporal a cujo log diz respeito (mais de 6 anos), a probabilidade desse nó não estar presente na amostra é muito baixa. Sendo assim isto significa que:

- Ou não existe a noção de logout na aplicação;
- Ou a acção de logout não é gravada no log.

Seja qual for a justificação, é aconselhável rever esta questão na aplicação.

Considerações finais

A avaliação de aplicações é um passo importante na melhoria e adaptação das mesmas aos trabalhos necessários. A aplicação desenvolvida e apresentada neste documento permite atacar o passo da avaliação relacionado com o modo como os utilizadores fazem uso das aplicações.

O “Viajante” aproveita os ficheiros de log das aplicações estudadas e esta abordagem tem vantagens quanto a algumas das formas de avaliação mais conhecidas pois não insere o factor da observação que pode alterar os resultados. Para além disso não ocupa tempo aos utilizadores e também não é dispendiosa pois a gravação de ficheiros de logs é prática comum hoje em dia em aplicações informáticas. Ao transformar os ficheiros de log em grafos o “Viajante” disponibiliza uma forma poderosa de análise

dos dados quer através do desenho dos próprios grafos quer através da aplicação de algoritmos da teoria de grafos. O desenho dos grafos revelou-se particularmente interessante para aplicações com poucos nós, tornando-se ilegível para aplicações com um número de nós mais elevado. Ainda assim, os vários tipos de grafo desenhados abrem a possibilidade a vários tipos de análise manual pois esquematizam os dados de formas diferentes e complementares. Nos casos em que a aplicação estudada apresentava um número de nós elevado a capacidade de aplicar algoritmos da teoria de grafos foi fundamental para gerar recomendações para as aplicações. Os algoritmos aplicados geraram resultados muito próximos dos resultados indicados pelo perito na aplicação breast.care e por isso foram também usados na aplicação intensive.care.

Para finalizar, é opinião do autor que tanto a aplicação desenvolvida como a abordagem utilizada apresentam uma mais-valia para a execução de análises de aplicações informáticas como foi demonstrado com as aplicações estudadas. Vários estudos na área da saúde estão já a usar alguns dos conceitos formalizados pelo “Viajante” o que pode fazer com que este seja adoptado em novos casos de estudo e melhorado para ser aplicado em novas situações.

7. Trabalho futuro

A não execução dos seguintes pontos deve-se à escassez de tempo mais do que a impossibilidades tecnológicas ou de qualquer outro tipo. Os pontos identificados onde esta ferramenta desenvolvida poderá vir a ser melhorada incluem:

- Comparação de grafos: Após definidas as métricas, esta funcionalidade poderá ser útil na descoberta de grafos equivalentes, que pode levar à descoberta de clusters nos dados utilizados;
- Algoritmos disponíveis out-of-the-box: A ferramenta inclui alguns algoritmos da teoria de grafos e foi construída de forma a ser simples adicionar outros algoritmos desde que programados na mesma linguagem que a aplicação. Caso venha a ser detectada uma possível utilidade de outro algoritmo poder-se-ia acrescentar este de base na aplicação;
- Análise de logs de aplicações diferentes: Foram analisados diferentes tipos de logs no decurso deste trabalho. Existem outros programas que têm a capacidade de gravar ficheiros de actividade com informação sobre as acções neles efectuadas. A análise de logs de outras aplicações tem a possibilidade de aumentar a credibilidade da aplicação caso as recomendações automáticas se verifiquem acertadas.
- Adicionar alguma inteligência aos algoritmos. O algoritmo dos “CLIQUEs”, por exemplo, apresenta vários cliques sem nenhuma ordem específica. Uma alteração que podia fazer sentido era ordenar a apresentação dos cliques por ordem decrescente relativamente à quantidade de transições de cada clique. Isto faria com que se pudesse caminhar para a descoberta de cliques mais significativos dentro do mesmo nível.
- Para simplificar alguns grafos pode existir uma opção de cut-off para o número de transições entre nós, estabelecendo um limite mínimo de transições para que a aplicação assuma que dois nós estão ligados;

- Desenvolver/aplicar um algoritmo que distinga nós quanto ao tipo de percurso onde normalmente estão inseridos, permitindo assim perceber se o nó é acessado aleatoriamente ou no decorrer de um percurso pré-definido e sequencial.

Referências

- Abedin, B. and B. Sohrabi (2009). "Graph theory application and web page ranking for website link structure improvement." Behaviour & Information Technology **28**(1): 63 - 72.
- Administração.Central.do.Sistema.de.Saúde (2010). "Sistema de apoio à prática de enfermagem (SAPE)." Retrieved 08-12-2010, from <http://www.acs.min-saude.pt/dis/2010/10/01/sistema-de-apoio-a-pratica-de-enfermagem-sape/>.
- Administração.Central.do.Sistema.de.Saúde (2010). "Sistema de Apoio ao Médico (SAM)." Retrieved 08-12-2010, from <http://www.acs.min-saude.pt/dis/2010/10/01/sistema-de-apoio-ao-medico-sam/>.
- Administração.Central.do.Sistema.de.Saúde (2010). "Sistema Informático de gestão da Lista de Inscritos para Cirurgia (SIGLIC)." Retrieved 08-12-2010, from <http://www.acs.min-saude.pt/dis/2010/09/30/sistema-informatico-de-gestao-da-lista-de-inscritos-para-cirurgia-siglic/>.
- Administração.Central.do.Sistema.de.Saúde (2010). "Sistema Integrado de Informação Hospitalar (SONHO)." Retrieved 08-12-2010, from <http://www.acs.min-saude.pt/dis/2010/09/30/sistema-integrado-de-informacao-hospitalar-sonho/>.
- Apache (2000). "Apache HTTP Server Version 1.3: Log Files." from <http://httpd.apache.org/docs/1.3/logs.html>.
- APPEAL, A. P. P. f. E. a. L. (2001). "Evaluation Methods, Advantages/Challenges and Resources Needed." from www.appealforcommunities.org.
- Asunka, S., H. S. Chae, et al. (2009). "Understanding Academic Information Seeking Habits through Analysis of Web Server Log Files: The Case of the Teachers College Library Website." The Journal of Academic Librarianship **35**(1): 33-45.
- Bertin, J. (1983). Semiology of Graphics, University of Wisconsin Press.
- Bertot, J. C., McClure, C.R., Moen, W.E., & Rubin, J. (1997). "Web Usage Statistics: Measurement Issues and Analytical Techniques." Government Information Quarterly **14**(4): 373-395.
- Blazewicz, J., E. Pesch, et al. (2005). "Novel representation of graph structures in web mining and data analysis. ." Omega **33**(1): 65-71.
- Caldwell, C. (1995). "Graph Glossary." Retrieved 16.08.2010, from <http://www.utm.edu/departments/math/graph/glossary.html>.
- Carroll, J. M. and M. B. Rosson (1987). Paradox of the active user. Interfacing thought: cognitive aspects of human-computer interaction, MIT Press: 80-111.

Carroll, S. and T. Daughtrey (2007). Fundamental Concepts for the Software Quality Engineer, American Society for Quality.

Cruz-Correia, R., J. C. Wyatt, et al. (2010). "Determinants of frequency and longevity of hospital encounters; data use." BMC Med Inform Decis Mak **10**(1): 15-15.

CSC.Healthcare.EMEA (2010). "CSC obtains performance certification for its Clinical Suite solution." Retrieved 08-12-2010, from http://www.csc.com/health_services/press_releases/54909-csc_obtains_performance_certification_for_its_clinical_suite_solution.

Draper, S. W. (2009). "The Hawthorne, Pygmalion, Placebo and other effects of expectation: some notes." Retrieved 08-12-2010, from <http://www.psy.gla.ac.uk/~steve/hawth.html>.

Eason, K., L. Yu, et al. (2000). "The use and usefulness of functions in electronic journals: The experience of Super Journal Project." Program **34**(1): 1-28.

Ellson, J., E. Gansner, et al. (2010). "Graphviz." Retrieved 07.10.2010, 2010, from <http://www.graphviz.org/>.

EuroRec (2010). "EuroRec Home Page." Retrieved 01-12-2010, from <http://www.eurorec.org>.

EuroRec (2010). "Voting Form EuroRec Seal 2010." Retrieved 08-12-2010, from www.eurorec.org/userfiles/file/EuroRec%20Seal2010%20Vote%2004.xls.

Gerald Stermsek, M. S., Gustaf Neumann (2007). A User Profile Derivation Approach based on Log-File Analysis. Proc. of the International Conference on Information and Knowledge Engineering.

Gunter T, T. N. (2005). "The Emergence of National Electronic Health Record Architectures in the United States and Australia: Models, Costs, and Questions." Journal of Medical Internet Research.

Guzdial, M. (1993). Deriving Software Usage Patterns from Log Files. T. R. G.-G.-.-. 41.

Healthcare.Informatics.Society.of.Ireland (2009). "Healthcare Informatics Society of Ireland Newsletter - Summer 2009." Retrieved 08-12-2010, from http://www.hisi.ie/media/HISI_Newsletter_June_2009_Edition.pdf.

Hegarty, M., P. A. Carpenter, et al. (1991). Diagrams in the comprehension of scientific text. New York, Longman.

Heijmen, B., J. Nuytens, et al. (2008). "Assessment of the Clinical Accuracy of Real-time Respiratory Motion Tracking by Analysis of Log Files." International Journal of Radiation Oncology*Biophysics*Physics **72**(1, Supplement 1): S626-S627.

Hoogeman, M., J.-B. Prévost, et al. (2009). "Clinical Accuracy of the Respiratory Tumor Tracking System of the CyberKnife: Assessment by Analysis of Log Files." International Journal of Radiation Oncology*Biophysics **74**(1): 297-303.

Ke, H.-R., R. Kwakkelaar, et al. (2002). "Exploring behavior of E-journal users in science and technology: Transaction log analysis of Elsevier's ScienceDirect OnSite in Taiwan." Library & Information Science Research **24**(3): 265-291.

Kozma, R. B. (1991). "Learning with Media." Review of Educational Research **61**(2): 179-211.

Learning.Technologies.at.Virginia.Tech. (2009). "Evaluation." Retrieved 06-12-2010, from <http://www.edtech.vt.edu/edtech/id/eval/eval.html>.

M. Guzdial, C. W., M. Konemann, E. Soloway (1993). Characterizing Process Change Using Log File Data. Georgia Institute of Technology Gvu Center Technical Report No. 93.

Mariner, V. (2002). "Logging Usability." Library Journal NetConnect.

Mark Guzdial, P. S., Albert Badre, Scott Hudson, Mark Gray (1994). Analyzing and Visualizing Log Files: A Computational Science of Usability. Gvu Center TR GIT-Gvu-94-8, Georgia Institute of Technology.

McNamara, C. (2010). "Basic Guide to Program Evaluation." Retrieved 06-12-2010, from http://managementhelp.org/evaluatn/fnl_eval.htm.

Microsoft (2003). "IIS Log File Format (IIS 6.0)." Retrieved 03-09-2010, from <http://www.microsoft.com/technet/prodtechnol/WindowsServer2003/Library/IIS/be22e074-72f8-46da-bb7e-e27877c85bca.mspx?mfr=true>.

Ministério.da.Saúde (2009). Retrieved 08-12-2010, from <http://www.min-saude.pt/portal/conteudos/a+saude+em+portugal/politica+da+saude/discussao/arquivo/viabilidade+sam.htm>.

Nicholas, D., Huntington, P., Lievesley, N., & Withey, R. (1999). "Cracking the Code: Web Log Analysis." Online & CD-ROM Review **23**(5): 263-269.

Ramadhan, H. A. (2009). "A Heuristic Based Approach for Improving Website Link Structure and Navigation." Journal of Emerging Technologies in Web Intelligence **1**.

Scott, J. (2000). Social Network Analysis: A Handbook, SAGE Publications Ltd.

Silva, V. (2010). Estudo do Workflow Clínico e Fluxo de Informação da Unidade de Diagnóstico de Patologia Mamária do Hospital de São João. Porto, Faculdade de Ciências Universidade do Porto | Faculdade de Medicina Universidade do Porto: 116.

SRC.Infonet (2010). "SRC Infonet awarded EuroRec Seal Level 2 Certificate." Retrieved 08-12-2010, from <http://www.infonet.si/en/news/src-infonet-awarded-eurorec-seal-level-2-certificate>.

Taha, A. (2004). Wired research: transaction log analysis of e-journal databases to assess the research activities and trends in UAE University. Proceedings of 12th Nordic Conference on Information and Documentation. Aalborg, Denmark.

Thelwall, M. and D. Wilkinson (2003). "Graph structure in three national academic webs: Power laws with anomalies." Journal of the American Society for Information Science and Technology **54**(8): 706-712.

Trochim, W. M. K. (2006). "The Planning-Evaluation Cycle." Retrieved 06-12-2010, from <http://www.socialresearchmethods.net/kb/pecycle.php>.

Vekiri, I. (2002). "What Is the Value of Graphical Displays in Learning?" Educational Psychology Review **14**(3): 261-312.

Warren, N. (2002). "Website Log Analysis: Approaches for the Library of the National Institute Of Environmental Health Sciences." A Master's Paper for the Masters of Science in Library Science.: 75.

Wikipedia (2010). "Graph theory." Retrieved 05-12-2010, from http://en.wikipedia.org/wiki/Graph_theory.

Winn, W. (1987). The Psychology of Illustration. New York, Springer-Verlag.

Winne, P., L. Gupta, et al. (1994). "Exploring individual differences in studying strategies using graph theoretic statistics." Alberta Journal of Educational Research **40**: 177-193.

Yan, H.-M., T. Qin, et al. (2005). Calculating webpage importance with site structure constraints. Paper presented at information retrieval technology. Lecture Notes in Computer Science. **3689**: 546-551.

Zawitz, M. W. (1998). Measuring User Activity: An Analysis of Bureau of Justice Statistics (BJS) Website Usage Statistics., Bureau of Justice Statistics Publication.

Zhou, B., J. Chen, et al. (2001). Website link structure evaluation and improvement based on user visiting patterns. 12th ACM conference on hypertext and hypermedia. Denmark: 241-244.