

Sistema de Recomendação Inteligente para uma Plataforma de *E-Learning*

Dissertação apresentada para a obtenção do grau de Mestre em
Informática e de Sistemas

Autor
Pedro Filipe Caldeira Neves

Orientadores
Doutor Carlos Manuel Jorge da Silva Pereira
Doutor Viriato António Pereira Marinho Marques
Instituto Superior de Engenharia de Coimbra

Coimbra, Novembro, 2014

Aos meus pais, irmãos e namorada
que sempre acreditaram no meu empenho,
pela amizade, motivação e carinho.

AGRADECIMENTOS

Ao professor Carlos Pereira pela orientação e aconselhamento dados ao longo das etapas deste trabalho;

Ao professor Viriato Marques pela ajuda, orientação e aconselhamento que também sempre me deu e pela sua excelente revisão da minha tese e apoio dado ao longo do percurso;

Ao Eduardo Pratas que sempre se disponibilizou para esclarecer todas as dúvidas que me surgiam no funcionamento da plataforma ADAPT;

Ao ISEC por possibilitar a realização deste trabalho;

À FCT por financiar o desenvolvimento do projeto ADAPT com a bolsa de referência PTDC/CPE-CED/115175/2009 FEDER - Eixo I de Programa Operacional Fatores de Competitividade (POFC) / QREN (COMPETE: FCOMP-01-0124-FEDER-014418).

Aos meus Pais, Irmãos, Namorada e restantes familiares que sempre estiveram do meu lado nos melhores e piores momentos e que sempre tiveram paciência para discutir assuntos do meu projeto, nunca me negaram ajuda e orientação psicológica e frequentemente me recargavam com novas ideias e sobretudo motivação. Foi sem dúvida graças a eles que o desenvolvimento deste trabalho foi possível;

Aos meus amigos que tiveram sempre um papel bastante importante durante o desenvolvimento deste trabalho;

A todos os restantes que contribuíram direta ou indiretamente para o desenvolvimento deste trabalho, quer seja por espontaneamente me fornecerem as suas opiniões ou motivação.

RESUMO

As plataformas de *e-Learning* são cada vez mais utilizadas na educação à distância, facto que se encontra diretamente relacionado com a possibilidade de proporcionarem aos seus alunos a valência de poderem assistir a cursos em qualquer lugar.

Dentro do âmbito das plataformas de *e-Learning* encontra-se um grupo especialmente interessante: as plataformas adaptativas, que tendem a substituir o professor (presencial) através de interatividade, variabilidade de conteúdos, automatização e capacidade para resolução de problemas e simulação de comportamentos educacionais.

O projeto ADAPT (plataforma adaptativa de *e-Learning*) consiste na criação de uma destas plataformas, implementando tutoria inteligente, resolução de problemas com base em experiências passadas, algoritmos genéticos e *link-mining*. É na área de *link-mining* que surge o desenvolvimento desta dissertação que documenta o desenvolvimento de quatro módulos distintos:

O primeiro módulo consiste num motor de busca para sugestão de conteúdos alternativos; o segundo módulo consiste na identificação de mudanças de estilo de aprendizagem; o terceiro módulo consiste numa plataforma de análise de dados que implementa várias técnicas de *data mining* e estatística para fornecer aos professores/tutores informações importantes que não seriam visíveis sem recurso a este tipo de técnicas; por fim, o último módulo consiste num sistema de recomendações que sugere aos alunos os artigos mais adequados com base nas consultas de alunos com perfis semelhantes.

Esta tese documenta o desenvolvimento dos vários protótipos para cada um destes módulos.

Os testes efetuados para cada módulo mostram que as metodologias utilizadas são válidas e viáveis.

Palavras-Chave: plataforma ADAPT, *e-Learning*, motor de busca, *link-mining*, *data mining*, sistema de recomendações.

ABSTRACT

E-Learning platforms are broadly used in distance education. This fact is related with the ability to provide to students the possibility to attend courses anywhere in the world through the internet.

Within the scope of Learning Management Systems (LMS), there is a group of special interest: the adaptive platforms that tend to replace the tutors (in-loco) through interactivity, content variability, automation, solving problems skills and by simulating educational behavior.

The ADAPT (an adaptive *e-Learning* platform) project is one of these platforms and implements intelligent tutoring systems, case-based reasoning, genetic algorithms and *link-mining*. It is within the scope of *link-mining* that this thesis was developed, consisting of four modules:

The first module is a search engine for suggesting additional contents; the second module identifies changes in the student's learning style; the third module consists of a platform for analyzing data that implements several *data mining* and statistical techniques to provide teachers/tutors with important information that may not be visible without the use of such techniques; finally, the last module consists of a recommender system that suggests to the students the suitable contents consulted by their colleagues with the closest related profile.

This thesis documents several prototypes for the developed modules.

The experiments performed for each module show that the methodologies are valid and viable.

Keywords: ADAPT platform, *e-Learning*, search engine, *link-mining*, data mining, recommender system.

ÍNDICE

CAPÍTULO I – INTRODUÇÃO1

1	Enquadramento	1
2	<i>E-Learning</i>	2
3	Plataforma ADAPT	5
4	Estilos de aprendizagem.....	6
5	Estrutura da tese	7

CAPÍTULO II – MOTOR DE BUSCA.....9

1	Introdução	9
2	Estado de arte	10
3	Conceitos gerais sobre motores de busca	10
3.1	Indexação	11
3.2	Pesquisa	11
4	Frameworks utilizadas no âmbito deste módulo	11
4.1	Lucene	12
5	Desenvolvimento do motor de busca	12
5.1	Indexação	12
5.2	Pesquisa.....	16
5.3	Chamada do módulo.....	20
5.4	Interação do módulo com o aluno	20
6	Performance	21
6.1	Precision, <i>recall</i> e <i>accuracy</i>	21
6.2	Ordenação de resultados.....	22
7	Conclusões	24

CAPÍTULO III – CLASSIFICADOR DE ESTILOS DE APRENDIZAGEM25

1	Introdução	25
2	Estado de arte	25
3	Desenvolvimento do classificador de estilos de aprendizagem.....	26
3.1	Campos da base de dados a ter em conta.....	26
3.2	Algoritmo implementado.....	27
4	Interação com a plataforma	29
5	Performance	29
6	Conclusões	30

CAPÍTULO IV – PLATAFORMA DE ANÁLISE DE DADOS31

1	Introdução	31
2	Estado de arte	31
3	Frameworks utilizadas no âmbito deste módulo	32
4	Desenvolvimento da plataforma.....	32
4.1	Objetivo do Módulo	32
4.2	Identificação de grupos de utilizadores	33
4.3	Média das notas por estilo de aprendizagem	35
4.4	Extração de regras	38

4.5	Percentagem de visualização de OAs por estilo de aprendizagem	41
4.6	Percentagem de mudanças de estilo de aprendizagem.....	42
4.7	Média das notas e predominância de estilos por disciplina e curso	43
4.8	Distribuição dos alunos por estilo de aprendizagem.....	45
4.9	Notas (discretizadas) por OA, por disciplina.....	47
4.10	Predição de notas	49
4.11	Identificação de grupos de utilizadores.....	52
4.12	Métodos adicionais	55
4.13	Organização de ficheiros	55
4.14	Interação com o projeto ADAPT	55
5	Conclusões	55
CAPÍTULO V – SISTEMA DE RECOMENDAÇÕES		57
1	Introdução	57
2	Estado de arte	57
3	Desenvolvimento do sistema de recomendações.....	58
3.1	<i>Dataset</i> criado.....	58
3.2	Algoritmo	60
4	Interação com o utilizador e plataforma	61
5	Performance	62
6	Conclusões	63
CAPÍTULO VI – PROTÓTIPOS DESENVOLVIDOS		65
1	Introdução	65
2	Motor de busca e sistema de recomendações	65
2.1	Construção de índices	65
2.2	Visualização de resultados.....	67
3	Plataforma de análise da base de dados do projeto.....	69
3.1	Criação de dados.....	69
3.2	Visualização de dados	70
CAPÍTULO VII – CONCLUSÕES E TRABALHO FUTURO		79
REFERÊNCIAS BIBLIOGRÁFICAS		81
ANEXOS.....		87

ÍNDICE DE FIGURAS

Figura 2.1 - Modo de funcionamento (geral) de um motor de busca.....	11
Figura 2.2 - Campos a indexar	13
Figura 2.3 - Campos a indexar para a criação dos índices de cada OA.....	13
Figura 2.4 - Ficheiro que relaciona cada conteúdo com a sua pontuação, tempo de visualização e estilo de aprendizagem. (idOA : PontuaçãoAssociada : TempoConsulta : EstiloAprendizagemPredominante).....	14
Figura 2.5 - Imagem contextual do módulo do motor de busca em funcionamento.	24
Figura 3.1 - Tabelas da base de dados a ter em conta e campos a endereçar.....	27
Figura 3.2 - Resultados obtidos pela <i>query</i> inicial.....	28
Figura 4.1 - Tabelas utilizadas para retirar os dados necessários para a pergunta 1.	33
Figura 4.2- <i>Dataset</i> da <i>query</i>	34
Figura 4.3 - Resposta à pergunta 1.	35
Figura 4.4 - Tabelas utilizadas para retirar os dados necessários à pergunta 2.	36
Figura 4.5 - <i>Dataset</i> correspondente à pergunta 2.....	36
Figura 4.6 - Resultados da pergunta 2 para a disciplina de sistemas digitais.	37
Figura 4.7 - Tabelas utilizadas para retirar os dados necessários para a pergunta 3.	39
Figura 4.8 - <i>Dataset</i> da pergunta 3.	39
Figura 4.9 – Resposta à pergunta 3.	40
Figura 4.10 - Tabelas utilizadas para retirar os dados necessários para a pergunta 4.	41
Figura 4.11 - <i>Dataset</i> da pergunta 4.....	41
Figura 4.12 - Resultado da pergunta 4. Disciplina de sistemas digitais.....	42
Figura 4.13 - Resultado do módulo de classificador de estilos de aprendizagem.	42
Figura 4.14 - Resposta à pergunta 5.	43
Figura 4.15 - Tabelas utilizadas para retirar os dados necessários para a pergunta 6.	44
Figura 4.16 - <i>Dataset</i> da pergunta 6.....	44
Figura 4.17 - Resposta à pergunta 6.....	45
Figura 4.18 - Tabelas utilizadas para retirar os dados necessários para a pergunta 7.	46
Figura 4.19 - <i>Dataset</i> da pergunta 7.....	46
Figura 4.20 - Resposta à pergunta 7.....	47
Figura 4.21 - Tabelas utilizadas para retirar os dados necessários para a pergunta 8.	48
Figura 4.22 - <i>Dataset</i> da pergunta 8.....	48
Figura 4.23 - Resultados da pergunta 8: disciplina de sistemas digitais.	49
Figura 4.24 - Tabelas utilizadas para retirar os dados necessários para a pergunta 9.....	50
Figura 4.25 - <i>Principal Component Analysis</i>	51
Figura 4.26 - Resposta à pergunta 9.....	52
Figura 4.27 - Tabelas utilizadas para retirar os dados necessários para a pergunta 10.....	53
Figura 4.28 - <i>Dataset</i> da pergunta 10.....	53
Figura 4.29 - <i>Clusters</i> formados na pergunta 10.	54
Figura 5.1 - Tabelas utilizadas para a geração do <i>dataset</i> do módulo de recomendações.....	59
Figura 5.2 - <i>Dataset</i> do módulo de recomendação.	60
Figura 5.3 - Regras geradas pelo algoritmo <i>Apriori</i>	61
Figura 5.4 - <i>Screenshot</i> do módulo de recomendação em operação.	63
Figura 6.1 - Interface de parametrização da indexação.	66
Figura 6.2 - Interface de parametrização da indexação: Aplicação a correr.	66
Figura 6.3 - Visualização de resultados do protótipo do motor de busca e sistema de recomendações.....	67

Figura 6.4 - Resultados do módulo do motor de busca.	68
Figura 6.5 - Resultados do sistema de recomendações.....	68
Figura 6.6 - Aspeto do protótipo depois de receber os resultados dos módulos de pesquisa e recomendação.	69
Figura 6.7 - Interface de parametrização da geração de dados.....	70
Figura 6.8 - Plataforma de análise de dados - Separação entre dados estatísticos e <i>data mining</i>	71
Figura 6.9 - Definição do período temporal para calcular a percentagem de mudanças de estilo de aprendizagem ao longo de um período temporal definido pelo utilizador.	72
Figura 6.10 - Percentagem de mudanças de estilo de aprendizagem ao longo de um período temporal definido pelo utilizador - Gráfico gerado.....	72
Figura 6.11 - Plataforma de análise de dados – Subcategorias da categoria <i>data mining</i>	73
Figura 6.12 - <i>Clustering – K-means</i>	74
Figura 6.13 - <i>Clustering</i> por períodos temporais – <i>X-means</i>	75
Figura 6.14 - Predição da média dos alunos não avaliados.	75
Figura 6.15 - Extração de regras de associação.	76
Figura 6.16 - Definições da plataforma de visualização de dados.....	77
Figura 6.17 - Pesquisa por id de utilizador.	77
Figura 6.18 - Resultado da pesquisa pelo ID do utilizador.....	77

ÍNDICE DE TABELAS

Tabela 2.1 - <i>N-Grams</i>	16
Tabela 2.2 - <i>Queries</i> geradas	17
Tabela 2.3 - Testes de performance – <i>Accuracy</i> , precisão e <i>recall</i>	22
Tabela 2.4 - Testes de performance – Ordenação.....	23
Tabela 3.1 - Teste de performance do módulo 2.....	30
Tabela 4.1 - <i>Dataset</i> explicativo do algoritmo <i>Apriori</i>	38
Tabela 4.2 - <i>Dataset</i> da pergunta 9.	50
Tabela 5.1 - Distribuição dos alunos no <i>dataset</i>	62

SIMBOLOGIA

Coord(q,d) – Fator de coordenação;

QueryNorm(q) – Fator normalizante;

Tf(t em d) – Frequência dos termos.;

Id(t) – Número de documentos onde o termo aparece;

T.GetBoost() – Importância do termo na *query*;

Norm(t,d) – Encapsulação de fatores normalizantes;

PontuaçãoFinal – Pontuação de um resultado no motor de busca;

OAPontuação – Pontuação média associada a um conteúdo;

OATempoNavegacao – Tempo de navegação médio associado a cada conteúdo;

EstiloAprendizagem – Factor de penalização que penaliza um conteúdo se este não pertencer ao mesmo estilo de aprendizagem que o aluno;

Q1 – Coeficiente do tempo de navegação médio de um conteúdo;

Q2 – Coeficiente da pontuação média de um conteúdo;

M_i – Valor da pontuação para a preferência VARK “i”;

l – Representa o número de artigos que o aluno consultou da preferência VARK “i”;

j – Representa o iterador de artigos consultados pelo aluno para a preferência VARK “i”;

i – Preferência VARK. Pode tomar o valor de visual, *aural*, *readwrite* e *kinesthetic*;

PontuacaoArtigo – Pontuação média associada a um conteúdo;

TempoVisualizacao – Tempo de navegação médio associado a cada conteúdo;

G-mean – Média geométrica;

ABREVIATURAS

ADAPT – Adaptive LMS. Projeto em desenvolvimento;
OA – Objeto de aprendizagem;
EA – Estilo de Aprendizagem;
LMS – Sistema de gestão de aprendizagem;
CAI – Computer-Aided Instruction;
ITS – Intelligent tutoring system;
ARFF – Attribute-Relation File Format;
CSV – Comma separated values;
VP – Verdadeiros positivos;
VN – Verdadeiros Negativos;
FP – Falsos positivos;
FN – Falsos negativos;
SCORM – Sharable Content Object Reference Model;
TEL – Technology Enhanced Learning;
CBI – Computer Based Instruction;
CAI – Computer Aided Instruction;
CMI – Computer Managed Instruction;
CBT – Computer Based Training;
IBT – Internet Based Instruction;
WBT – Web Based Training;
LM – Link mining;
DB – Base de dados na sigla inglesa;

CAPÍTULO I – INTRODUÇÃO

Este capítulo aborda a introdução desta dissertação especificando o seu enquadramento e objetivos, o estado de arte de *e-Learning*, a arquitetura do projeto ADAPT e por fim um breve levantamento de estado de arte de modelos de estilo de aprendizagem e a estrutura geral da tese.

1 Enquadramento

Este trabalho, de carácter relativamente técnico e orientado ao público académico, insere-se no âmbito do projeto ADAPT (ADAPT, 2014), uma plataforma de *e-Learning* (e-Learning, 2014) inteligente de aprendizagem assistida e automática. A plataforma pretende ajudar os alunos, inicialmente do Departamento de Engenharia Informática e Sistemas (DEIS) do Instituto Superior de Engenharia de Coimbra (ISEC) (ISEC, 2014), ao longo do seu percurso académico, através da disponibilização de conteúdos programáticos ajustados às necessidades de cada aluno, adequando-se ao seu método de aprendizagem.

Esta dissertação documenta o desenvolvimento de quatro módulos no âmbito desta plataforma que tem vindo a ser desenvolvida pelo Instituto Superior de Engenharia do Porto (ISEP) (ISECP, 2014), Instituto Superior de Engenharia de Coimbra (ISEC), Universidade de Trás-os-Montes e Alto Douro (UTAD) (UTAD, 2014), Grupo de Investigação em Engenharia do Conhecimento e Apoio à decisão (GECAD) (GECAD, 2011) e por fim, suportado pela Fundação para a Ciência e a Tecnologia (FCT) (FCT, 2014) e FEDER (FEDER, 2011). A sua motivação prende-se com a necessidade de desenvolvimento de alguns módulos que permitam a sugestão de conteúdos adicionais aos alunos da plataforma. Nos módulos previstos encontram-se também dois que analisam *links* e que capacitam a plataforma para adaptação a mudanças nos estilos de aprendizagem dos alunos, fornecendo também aos professores/tutores elementos destinados a facilitar a compreensão dos dados relativos a utilização e automaticamente produzidos pela plataforma quando em uso.

Dos quatro módulos desenvolvidos o primeiro prende-se com o desenvolvimento de um motor de busca que permite sugerir aos alunos melhores conteúdos com base em parâmetros sociais, como a média das notas dos alunos que consultaram os artigos e o seu tempo de visualização médio; o segundo módulo prende-se com o desenvolvimento de um sistema de classificação que avalie se os alunos têm vindo a mudar o seu estilo de aprendizagem e, no caso de a resposta ser positiva, sugira o ajuste dos guiões que o aluno segue; o terceiro módulo consiste no desenvolvimento de uma plataforma de extração de conhecimento da base de dados de modo a que os professores/tutores possam entender o progresso dos seus alunos, o estilo de aprendizagem associado a melhores notas, entre outras funcionalidades; por fim, o quarto módulo consiste no desenvolvimento de um sistema de recomendações que sugira aos alunos os melhores conteúdos consultados pelos seus colegas com perfis académicos semelhantes.

Os quatro módulos estão dissociados e por isso, também nesta tese são descritos em capítulos diferentes.

2 *E-Learning*

O conceito de *e-Learning* aplica conteúdos eletrônicos e tecnologias da informação e comunicação à educação. Este conceito surgiu nos anos 60 na universidade de Illinois onde na altura foi criada uma rede de terminais que permitia aos alunos ter acesso a informação extra (Wooley, 2013) e.g.: aulas previamente gravadas. O trabalho nesta área foi sendo desenvolvido e em 1976, Bernard Luskin utilizando como meio a TV, lançou “Coastline Community College” como o “colégio sem paredes”.

Em meados de 1980 já era possível aceder a este tipo de cursos em muitas bibliotecas universitárias. Com a chegada dos anos 90 vieram melhoramentos tecnológicos que levaram ao desenvolvimento de plataformas web que forneciam cursos online (bates, 2005).

Na década de 2000 surgiram modelos standard de padrões e especificações para *e-Learning*, dos quais se destaca a norma SCORM (“Sharable Content Object Reference Model” na sigla inglesa) que foi criada ainda em 1999 (aprofundada nos anos seguintes) e define padrões e especificações para *e-Learning* baseados na web (SCORM, 2014).

O desenvolvimento da internet nesta década veio cimentar a aceitação deste tipo de plataformas pelos professores e tutores que viram nelas um grande potencial pedagógico. De 2002 a 2005 o número de estudantes utilizadores de *e-Learning* subiu cerca de 65% e em 2013, segundo Christopher Pappas afirma no seu *survey* (Estatísticas *e-Learning*, 2014), o *e-Learning* é um negócio de 56.2 mil milhões de dólares que se pensa duplicar até ao ano de 2015. Segundo Christopher, só na Europa e nos Estados Unidos concentram-se cerca de 70% dos utilizadores de *e-Learning* embora a Ásia esteja a ganhar terreno. O autor assegura ainda que cerca de 72% das empresas que responderam ao *survey*, afirmam que o *e-Learning* as mantém no topo empresarial, na medida em que esta área está em constante evolução e que através de *e-Learning* mais facilmente conseguem acompanhar essas mudanças, fornecendo formação aos seus trabalhadores.

Atualmente, *e-Learning* é um conceito com várias designações equivalentes: multimedia Learning, Technology-enhanced learning (TEL), Computer-Based Instruction (CBI), Computer managed instruction (CMI), Computer-Based training (CBT), Computer-Aided instruction (CAI) ou Computer-Based instruction, Internet-Based Training (IBT), Web-Based training (WBT), Online education, Virtual education, Virtual learning environments (VLE) e M-Learning. Contudo subdivide-se em dois tipos distintos (Hrastinski et al, 2008):

- Sistemas síncronos: tipo de plataformas que permitem a partilha de conteúdos em tempo real em que os alunos interagem em simultâneo. Este tipo de plataformas exige um ou mais participantes durante um determinado período de tempo, como por exemplo uma discussão face-a-face em que são necessários dois ou mais intervenientes.

Alguns exemplos deste tipo de sistemas são: BlackboardCollaborate (Blackboard Collaborate, 2014), Adobe Connect (Adobe Connect, 2014) e WebEx (Cisco WebEx, 2014). Estas ferramentas são normalmente utilizadas em conversações.

- Sistemas assíncronos: um tipo de plataformas que permite a partilha de conhecimento e a interação de alunos sem que haja necessidade de outros estarem presentes.

Alguns exemplos deste tipo de sistemas são: Desire2Learn (criado em 1999 por John Baker) (Desire2Learn, 2013), sistemas Blackboard, Moodle (um dos sistemas mais utilizados - criado em 2002 por Martin Dougiamas) (Moodle, 2014) e Sakai (criado em 2005 por uma comunidade que envolve instituições acadêmicas e organizações comerciais) (Sakai, 2014).

A discussão sobre os benefícios e limitações dos dois tipos de sistemas de *e-Learning* limitou-se a um curto período temporal no qual os investigadores tentaram determinar o tipo de plataforma que melhor funcionava. Porém, os estudos criados nesse âmbito não registaram diferenças significativas (Nirmalani, 2014). Assim, em vez de tentar determinar o melhor tipo de plataformas, a comunidade de *e-Learning* começou a entender que tanto um como outro têm as suas vantagens e desvantagens e que tudo depende da função requerida. Algumas das vantagens dos sistemas assíncronos que são desvantagens dos sistemas síncronos são:

- Os alunos podem descarregar os conteúdos quando desejarem e estudar quando pretenderem;
- Os alunos podem conciliar o estudo com as suas outras atividades;
- Os alunos não têm limite temporal (salvo algumas exceções de avaliações) sobre o tempo de consulta de conteúdos.

Algumas das vantagens dos sistemas síncronos que acabam por ser desvantagens dos sistemas assíncronos são:

- Os alunos sentem-se integrados numa turma como se estivessem a ter aulas presenciais;
- Os alunos têm horários restritos que lhes dão a sensação de estarem em aulas presenciais;
- Os alunos recebem apoio direto dos professores;
- Estes sistemas suportam normalmente muito mais tipos de conteúdo que os assíncronos.

Os dois sistemas são bastante úteis dependendo da função a desempenhar, embora, pelas suas características flexíveis, normalmente os sistemas assíncronos sejam os preferidos: a maioria das pessoas utiliza sistemas de *e-Learning* precisamente por lhes proporcionar mais flexibilidade, permitindo que continuem a efetuar as suas tarefas quotidianas sem as restrições horárias que um curso com aulas presenciais, ou um sistema síncrono envolvem.

Os sistemas referidos anteriormente têm uma característica em comum: são sistemas não inteligentes. A área de sistemas de *e-Learning* inteligentes é uma área muito vasta que surgiu por volta da década de 70 com a ascensão dos sistemas de *e-Learning*. Jaime Carbonell sugeriu que os computadores poderiam ser “professores” em vez de apenas mais uma ferramenta. Surgiu assim uma nova área nos sistemas de *e-Learning* que se focava no ensino inteligente de alunos. A esta nova tecnologia denominou-se ICAI (Intelligent Computer Assisted Instruction na sigla inglesa) ou ITS (Intelligent Tutoring Systems na sigla inglesa). Contudo, nesta década os requisitos técnicos da implementação de um ITS eram maiores e mais complexos do que os

de um sistema CAI e o desenvolvimento de ITSs ficou quase estagnado (Chambers e Spreacher, 1983).

Com a criação de computadores como o Apple 2, Commodore PET e TRS-80 o desenvolvimento de ITS e CAI ganhou de novo interesse. Com os novos desenvolvimentos, os sistemas CAI e ITS divergiram um pouco: os sistemas CAI focaram-se em interações mais profundas com os conteúdos criados para uma área específica, enquanto a área de desenvolvimento de sistemas ITS procurou incidir sobre o conhecimento e a capacidade de generalizar esse conhecimento em formas não-específicas (Larkin e Chabay, 1992). O conceito de ITS tal como se pensava na altura, estabelecia como metas não só a capacidade de ensinar uma tarefa a um aluno como também de a realizar, adaptando-se dinamicamente a cada situação. Na transição dos sistemas CAI para sistemas ICAI, um computador deveria distinguir não só entre a resposta correta e incorreta, mas também o tipo de resposta incorreta de modo a ajustar o tipo de instrução a dar ao aluno. As primeiras implementações de sistemas ITS consistiam em diagnosticar este tipo de erros e adaptar-se de modo a remediar os erros ocorridos com base no diagnóstico efetuado – tipo de implementação que ainda hoje em dia é utilizada (Shute e Psotka, 1994).

Um marco histórico no desenvolvimento de sistemas de ITS foi a criação do sistema LISPITS, criado em 1983 para ensinar aos alunos a linguagem de programação LISP (Corbett e Anderson, 1992) que ajudava na identificação de erros e fornecia feedback construtivo aos seus utilizadores.

Outros sistemas como o TUTOR descrito em (Ford, 2008) e PARNASSUS (Bailin e Levin, 1989) foram ainda desenvolvidos durante a década de 80 para ensino geral e de linguagens de programação respetivamente.

Durante as décadas de 1990 e 2000 surgiram mais alguns ITSs baseados num sistema de ITTs (Intelligent Tutoring Tools na sigla inglesa).

Atualmente o desenvolvimento de sistemas CAI e ICAI continua a ser uma área bastante ativa.

O artigo (Kawakami et al, 2010) apresenta o desenvolvimento de um sistema de *e-Learning* denominado ELSEC que tem a função de ensinar segurança aos seus alunos. i.e.: o sistema, através de imagens e animações, ensina alguns parâmetros de segurança que poderão evitar alguns tipos de ataques eletrónicos.

Em (Venkatesan et al, 2013) os autores apresentam um sistema de *e-Learning* adaptável baseado numa ontologia híbrida relacionada com o ensino de crianças autistas. Criado a partir do conhecimento sobre a doença e metodologias de ensino associadas, o sistema adapta-se às necessidades de aprendizagem de cada criança, facilitando-lhe conteúdos de aprendizagem mais adequados às suas capacidades.

Por fim, o artigo (Wei et al, 2009) descreve um sistema CAI que combina técnicas de *data mining* para fornecer aos alunos cursos didáticos personalizados. O sistema recomenda ainda conteúdos úteis ao aluno tendo como base o seu desempenho.

Os três trabalhos referidos são apenas alguns das dezenas de trabalhos desenvolvidos nas últimas décadas. A área de *e-Learning* e também b-learning (blended-learning – um tipo de ensino onde a maioria dos conteúdos é transmitida online mas que inclui também situações presenciais) é uma área em crescente mudança que, como se afirmou, é também um negócio bastante rentável. O facto de cada vez mais as pessoas terem menos tempo cria a necessidade de utilização deste tipo de sistemas. Os sistemas de *e-Learning* têm as suas vantagens e desvantagens mas permitem a internacionalização de muitas universidades e empresas fornecedores de ensino.

3 Plataforma ADAPT

Os sistemas de *e-Learning* atuais apresentam grandes limitações especialmente no que toca à pedagogia e interatividade. Estes sistemas são normalmente baseados em conteúdos estáticos sem grande interatividade (e.g.: enunciados de laboratórios práticos, fichas de exercícios, artigos, etc...) onde o professor continua a desempenhar um papel fundamental na avaliação do desempenho e capacidades cognitivas de cada aluno, construindo conteúdos, dando aulas e respondendo a dúvidas. Os sistemas atuais pouco mais permitem que a criação de eventos e conteúdos estáticos, oferecendo eventualmente a possibilidade de comunicação com os colegas/professor via *chat* ou mensagem (como é o caso do moodle).

A incorporação de interatividade e automatização numa plataforma de *e-Learning* são muito importantes. É necessário considerar as seguintes funcionalidades:

- Considerar o desempenho, estilo de aprendizagem, casos de sucesso e falha de cada aluno;
- Adaptar os conteúdos mostrados a cada aluno de acordo com o seu perfil cognitivo e preferências de aprendizagem;
- Fornecer interatividade entre o aluno e a plataforma;
- Ser capaz de aprender como “ensinar o aluno” baseando-se em experiências passadas.

As limitações dos sistemas de *e-Learning* convencionais, constituem na motivação do desenvolvimento do projeto ADAPT - *Adaptive Learning Management System* (Plataforma Adaptativa de Ensino à Distância) (Marques et al, 2010).

Para endereçar as 4 funcionalidades acima descritas, a plataforma implementa:

- *Intelligent Tutoring System* (ITS): tenta substituir os professores criando sequências de ações simulando uma atitude educacional;
- *Case-Based Reasoning* (CBR): um paradigma que se baseia em ocorrências passadas para resolver questões atuais;
- *Link-Mining* (LM): este termo representa uma área de investigação que envolve várias subsecções como por exemplo a análise de *links*, hipertexto, aprendizagem relacional e lógica indutiva. Ainda dentro da área, encontra-se também o desenvolvimento de motores de busca, sistemas de sugestões e análise de dados.

No âmbito do ADAPT, esta área é explorada nos módulos desenvolvidos nesta dissertação, sugerindo conteúdos alternativos ao aluno (respetivamente no motor de busca e sistema de sugestões) e explorando as ligações e dados produzidos (classificador de estilos de aprendizagem e plataforma de análise de dados);

- *Evolutionary Computation* (EC): envolve a otimização de problemas combinacionais. Uma subárea de EC utilizada no âmbito do projeto ADAPT são os algoritmos genéticos (AG) que se baseiam na teoria evolutiva proposta por Charles Robert Darwin (1809–1882) em que apenas os espécimens mais aptos sobrevivem (“The Origin of Species”, publicado em 1859). No caso do ADAPT, estes algoritmos são utilizados para responder a problemas “What-if” colocados pelos alunos de uma forma interativa.

Aplicando estas funcionalidades o ADAPT implementa interatividade e automatização superando os sistemas convencionais.

4 Estilos de aprendizagem

Uma vez que nem todos os alunos partilham o mesmo estilo de aprendizagem, a definição de estilos padrão em que cada aluno possa ser integrado, torna-se uma necessidade.

O modelo proposto por Kolb é um dos mais comuns e baseia-se num ciclo de aprendizagem de quatro níveis (Kolb et al, 2005), (Stash, 2005):

- Experiência concreta – (CE na sigla inglesa);
- Observação reflexiva – (RO na sigla inglesa);
- Conceitualização abstrata – (AC na sigla inglesa);
- Experiência ativa – (AE na sigla inglesa)

O modelo Honey and Mumford (Modelo de Honey e Mumford, 2014) é baseado no modelo de Kolb e utiliza um questionário (Learning Style Questionnaire) para mapear as preferências dos alunos em 4 estilos de aprendizagem:

- Ativista – Aluno que prefer aprender recorrendo a actividades experimentais;
- Teórico – Aluno que gosta de entender a teoria por detrás das acções, recorrendo frequentemente a modelos e conceitos para se esclarecer;
- Pragmático – Aluno que têm de entender como os conhecimentos que aprendem se aplicam na realidade;
- Refletor – Aluno que aprende por observação;

Anthony F. Gregorc criou uma ferramenta conhecida como delineador de Gregorc (Modelo de Gregorc, 2014), um questionário que conjuga as percepções dos alunos com as suas capacidades de ordenação determinando assim o seu estilo de aprendizagem. O modelo de Gregorc considera as percepções concreta e abstrata e as capacidades sequencial e aleatória considerando 4 estilos de aprendizagem diferentes:

- Abstrato sequencial – Aluno que tem uma forte componente de leitura e utiliza a lógica para compreender as situações com que se depara;

- Concreto sequencial – Aluno que prefere cenários experienciais e por isso prefere conteúdos relacionados com actividades práticas.
- Abstrato aleatório – Aluno que prefere receber instrução de um professor/tutor;
- Concreto aleatório – Entende rapidamente a essência das ideias e tem a capacidade de tomar saltos intuitivos. Os alunos associados a este estilo preferem, segundo o modelo a abordagem de tentativa e erro.

O modelo Felder-Silverman (Felder, 1988), (Gomes, 2000) é outro modelo de estilos de aprendizagem considerado particularmente útil pelo facto de o seu criador ter experiência na área de engenharia e ter focado o modelo mais nesta área. Este modelo considera 4 dimensões para os estilos de aprendizagem:

- *Sensing/Intuitive*; ;
- *Visual/Verbal*;
- *Active/Reflective*;
- *Sequential/Global*;

As 4 dimensões do modelo Felder-Silverman podem ser determinadas através do questionário ILS (Index of Learning Styles) – composto por 44 perguntas de duas escolhas – que mapeia as respostas do aluno nas dimensões do modelo.

O modelo VARK (Vark, 2014) parte de um questionário inicial efetuado aos alunos, e de acordo com as suas respostas atribui-lhes um dado perfil de aprendizagem caracterizado em quatro dimensões:

- *VISUAL* – preferências visuais;
- *AURAL* – preferências auditivas;
- *READWRITE* – preferências de leitura e escrita;
- *KINESTHETIC* – preferência por conteúdos com movimento;

No ADAPT o estilo de aprendizagem inicial de cada aluno é determinado pelo ILS e através de uma inferência difusa as 4 dimensões do modelo Felder-Silverman são mapeadas em preferências dos alunos por conteúdos dos tipos “*Visual*”, “*Aural*”, “*Readwrite*” ou “*Kinesthetic*”, isto é, segundo o modelo VARK. Cada OA e cada utilizador têm uma determinada percentagem de cada estilo de aprendizagem que lhes confere uma predominância de um ou alguns dos estilos. Um OA que tenha uma percentagem mais elevada no parâmetro *visual* que nos restantes pode ser considerado um OA predominantemente do tipo visual.

5 Estrutura da tese

A presente tese está dividida em sete capítulos. O Capítulo I introduz esta dissertação; os Capítulos II, III, IV e V apresentam os 4 módulos desenvolvidos, respetivamente o módulo do motor de busca, o classificador de estilos de aprendizagem, a plataforma de análise de dados e o sistema de recomendações; o Capítulo VI descreve os protótipos desenvolvidos no âmbito desta dissertação; por fim, o Capítulo VII apresenta as conclusões, trabalho futuro e considerações a referir.

CAPÍTULO II – MOTOR DE BUSCA

Este capítulo descreve a implementação do motor de busca desenvolvido estando organizado do seguinte modo: Introdução; levantamento do estado de arte; conceitos gerais sobre motores de busca; frameworks utilizadas; desenvolvimento do módulo; performance e conclusões retiradas.

1 Introdução

Este capítulo descreve o desenvolvimento de um motor de busca criado para pesquisar na base de dados da plataforma conteúdos complementares que possam ser de interesse para o aluno, habilitando a plataforma a apresentar-lhe não só os conteúdos que lhe estavam previstos nos guiões mas também conteúdos complementares pesquisados localmente.

No seu desenvolvimento usou-se a *framework Lucene* (Lucene, 2012a), (Lucene, 2012b) para os métodos de indexação e pesquisa. Contudo foram endereçados alguns parâmetros adicionais com o objetivo de criar uma nova fórmula de ordenação de resultados que tem em conta os seguintes parâmetros:

1. Similaridade de cosseno – proximidade entre os termos da *query* de pesquisa e os conteúdos indexados;
2. Pontuação de cada conteúdo – a média das notas dos alunos que consultaram o conteúdo por um tempo superior a 2 minutos;
3. Tempo de navegação de um conteúdo – a média do tempo de navegação de cada conteúdo por um período de tempo superior a 2 minutos;
4. Estilo de aprendizagem predominante do conteúdo;

O primeiro critério de ordenação, a similaridade de cosseno, fornecida pelo Lucene, mede a proximidade de cada conteúdo encontrado com aquele que o utilizador se encontra de momento a visualizar. No entanto, se existir um empate entre os valores da similaridade de cosseno dos conteúdos é utilizada uma fórmula que considera os três restantes parâmetros. Normalmente um empate entre os valores de similaridade de cosseno raramente acontece, mas acreditamos que um motor de busca para um LMS não deve considerar apenas esta heurística. Assim, este valor é limitado até quatro casas decimais, de modo a criar mais homogeneidade e ter a certeza que os restantes parâmetros também são utilizados. Este fator é bastante importante, pois significa que os alunos têm acesso aos melhores OAs (objetos de aprendizagem): aqueles que se assemelham mais ao que estão atualmente a consultar e aqueles associados a um maior tempo de navegação, melhores médias e maior proximidade ao seu estilo de aprendizagem.

Como o projeto ADAPT ainda não está totalmente implementado foi utilizado um *dataset* de experiência reduzido. No entanto, os resultados preliminares mostram que esta metodologia é viável. O motor de busca tem uma *accuracy* de 27%, precisão de 73,14% e *recall*

de 40,50% na procura dos OAs mais adequados a cada aluno. A nível da ordenação de resultados apresenta uma distância de *Hamming* ao modelo criado de 108, a sua média é de 3,6 e o desvio padrão é de 4,30.

2 Estado de arte

Existe de facto um extenso trabalho na área de motores de busca, sua personalização e alteração da função de pontuação que classifica documentos ou páginas web.

O artigo (Rosas et al, 2011) documenta um estudo que produziu um motor de busca especializado baseado no *page rank* da Google mas que, segundo os autores, melhora o método de ensino na sala de aula. O motor de busca implementa sugestões de termos e pesquisas bilinguais e os autores afirmam ainda que o seu sucesso na procura de objetos de aprendizagem é melhor que usando o Google.

Em (Coffman et al, 2011), os autores analisam fatores de algumas funções de pontuação. Os autores consideram apenas alguns desses fatores e utilizam machine learning de modo a criar uma nova função de pontuação que fornece resultados significativamente melhores que as abordagens já existentes. A solução passa por simplificar a função de pontuação, removendo fatores menos importantes. Os autores concluem que a sua abordagem supera as já existentes.

O estudo (Ozmutlu et al, 2004) apresenta um algoritmo de identificação de tópicos em sessões de pesquisa. De acordo com os autores, os utilizadores apenas efetuam algumas consultas e pesquisam temas que, muitas vezes, estão relacionados. O algoritmo que apresentam utiliza redes neuronais a fim de identificar as mudanças de tópicos nas pesquisas efetuadas.

Nos testes efetuados sobre os dados das sessões de pesquisa, os autores identificam 76% de alterações de tópicos e 92% de continuações.

No artigo (Preethi et al, 2013), os autores apresentam uma nova abordagem para o algoritmo Page Rank que passa por armazenar os resultados de pesquisas anteriores numa base de dados. O algoritmo utilizará então esses dados para classificar os resultados.

Em (Harb et al, 2009) os autores introduzem o conceito de um motor de busca personalizado através do armazenamento, estudo e utilização dos interesses dos utilizadores ao efetuar uma pesquisa. O motor de busca assenta em três fatores principais: o grau de importância da categoria do documento para o utilizador, o interesse do utilizador (que depende das suas pesquisas), e o grau de relevância do documento.

Por fim, o livro (McCandless et al, 2010) foi também uma referência importante para o desenvolvimento deste módulo.

3 Conceitos gerais sobre motores de busca

Cada motor de busca implementa os seus próprios algoritmos de indexação e pesquisa, contudo o seu modo de funcionamento é bastante semelhante. Regra geral, um motor de busca tem a configuração apresentada na figura 2.1.

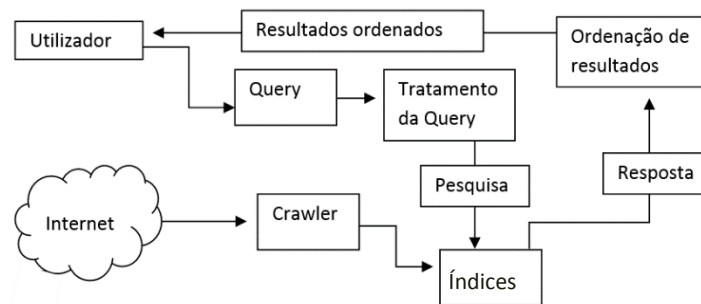


Figura 2.1 - Modo de funcionamento (geral) de um motor de busca

De acordo com o diagrama representado, o utilizador começa por fazer uma pesquisa (*Query*) que é tratada de modo a eliminar caracteres especiais e palavras sem significado (stop words, e.g.: se, então, como...). O conteúdo da *query* é então pesquisado nos índices (Índices) que foram previamente criados pelo *crawler* que indexou a informação de várias páginas *web*. Após a pesquisa, os resultados são ordenados e mostrados ao utilizador.

3.1 Indexação

A indexação corresponde ao processo de catalogação da informação e é necessária para que a pesquisa seja mais simples de efetuar. Assim, quando é feita uma pesquisa, a consulta incidirá sobre os índices criados de modo a que se consigam obter resultados de pesquisa mais rápidos.

O índice de um motor de busca funciona de modo semelhante ao índice de um livro, onde habitualmente se organiza a informação por capítulos. Neste campo da computação, a informação também é organizada por campos que definem como se deseja indexar a informação disponível para pesquisa.

3.2 Pesquisa

A pesquisa consiste na procura de informação sobre os índices criados. A pesquisa é efetuada da mesma forma que seria se ao invés de se pesquisar nos índices de páginas *web*, ou conteúdos de uma base de dados, se pesquisasse por determinado capítulo num livro. Da mesma forma que um capítulo tem normalmente um título sugestivo que define o seu conteúdo, uma entrada de um índice criado por um motor de busca contém informação sobre os dados que referencia.

4 Frameworks utilizadas no âmbito deste módulo

Existem várias *frameworks* que fornecem a possibilidade de implementar um motor de busca. Algumas destas são comerciais e outras de código aberto. A opção tomada passou por utilizar uma *framework* de código aberto e alterar o seu funcionamento de modo a corresponder às exigências do ADAPT. A *framework* utilizada na indexação e pesquisa deste módulo foi o *Lucene* e é descrita na subsecção seguinte.

4.1 Lucene

Implementado sobre a licença Apache (Apache Foundation, 2012) o projeto desta *framework* conta com a contribuição de mais de 80 pessoas no seu desenvolvimento e manutenção, disponibilizando uma panóplia de funcionalidades¹ tais como:

- Indexação e pesquisa de resultados;
- Pesquisa ordenada, onde os melhores resultados são mostrados em primeiro lugar;
- Variados tipos de *queries*;
- Pesquisa e ordenação por qualquer campo;
- Pesquisa em vários índices e agregação de resultados

Implementada em Java, embora existam versões em C++, .NET, C, Objective-C, PHP, Perl, Python, Delphi, Ruby e Common Lisp, a *framework* é multiplataforma. Dado que é código aberto, esta *framework* permite a alteração do código de maneira a ajustá-la à medida das necessidades de cada projeto, o que é uma vantagem quando se deseja alterar o modo de indexação dos objetos ou o algoritmo de ordenação dos resultados.

5 Desenvolvimento do motor de busca

Este subcapítulo descreve o desenvolvimento do motor de busca.

O diagrama de casos de uso e respetiva descrição podem ser consultados no anexo II, o seu diagrama de actividades e especificação no anexo III.

5.1 Indexação

A indexação, como já foi explicado, consiste na catalogação da informação existente de modo a formar índices segundo os quais é muito mais rápido realizar pesquisas do que sobre a totalidade da informação. Contudo, a indexação efetuada neste projeto é ligeiramente diferente da explicada anteriormente. Embora o conceito seja similar, no caso deste projeto indexou-se a informação contida numa base de dados MySQL e não a informação proveniente de páginas *Web* (um dos objetivos do projecto: Fornecer ao aluno conteúdos locais alternativos – da base de dados da plataforma ADAPT).

A indexação de conteúdos foi desenvolvida com base no método de indexação da *framework* Lucene, utilizando os métodos *IndexWriter* e *Document* (Lucene, 2013).

Os campos da base de dados do projeto a indexar pertencem às tabelas OA (onde são guardados os dados descritivos de cada conteúdo como por exemplo as suas preferências *vark* – relativas ao EA do OA) e OADetalhe (onde são guardados os detalhes de cada conteúdo):

- Id do OA – idOA – id do Objeto de aprendizagem;
- Designação do OA – designação – Designação do OA;
- Descrição do OA – texto – Descrição do OA.

¹ As funcionalidades descritas são apenas alguns exemplos de implementações da *framework* e de maneira nenhuma correspondem à totalidade das suas funcionalidades.

A figura 2.2 mostra as tabelas e campos em causa.

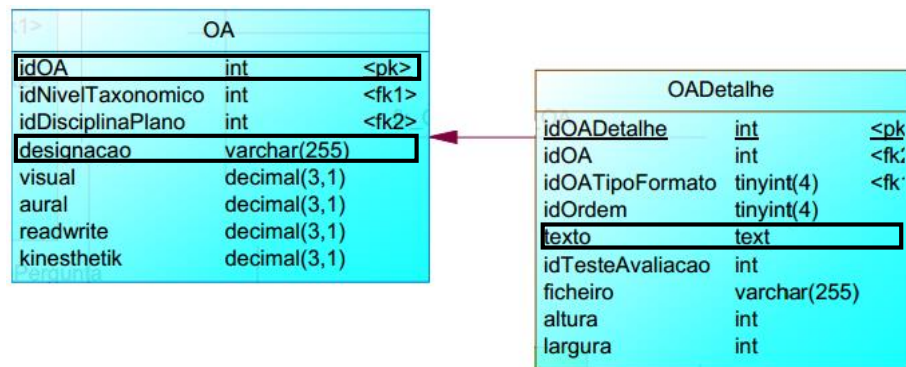


Figura 2.2 - Campos a indexar

Apesar da informação necessária a uma pesquisa estar agora indexada, falta relacionar a informação relativa à importância de cada OA. Para tal recorreu-se a *queries* SQL que têm em conta os campos da base de dados correspondentes à pontuação, dataInicio, dataFim, idOA, *visual*, *aural*, *readwrite* e *kinesthetic*, a fim de conseguir calcular a pontuação, tempo de navegação e estilo de aprendizagem predominante associados a cada artigo. A figura 2.3 ilustra os campos utilizados pertencentes às tabelas Utilizador (contém os dados dos utilizadores inscritos na plataforma), NavegacaoLog (contém os dados relativos às consultas de conteúdos por parte dos alunos), UtilizadorTesteAvaliacao (tabela que contém a nota de um utilizador) e OA:

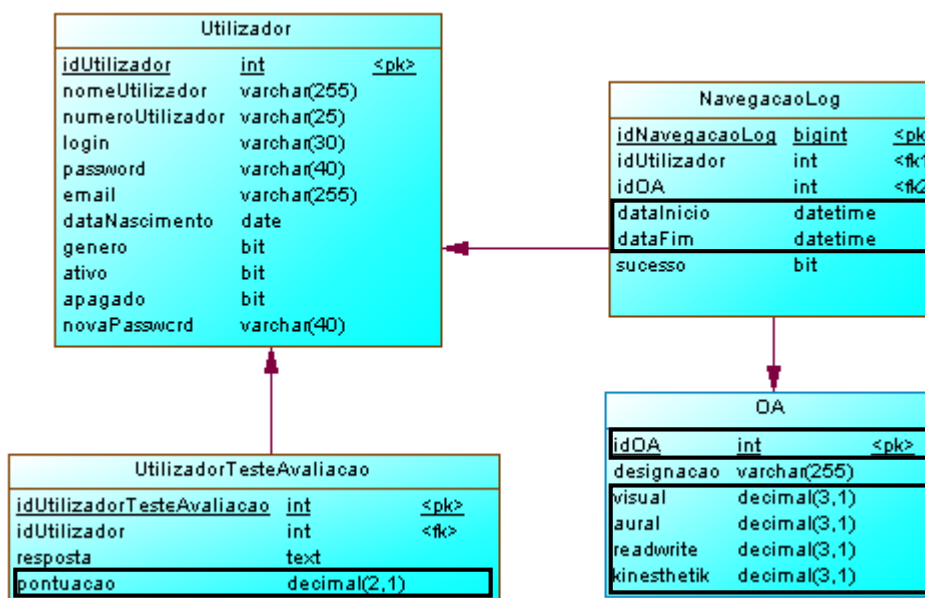


Figura 2.3 - Campos a indexar para a criação dos índices de cada OA.

Os resultados das *queries* SQL são utilizados num algoritmo que faz corresponder os valores de id, pontuação média e tempo médio de navegação de cada OA (em segundos), produzindo um ficheiro semelhante ao que se encontra na figura 2.4.

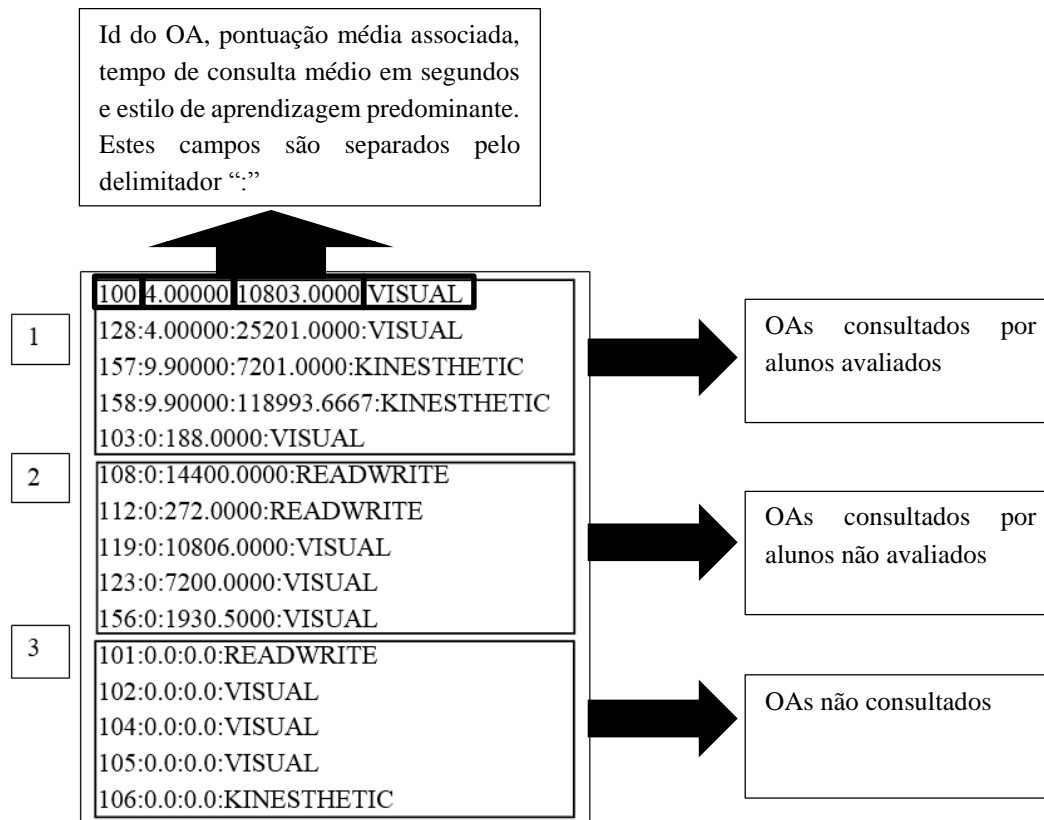


Figura 2.4 - Ficheiro que relaciona cada conteúdo com a sua pontuação, tempo de visualização e estilo de aprendizagem. (idOA : PontuaçãoAssociada : TempoConsulta : EstiloAprendizagemPredominante).

O ficheiro representado na figura 2.4 pode ler-se do seguinte modo:

- O OA de id 100 (predominantemente *visual*) foi visualizado por alunos que já foram avaliados.
- O OA de id 101 (predominantemente *readwrite*) é recente, dado que não foi visualizado por nenhum aluno;
- O OA de id 108 é predominantemente *readwrite* e foi visualizado por alguns alunos mas estes ainda não têm pontuação associada pois ainda não foram avaliados;

Este ficheiro relaciona o ID de cada OA com a sua pontuação, o seu tempo de visualização e o seu estilo de aprendizagem predominante em três passos distintos, obtendo os parâmetros correspondentemente aos seguintes passos:

1. Obter os OAs que foram consultados por alunos que foram avaliados;
2. Obter os OAs que foram consultados por alunos que não foram avaliados;
3. Obter os OAs que ainda não foram consultados.

Os três passos são necessários porque ao efetuar uma *query* SQL que obtenha os OAs consultados por alunos avaliados, não se consideram os OAs consultados por alunos não avaliados nem OAs recentes que ainda não foram consultados. Quer os OAs consultados por

alunos avaliados quer os OAs consultados por alunos por avaliar ou até os OAs que ainda não foram consultados poderão ser uma boa fonte de informação alternativa para o aluno e por isso têm de ser endereçados.

Especificando os três passos do algoritmo e seguindo a ordem crescente apresentada, o algoritmo consiste no seguinte:

1. Primeiramente obtêm-se os tempos de navegação dos OAs que foram consultados por alunos que foram já avaliados e portanto têm pontuação associada. É também calculado o estilo de aprendizagem predominante do OA e criada uma lista de inteiros (lista X) com os IDs dos OAs obtidos;
 - a. É criado o ficheiro apresentado na figura 2.4 onde são registados os OAs presentes na lista X;
2. Seguidamente o algoritmo obtém os OAs que já foram consultados por alunos que não foram ainda avaliados. É calculado o estilo de aprendizagem predominante do OA. Se os OAs obtidos ainda não existirem na lista X então são adicionados e são salvaguardados no ficheiro;
3. Por fim, o algoritmo verifica se o número de OAs na lista corresponde ao número total de OAs existentes. Se não corresponder, então os OAs que não constam na lista são adicionados ao ficheiro. Estes OAs (os que não foram ainda consultados e por isso não têm pontuação nem tempo de navegação associados) não são já adicionados á lista X pois esta servia apenas para saber quais os OAs já obtidos e quais os que faltam obter. Neste passo é ainda calculado o estilo de aprendizagem predominante dos OAs adicionando-os ao ficheiro;

O algoritmo tem como resultado o ficheiro apresentado pela figura 2.4 cuja informação, em conjunto com os índices fornecidos pelo *Lucene* formam os índices deste motor de busca.

Há que realçar que as notas de um determinado aluno X que consultou o OA Y apenas são somadas á pontuação do OA Y se o tempo de visualização deste por parte do aluno X for superior a 2 minutos (120 segundos – tempo mínimo para se considerar uma consulta intencional). O mesmo se passa com o tempo de visualização que apenas é somado ao tempo de visualização total do OA se o tempo de visualização do conteúdo Y pelo aluno X for maior que 120 segundos. Esta “limitação” prende-se com a garantia mínima de que o OA foi consultado propositadamente e não acidentalmente.

A indexação consiste pois num projeto á parte que permite ao administrador da plataforma fazer a indexação da base de dados quando desejar, ou programar a indexação por períodos de tempo. O agendamento do serviço de indexação é com certeza uma mais-valia, tendo em conta que muitas vezes os OAs são inseridos remotamente e, para que o motor de busca os possa “encontrar” durante uma pesquisa, estes terão obrigatoriamente de ser indexados.

5.2 Pesquisa

A pesquisa consiste na procura/consulta de informação sobre os índices criados. Neste caso específico, a pesquisa é efetuada em dois passos: a construção da *query* de pesquisa e a pesquisa nos índices pela informação relevante de acordo com a *query* em questão.

5.2.1 Construção de *queries*

Para entender o porquê de algumas opções tomadas, é essencial ter algum conhecimento geral sobre *N-Grams* e *Stemming*.

5.2.1.1 *N-Grams*

N-Gramming é o processo de subdivisão de uma determinada frase por palavras criando várias combinações possíveis. Nos algoritmos de *N-Gram*, N poderá tomar qualquer valor no intervalo de $[1, +\infty[$, estando apenas limitado aos recursos físicos da máquina (no entanto, há que ser razoável). Assim poderão ter-se *1-Gram (Unigram)*, *2-Gram (Bigram)*, *3-Gram (trigram)*, ..., *N-Gram*. Aplicando o algoritmo *N-Gram* com N igual a 3 na frase: “Simplificação Funções teoremas” o resultado é o seguinte:

Tabela 2.1 - *N-Grams*

Queries geradas
Simplificação
Simplificação funções
Simplificação funções teoremas
Funções ²
Funções teoremas
Teoremas

A implementação de algoritmos de *N-Gramming* é fundamental uma vez que permite abranger uma maior quantidade de combinações de palavras. Este algoritmo permite a criação de várias *queries* que poderão fazer sentido pesquisar, uma vez que estas combinações são compostas por palavras provenientes da frase original.

5.2.1.2 *Stemming*

Stemming consiste na redução das palavras á sua raiz e eliminação de plurais. Por exemplo, para a palavra “muitíssimo, o resultado deste algoritmo é “muito”. O algoritmo tem mais significado em inglês devido ao facto de esta língua usar frequentemente o tempo verbal gerúndio, embora possa ser utilizado em várias línguas.

² A implementação utilizada (https://lucene.apache.org/core/4_4_0/analyzers-common/org/apache/lucene/analysis/ngram/NGramTokenizer.html) não tem em conta todas as combinações, apenas junta as palavras mais próximas, neste caso, está em falta a sub-query “Simplificacao teoremas”

5.2.1.3 Processo de criação de *queries*

Ao receber uma frase de pesquisa (gerada pelo sistema a partir do título do OA que o aluno se encontra a visualizar no momento) o módulo tem, inicialmente, que tomar algumas precauções tendo em vista uma melhor performance. Assim, optou-se por em primeiro lugar retirar todos os caracteres especiais, palavras sem significado relevante (implementou-se um método para a remoção de palavras como “ou”, “mas”, etc...) e excessivos espaços em branco à *query* antes de esta poder estar apta para prosseguir. Seguidamente, a *query* passa por um processo de derivação que consiste na criação de vários subconjuntos de *queries*, formadas com base na frase inicial. As novas *queries* compõem 3 subconjuntos em que os dois últimos são construídos através do uso de *N-grams* (neste caso *3-Gram*, pois *2-Gram* ou *4-Gram* geram combinações ou muito grandes ou muito pequenas e, como tal, desinteressantes) e *Stemming*.

Para uma pesquisa pela frase: “Simplificação de Funções pelos teoremas”, as *queries* geradas seriam:

Tabela 2.2 - *Queries* geradas

<i>Queries</i> geradas	Sub-Conjunto	Precedência	Observações
simplificacao funcoes teoremas	1	1	Frase inicial
simplificacao	2	2	Obtido por “ <i>N-Gram</i> ”
simplificacao funcoes			
simplificacao funcoes teoremas ³			
funcoes			
funcoes teoremas			
teoremas	3	3	Obtido por “ <i>N-gramming</i> ” da frase de pesquisa “ <i>Stemmed</i> ”
simplificaca			
simplificaca funco			
simplificaca funco teorem			
funco			
funco teorem			
teorem			

As sub-*queries* criadas pertencem a 3 subconjuntos distintos:

- O primeiro subconjunto corresponde à frase de pesquisa por inteiro, permitindo pesquisar em primeiro lugar pela frase de pesquisa, que é o alvo mais significativo. Esta *query* é obviamente a que tem maior precedência, uma vez que consiste no título do OA que o utilizador está de momento a visualizar;

³ Neste caso, esta query gerada através de N-Gramming é a query original porque esta é composta por 3 palavras. Se a query tivesse 4 palavras, tal já não aconteceria.

- O segundo subconjunto corresponde a várias sub-*queries* obtidas por *N-Gramming* da frase de pesquisa original. Este subconjunto permite pesquisar vários conjuntos de combinações de três palavras pertencentes à frase de pesquisa original, aumentando a possibilidade de encontrar OAs relacionados com o que o aluno está a visualizar;
- O terceiro subconjunto corresponde também a sub-*queries* obtidas através de *N-Gramming* da frase de pesquisa mas apenas depois de sofrer um processo de *stemming*. Este terceiro sub-conjunto não é tão importante como os seus antecessores mas permite pesquisar por combinações de palavras na sua forma reduzida.

Após a criação destas *queries*, as mesmas são adicionadas a uma *boolean query* (Lucene API) – *query* que verifica de forma binária se um termo se encontra ou não num texto – que será utilizada posteriormente pela *framework Lucene* como objeto de pesquisa.

5.2.2 Algoritmo de pesquisa

Os OAs devolvidos pelo *Lucene* como “*match*” sofrem uma ordenação segundo os seguintes parâmetros:

- **Similaridade de Cosseno** – Esta medida representa a semelhança entre dois vetores de documentos, medindo o cosseno do ângulo entre si. Neste caso esta medida calcula a distância entre um termo de uma *query* e um documento (nome e descrição de um OA presente na base de dados). Este valor é calculado pela *framework Lucene* segundo a fórmula (Lucene, 2013):

$$\text{Score}(q, d) = \text{Coord}(q, d) * \text{QueryNorm}(q) * \sum_{t \text{ em } q} (tf(t \text{ em } d) * id(t)^2 * t.\text{GetBoost}() * \text{Norm}(t, d)) \quad (1)$$

- **Coord(q,d):** fator de coordenação que permite definir se um documento tem de fazer *match* com todos os termos de uma *query* multitermo ou não;
- **QueryNorm(q):** fator normalizante que não afeta o score, visto que todos os documentos são multiplicados pelo mesmo valor;
- **Tf(t em d):** frequência dos termos. I.e.: Número de vezes que o termo *t* aparece no documento *d*;
- **Id(t):** Número de documentos onde o termo aparece;
- **T.GetBoost():** importância do termo na *query*;
- **Norm(t,d):** encapsula alguns fatores normalizantes como por exemplo a normalização do tamanho do parágrafo (caso se aplique) e a potenciação (importância) de um dado termo;

- **Pontuação do Conteúdo** – Média das notas de todos os alunos que consultaram o OA. Esta pontuação pode ser zero, uma vez que o OA pode ou não ter sido consultado ou ser consultado por alunos que ainda não foram avaliados;
- **Tempo de consulta** – Média do tempo de consulta do OA por parte de todos os utilizadores da plataforma. O tempo de consulta pode ser zero, significando assim que o OA ainda não foi consultado;
- **Estilo de aprendizagem do OA** – Um OA do mesmo estilo de aprendizagem do aluno é mais importante que um que seja diferente. Assim, um OA de estilo diferente do aluno verá a sua pontuação penalizada em 50% face a um que seja do mesmo estilo;

A ordenação dos resultados de uma pesquisa é efetuada tendo em conta os seguintes critérios:

1. Ordenação dos OAs pelo seu valor de similaridade de cosseno – Assim, obtêm-se primeiramente os OAs que estão mais próximos do OA que o utilizador está de momento a consultar;
2. Em caso de empate dos valores da similaridade de cosseno de vários OAs (limitados a 4 casas decimais propositadamente a fim de possibilitar a utilização dos restantes parâmetros que acreditamos serem importantes na ordenação de resultados de um motor de busca para um LMS) o critério de desempate utilizará uma fórmula que relaciona os valores da pontuação, tempo de navegação e estilo de aprendizagem predominante do OA. A fórmula utilizada na criação da pontuação final para um artigo é a seguinte:

$$PontuaçãoFinal = \frac{Q1 * OATempoNavegacao + Q2 OAPontuacao}{Q1 + Q2 + EstiloAprendizagem} \quad (2)$$

O parâmetro estilo de aprendizagem pode tomar os valores de 0 ou 1 consoante é ou não do mesmo estilo de aprendizagem do aluno em questão, penalizando assim os artigos de estilo de aprendizagem diferente ao do aluno, face aos artigos do mesmo estilo de aprendizagem.

A Pontuação Final é um valor normalizado.

Não seria justo considerar apenas a pontuação ou o tempo total de navegação de um OA como critério adicional de desempate à similaridade de cosseno, pois se qualquer um dos dois campos fosse utilizado sozinho poderiam verificar-se os seguintes factos:

- O facto de a pontuação ser elevada pode não significar a boa qualidade do OA, mas sim a existência de um falso positivo, o que significa que o OA aparenta ser bom mas existe a possibilidade de muitos alunos o terem consultado durante pouco mais de dois minutos. O que torna o OA menos bom do que inicialmente se pensaria, ainda que consultado muitas vezes;
- O facto do tempo de navegação de determinado OA ser elevado pode, uma vez mais, não significar a sua qualidade, mas sim que o OA foi consultado imensas vezes mesmo

por alunos com notas menos boas, ou seja, o conteúdo que foi visualizado muitas vezes ou por um elevado período de tempo pode até estar na origem de maus resultados obtidos na disciplina.

Existe a possibilidade de um OA introduzir entropia no sistema i.e., ser considerado bom quando de facto não o é. Um exemplo é o caso de um OA ter uma designação sugestiva e levar os utilizadores a verificarem se é ou não interessante (o que em alguns casos poderá até levar mais que dois, três ou cinco minutos). Assim, o seu tempo de visualização acaba por se revelar bastante elevado, sobrevalorizando-o. A designação do OA é também por isso bastante importante e deve ser atribuída com precaução.

Através da junção dos dois parâmetros obtém-se um bom resultado pois se a pontuação e o tempo de navegação do OA forem altos, este terá uma boa cotação/qualidade, e se forem baixos terá uma cotação/qualidade mais baixa. A junção dos dois parâmetros traz algum balanceamento á escolha, tornando-a menos drástica.

5.3 Chamada do módulo

Ao invocar o módulo do motor de busca, a plataforma tem de especificar cinco parâmetros:

- **Frase a pesquisar** – Corresponde à pesquisa a efetuar;
- **Linguagem** – Corresponde à linguagem na qual a pesquisa será feita. Este parâmetro é essencial para saber que língua deve ser considerada ao efetuar o *stemming* da frase de pesquisa;
- **Coefficiente Q1** – Corresponde ao coeficiente a multiplicar pelo tempo de consulta;
- **Coefficiente Q2** – Corresponde ao coeficiente a multiplicar pela pontuação;
- **Id do aluno autenticado na plataforma** – Permitirá saber qual o aluno e qual o seu estilo de aprendizagem para o qual a sugestão de novos OAs será feita.

Ao invocar o módulo do motor de busca sobre determinada frase, este devolverá sempre resultados, dado que a pesquisa é efetuada com base no OA que o utilizador está de momento a visualizar e portanto, pelo menos um resultado será sempre mostrado. Os resultados devolvidos serão posteriormente apresentados ao utilizador sob a forma de *links*, permitindo assim ao aluno navegar para o conteúdo que desejar.

5.4 Interação do módulo com o aluno

O aluno em si não interage com o módulo em causa, limitando-se apenas a visualizar a listagem de conteúdos que lhe são sugeridos pelo módulo com base naquele que se encontra de momento a visualizar. A única opção que lhe é fornecida é o facto de poder reordenar os resultados como desejar.

5.4.1 Reordenação de resultados

A função de reordenação permite ao utilizador final reordenar a listagem fornecida pelo motor de busca como pretender:

- Modo por defeito: modo de ordenação por defeito do motor de busca – ordenação segundo a similaridade de cosseno e, em caso de empate, segundo a eq. (2);
- Combinação dos três parâmetros: consiste na ordenação por um ou mais parâmetros fornecidos pelo motor de busca. Em caso de combinação de parâmetros, estes são somados (e.g.: tempo de navegação + pontuação do OA) e os resultados ordenados de forma decrescente.

6 Performance

Os testes de performance deste motor de busca assentam essencialmente em dois testes. O primeiro testa a capacidade deste em encontrar conteúdos próximos ao que o aluno está a consultar e o segundo testa a ordenação dos resultados. O *dataset* utilizado nas experiências está especificado no anexo X.

6.1 Precision, *recall* e *accuracy*

A performance de um motor de busca é normalmente testada recorrendo a três medidas: *accuracy*, precisão e *recall*. Estas medidas são calculadas através da construção de uma tabela de confusão onde figuram os valores de VP (verdadeiros positivos), VN (verdadeiros negativos), FP (falsos positivos) e FN (falsos negativos) dos testes efetuados.

Para este teste, criaram-se 30 *queries* de 5 classes diferentes. O teste efectuado passa por pesquisar cada *query* (cada linha) e calcular o número de resultados VP, VN, FP e FN calculando posteriormente a *accuracy*, precisão e *recall* associadas ao motor busca.

Correntemente existem 78 OAs de 5 classes diferentes na BD do projeto. Entenda-se como classe a palavra-chave associada a cada OA (ver campo *palavrachave* da tabela OAPalavrachave do Anexo I) – que não é indexada ou pesquisada. Uma vez que o motor de busca pesquisa no título e descrição dos conteúdos (que nem sempre foram inseridos por peritos na matéria) não contextualiza um assunto, sendo de carácter binário (encontra ou não encontra as *queries* a pesquisar). Assim é natural que o número de falsos negativos ou falsos positivos seja elevado pois imagine-se os seguintes exemplos:

- Pesquisar pela *query* “tabelas de verdade”: Como as tabelas de verdade estão implicitamente associadas aos mapas de karnaugh, a sua classe é efetivamente mapas de karnaugh, o que não quer dizer que na pesquisa sejam contabilizados todos os conteúdos da classe “mapas de karnaugh”. O que foi pesquisado foi efetivamente “tabelas de verdade” e não qualquer outro conteúdo pertencente à classe “mapas de karnaugh”. Assim o valor dos FN neste caso poderia elevar-se, baixando a *recall*;
- Imagine-se agora que a pesquisa é por exemplo “soma lógica”. Neste caso a classe seria “componentes” mas poderia surgir também algum conteúdo que não pertencesse à classe que porventura tivesse as palavras “soma lógica”. Logo, o número de FP poderia também ser elevado, baixando a precisão.

Os dois exemplos descritos são dois exemplos que podem de facto alterar a performance do motor de busca baixando a precisão, a *recall* e a *accuracy*. O facto de o número de OAs ser reduzido, não existirem peritos indiferentes ao desenvolvimento do projecto e a falta de introdução de dados na BD por peritos na matéria poderão ser factores condicionantes na performance do módulo.

A precisão é a proporção de casos positivos previstos que estavam corretos e é calculada de acordo com a seguinte fórmula:

$$Precisão = \frac{VP}{FP + VP} \quad (3)$$

A *recall* é a proporção de casos positivos que foram corretamente identificados e é calculada de acordo com a seguinte fórmula:

$$Recall = \frac{VP}{FN + VP} \quad (4)$$

A *accuracy* é o valor global de valores que o motor de busca acerta e é dada pela seguinte fórmula:

$$Accuracy = \frac{VP + FN}{VP + VN + FP + FN} \quad (5)$$

Tendo em conta as fórmulas anteriores, a *accuracy*, precisão e *recall* do motor de busca na pesquisa das 30 *queries* em 78 OAs diferentes é a seguinte:

Tabela 2.3 - Testes de performance – *Accuracy*, precisão e *recall*

Precisão	73,14%
<i>Recall</i>	40,50%
<i>Accuraccy</i>	27%

6.2 Ordenação de resultados

Uma vez testada a *accuracy*, precisão e *recall* do motor de busca na pesquisa de documentos relacionados com o que o utilizador está a consultar, é necessário testar a ordenação dos resultados a fim de verificar se esta é a melhor.

Para testar o desempenho na ordenação de resultados levada a cabo por este motor de busca, foi criado um modelo onde constam os 10 primeiros resultados – de acordo com (Silverstein et al, 1999), um utilizador de um motor de busca geralmente consulta apenas os dez primeiros resultados – esperados para 30 *queries*. Os 10 primeiros resultados são ordenados pela proximidade à frase de pesquisa e cuja pontuação e tempo de navegação esteja em concordância com o seguinte:

- A pontuação não deve ser tão alta quanto o tempo de navegação: isto poderia significar que uma grande quantidade de alunos consultaram o artigo, mas por um curto período de tempo;
- A diferença entre o tempo de navegação e a pontuação não deve ser muito elevado. Um elevado tempo de navegação em comparação com uma pontuação baixa pode significar que a OA foi consultado por apenas alguns alunos com notas baixas por um período de tempo alongado;
- De preferência, o tempo de navegação e pontuação são tão elevados quanto possível mas a diferença entre eles deve ser pequena. Isto assegura que a consulta do OA é feita regularmente.

As afirmações anteriores consistem em regras seguidas na criação do modelo. No entanto, nem todos os OAs se encaixam neste modelo. Alguns casos podem ser complexos: um OA pode ter uma boa relação entre os valores de pontuação e tempo de navegação e aparentemente ser um OA que foi consultado várias vezes por um longo período de tempo e por estudantes com boas notas mas no entanto pode nem sequer ter sido lido, apenas “deixado aberto”.

Para testar a ordenação de resultados levada a cabo pelo motor de busca do ADAPT comparou-se a ordem dos resultados (OAs) no modelo com a ordem dos resultados do motor de busca calculando a distância de *Hamming* entre a posição (nos 10 resultados) onde o resultado X aparece e onde deveria aparecer no modelo.

Considere-se o seguinte exemplo em que cada linha apresenta os resultados esperados para uma *query* (modelo) e os resultados que efetivamente se verificaram:

Tabela 2.4 - Testes de performance – Ordenação

Modelo	Resultados	Distância de Hamming
1, 2, 3	2, 1, 3	2
4, 5, 6	6, 5, 4	4
		6

A distância de *Hamming* entre a primeira linha dos resultados e a primeira linha do modelo é 2 porque ambos os resultados 1 e 2 têm as suas posições trocadas estando a uma posição de distância da sua posição devida.

Na segunda linha do modelo a distância é 4 porque ambos os resultados 4 e 6 se encontram a uma distância de 2 posições da sua posição no modelo. A distância dos resultados neste exemplo ao modelo é efetivamente 6 e corresponde à soma das distâncias entre as linhas de resultados e do modelo.

Neste teste tem ainda de se ter em conta que poderão existir resultados que não constam no modelo e por isso, tais resultados têm de ser penalizados com uma penalização “*p*”.

Para um modelo com 30 *queries* e uma penalização $p=1$, a distância de *Hamming* dos 10 primeiros resultados ao modelo é 108, a sua média é de 3,6 e o desvio padrão é de 4,30.

Estes testes foram realizados num reduzido mas crescente conjunto de dados. Por enquanto, este *dataset* de teste tem apenas dados experimentais e é constituído por 78 OAs de 5 classes. No entanto, pelos testes anteriores, pode concluir-se que a nossa metodologia é viável.

De futuro espera-se melhorar os testes efectuados, comparando os resultados do motor de busca com um modelo gerado por consenso de vários curadores.

A figura 2.5 mostra o resultado de uma pesquisa efetuada. Os números à esquerda representam os Ids dos OAs (para efeitos de teste). Em baixo pode ver-se a interface de reordenação de resultados:

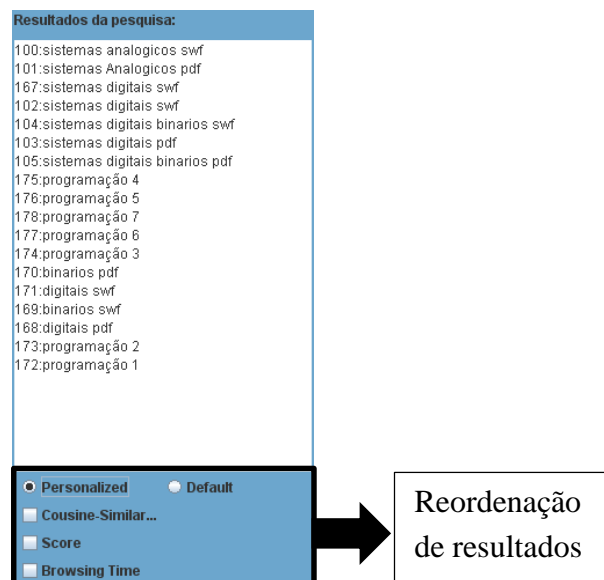


Figura 2.5 - Imagem contextual do módulo do motor de busca em funcionamento.

7 Conclusões

O contexto social de um OA no âmbito de um sistema de *e-Learning* é muito importante pois encontra-se diretamente relacionado com a qualidade do mesmo. No desenvolvimento de um motor de busca para este tipo de sistemas, a necessidade da incorporação desta medida torna-se imperativa pelo facto de juntar alguma pedagogia na sua função de ordenação, evitando assim uma busca naïve que procure apenas por OAs com conteúdos semelhantes à frase de pesquisa.

Embora tenham sido usados apenas dados experimentais, pode concluir-se que a metodologia é viável: O motor de busca apresenta uma *accuracy* de 27%, precisão de 73,14% e *recall* de 40,50% na procura dos OAs mais adequados a cada aluno. Na ordenação de resultados o módulo apresenta uma distância de *Hamming* ao modelo criado de 108, a sua média é de 3,6 e o desvio padrão é de 4,30.

CAPÍTULO III – CLASSIFICADOR DE ESTILOS DE APRENDIZAGEM

Este capítulo descreve o classificador de estilos de aprendizagem desenvolvido e está organizado da seguinte forma: introdução; estado de arte; desenvolvimento do classificador; interação com a plataforma; performance e por fim as conclusões retiradas.

1 Introdução

Ao longo da sua vida os indivíduos podem ir modificando o seu estilo de aprendizagem mesmo que disso não se apercebam. Por exemplo, uma pessoa que inicialmente preferia aprender através de conteúdos com uma componente de leitura mais forte, pode começar a preferir aprender através de conteúdos mais dinâmicos tais como, por exemplo, vídeos.

Este capítulo descreve o desenvolvimento de um módulo para a plataforma ADAPT que proporciona um modo de verificar se os utilizadores estão ou não a modificar o seu estilo de aprendizagem. O módulo classifica um determinado conteúdo consultado numa das quatro dimensões previstas pelo modelo VARK (*visual, aural, readwrite, kinesthetic*), verificando se o perfil de aprendizagem de cada aluno está em concordância com os conteúdos que ele tem consultado nas semanas antecedentes. Se não estiver em concordância, o módulo alerta para esse facto e sugere ao ITS do sistema ADAPT uma adaptação (mudança) do guião do curso para aquele indivíduo.

A motivação para o desenvolvimento deste módulo está portanto diretamente relacionada com as características adaptativas do ADAPT, conferindo um certo grau de inteligência à plataforma. O objetivo é não limitar o aluno apenas ao seu estilo de aprendizagem inicial mas acompanhá-lo ao longo do seu percurso académico.

O método atinge uma *accuracy* de 75%.

2 Estado de arte

A documentação existente é vasta, pois a área de *link-mining*, na qual este módulo se insere, é uma área em desenvolvimento com vários ramos de bastante interesse, como por exemplo: bioinformática, estudo de comportamentos criminosos, avaliação de preferências em redes sociais (este tópico é bastante vasto), etc. O tema sob o qual nos focamos é *link-mining* em *Learning Management Systems* (LMS – Sistemas de gestão de aprendizagem). Existe de facto muita atividade na aplicação de *link-mining* e *data mining* a LMS's. Alguns dos trabalhos que se destacam são os seguintes:

O artigo (Getoor et al, 2005) consiste num survey sobre *link-mining* e especifica que habitualmente é usado *clustering* aglomerativo para criar grupos de diferentes interesses chamados *positions*. Posteriormente é utilizado um algoritmo estocástico descrito em (Nowicki et al, 1997) para dividir os utilizadores segundo esses grupos. O mesmo trabalho refere ainda que os dados são normalmente apresentados em formato de grafo e nas tarefas de classificação são utilizados algoritmos de aprendizagem supervisionada.

Em (Romero et al, 2008) os autores apresentam uma ferramenta que corre sob o Moodle para classificar os alunos e predizer as suas notas. Segundo eles, os melhores resultados foram adquiridos com os algoritmos CART (aplicado em árvores de decisão), GAP (aplicado em indução de regras difusas), GGP (utilizado na indução de regras) e NNEP (utilizado em redes neuronais para geração de regras de associação). Os melhores algoritmos, utilizando dados categóricos foram os algoritmos CART e C4.5 (65% de g-mean). Por fim, afirmam que os resultados não foram superiores a 70% porque utilizaram dados incompletos. O trabalho não explica como foram aplicados os algoritmos, embora forneça conhecimento de que foram utilizadas árvores de decisão no processo.

O estudo (Graf et al, 2008) propõe um método automático para identificar mudanças de estilos de aprendizagem baseados no modelo Felder-Silverman. O seu algoritmo consiste em primeiramente contabilizar as visitas aos *links*. Estes resultados são designados por pistas. Estas pistas são classificadas posteriormente em quatro valores: “3” - totalmente de acordo com o seu EA (Estilo de aprendizagem); “2” - mais ou menos de acordo; “1” - em desacordo com o seu EA; “0” - não existe qualquer informação disponível. Para classificar os comportamentos dos alunos os autores recorreram a *thresholds* propostos pela literatura tal como por exemplo o número de vezes que um aluno deve participar num fórum. Por fim somam todas as pistas e dividem pelo número de padrões (apenas dos quais existe informação disponível). O respetivo resultado é normalizado e o resultado final representa a percentagem de preferência do aluno pelo seu EA.

O artigo (Zorrilla et al, 2010) apresenta um sistema para aplicar sobre LMS's e efetuar *data mining* sobre os seus dados. O sistema utiliza uma *DataWarehouse* para guardar as informações dos utilizadores, utiliza o *K-means* (ou algoritmo de Lloyd) como algoritmo de *clustering* e por fim o algoritmo *Apriori* para obter algumas regras sobre os utilizadores.

Os estudos descritos são apenas os que maior relevância tiveram para o desenvolvimento deste módulo. Conforme referido, esta área de investigação é bastante vasta.

3 Desenvolvimento do classificador de estilos de aprendizagem

Esta secção explica a implementação deste algoritmo. São explicados quais os campos da base de dados que são utilizados na construção deste módulo bem como o algoritmo implementado e os seus testes de performance.

O diagrama de casos de uso e respetiva descrição estão especificados no anexo IV. O seu diagrama de actividades e especificação podem ser consultados no anexo V.

3.1 Campos da base de dados a ter em conta

Os principais campos tomados em linha de conta neste módulo foram “*visual*”, “*aural*”, “*readwrite*” e “*kinesthetic*” das tabelas “Vark” (definem a percentagem de cada estilo de aprendizagem do aluno) e “OA” (que contém os atributos dos conteúdos existentes). O campo “*idNivelTaxonomico*” da tabela *NivelTaxonomico* é também importante, pois especifica a

dificuldade técnica de um OA de acordo com as taxonimias de Bloom⁴ (Ferraz et al, 2010). Por fim, o campo “idUtilizador” permite relacionar todos os outros, uma vez que identifica o aluno. Os campos “dataInicio” e “dataFim” da tabela NavegacaoLog (contém os dados relativos às consultas efetuadas pelos utilizadores) são utilizados para calcular o tempo de permanência em cada OA (endereçado por um *link*). A tabela LearningStyle corresponde às características dos alunos no modelo Felder-Silverman que são traduzidas nas preferências Vark, as quais são utilizadas neste módulo.

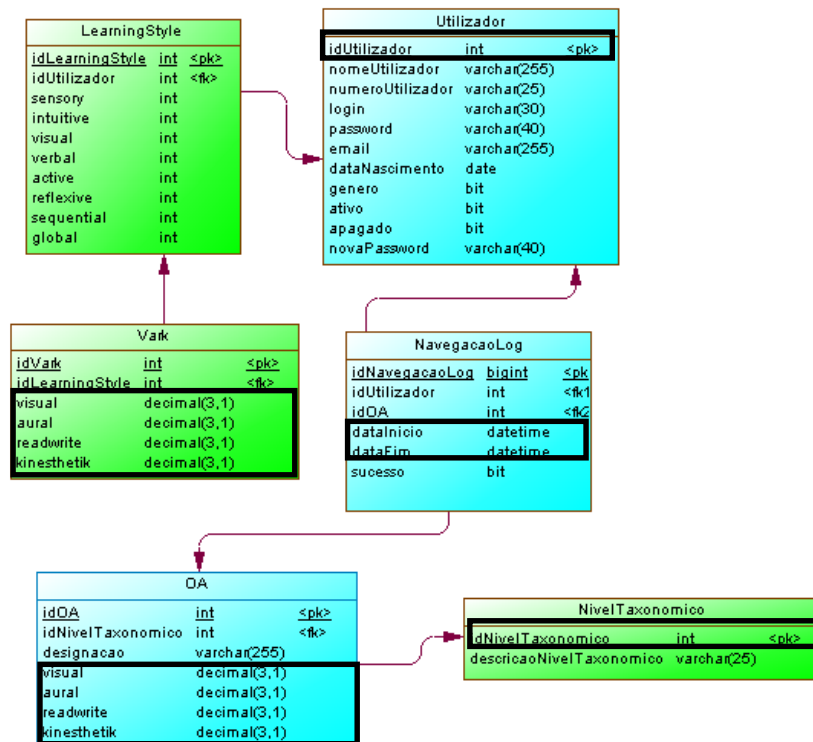


Figura 3.1 - Tabelas da base de dados a ter em conta e campos a endereçar.

A figura anterior mostra as tabelas e respetivos campos da base de dados do projeto que foram utilizados neste módulo.

3.2 Algoritmo implementado

Em primeiro lugar, quando o módulo é chamado, é efetuada uma *query* SQL de modo a verificar quais os OAs que cada utilizador tem consultado (uma consulta implica um tempo de permanência no mínimo de 2 minutos no OA em questão) em X semanas (sendo que X é definido na chamada do módulo), obtendo-se a informação representada na figura 3.2.

⁴ Taxonomia que compreende 6 níveis: 1 – conhecimento; 2 – compreensão; 3 – aplicação; 4 – análise; 5 – Síntese; 6 – avaliação.

IDUSER	VISUAL	AURAL	READWRITE	KINESTHETIK	TempoPermanecidoNoLink	BLOOM
26	50.0	0.0	30.0	20.0	4500	6
26	0.0	99.9	0.0	0.0	36482	5
26	50.0	0.0	30.0	20.0	1986	6
26	30.0	0.0	20.0	50.0	4080	3
26	35.0	0.0	30.0	35.0	1206	5
26	50.0	50.0	0.0	0.0	1380	4
26	15.0	0.0	90.0	50.0	7800	1
27	65.0	0.0	50.0	30.0	30780	3
27	50.0	0.0	30.0	20.0	5100	6

Figura 3.2 - Resultados obtidos pela *query* inicial.

Depois de obter os resultados da *query* anterior, determina-se o EA predominante de cada OA, que é calculado com base no parâmetro VARK (*visual*, *aural*, *readwrite* ou *kinesthetic*) mais elevado. Deste modo, o primeiro OA consultado terá a classificação de *visual*. O segundo OA será *aural* e assim sucessivamente. Na fig. 3.2, cada linha corresponde a uma consulta efetuada por um utilizador e portanto a um OA visualizado.

Os OAs consultados por cada aluno são separados por estilo de aprendizagem e os parâmetros “TempoPermanênciaLink” (tempo de consulta do OA) e “BLOOM” (classificação do OA de acordo com as taxonomias de Bloom) são somados ao contador de OAs de cada estilo de aprendizagem para cada aluno de acordo com a eq. (6):

$$M_i = \sum_{j=0}^l (\text{TempoPermanênciaLink} * \text{bloom}) \quad (6)$$

- “ M_i ” – Valor da pontuação para a preferência VARK “ i ”.
- “ I ” – Representa o número de artigos que o aluno consultou da preferência VARK “ i ”;
- “ j ” – Representa o iterador de artigos consultados pelo aluno para a preferência VARK “ i ”;
- “ i ” – Preferência VARK. Pode tomar o valor de *visual*, *aural*, *readwrite* e *kinesthetic*;

A eq. (6) gera um valor que corresponde à contribuição “ M ” do OA “ j ” para a preferência VARK “ i ” de cada aluno. Desta equação resultam as novas preferências VARK do aluno obtidas através da multiplicação do “TempoPermanênciaLink” e “bloom” de cada OA consultado.

Ambos os parâmetros “TempoPermanência” e “bloom” são normalizados e surgem porque é importante analisar o tempo de permanência no *link*, que pode ser influenciado pelo grau de dificuldade do mesmo.

Por fim é verificado se o estilo de aprendizagem de cada utilizador está de acordo com o tipo de OAs que predominantemente consultou durante as X semanas anteriores atrás referidas. Se não estiver de acordo com o seu estilo de aprendizagem, o módulo sugere ao ITS da plataforma ADAPT uma adaptação do guião do curso segundo os novos parâmetros VARK

do aluno. O novo estilo de aprendizagem do aluno é dado por M_j ($w\%$ *visual*, $x\%$ de *aural*, $y\%$ de *readwrite* e $z\%$ de *kinesthetic*) calculados na eq. (6).

4 Interação com a plataforma

O classificador é chamado periodicamente pela plataforma e devolve um vetor da classe “Utilizadores” que tem como atributos o id de cada aluno, uma *flag* que especifica se o aluno está ou não a mudar de EA e os novos parâmetros VARK do utilizador. A plataforma verifica os utilizadores que necessitam de um novo EA e procede à mudança⁵, baseando-se nos dados dos parâmetros VARK dos utilizadores.

5 Performance

Para calcular a performance deste módulo, criou-se um *dataset*⁶ de teste composto por consultas a OAs por parte de 12 utilizadores da plataforma que à falta de um modelo curado por especialistas permite testar o classificador. De notar que devido à falta de dados reais, pelo facto de até ao momento existir pouca utilização da plataforma, muitos foram introduzidos deliberadamente a fim de criar um *dataset* com um volume de informação adequado.

Para efeitos de teste, foi criado um modelo que definisse se um utilizador está ou não a mudar o seu perfil de aprendizagem. Este modelo verifica a quantidade de OAs, de cada EA predominante (preferências VARK predominantes), consultados por cada utilizador segundo a eq. (7):

$$M_i = \sum_{j=0}^l (M_i + 1) \quad (7)$$

- “ M_i ” – Valor da pontuação para a preferência VARK “ i ”.
- “ I ” – Representa o número de artigos que o aluno consultou da preferência VARK “ i ”;
- “ j ” – Representa o iterador de artigos consultados pelo aluno para a preferência VARK “ i ” ;
- “ i ” – Preferência VARK. Pode tomar o valor de *visual*, *aural*, *readwrite* e *kinesthetic*;

O teste efetuado avalia a performance da fórmula do classificador desenvolvido contra este modelo, efetuando um juízo de relevância binária, i.e., verifica se o modelo e o classificador estão em concordância no facto de um aluno modificar o seu estilo de aprendizagem, e se estiverem, se estão de acordo na predominância do seu novo estilo de aprendizagem. E.g.: o

⁵ A criação de um novo guião ou a alteração do atual está a cargo do ITS do sistema ADAPT, que procederá de acordo com estes dados na fase de adaptação do ciclo CBR recorrendo a algoritmos genéticos para encontrar uma solução ótima, próxima do perfil de aprendizagem encontrado.

⁶ Anexo XI.

modelo afirma que um determinado aluno predominantemente *aural* tem consultado – segundo a eq. (7) – 60% de OAs predominantemente *visuais*, 10% *aurais*, 20% *readwrite* e 10% *kinesthetick* e portanto, com base nestes valores, terá de modificar o seu estilo de aprendizagem. Se o classificador apresentar – segundo a eq. (6) – que o utilizador consultou – 70% de OAs predominantemente *visuais*, 5% *aurais*, 20% *readwrite* e 5% *kinesthetic*, os resultados do modelo e do classificador estão em concordância (o utilizador veria o seu EA modificado para um predominantemente *visual*), pelo que resultado é “1”. Se o classificador apresentar 30% *visuais*, 10% *aurais*, 50% *readwrite* e 10% *kinesthetick*, o resultado da experiência seria “0”.

O modelo e classificador foram aplicados sobre o *dataset* composto por algumas consultas a OAs. Os resultados de *accuracy* são calculados através do número de concordâncias entre o classificador e o modelo, sobre o número total de tuplos. Os resultados para 2, 5 e 10 semanas anteriores à atual são apresentados na tabela 3.1.

Tabela 3.1 - Teste de performance do módulo 2.

Semanas	Accuracy
2	75%
5	65%
10	65%

Os resultados mostram a viabilidade desta abordagem, especialmente se se tiver em conta que o modelo criado por si só é bastante naïve, pois apenas verifica o número de OAs consultados de um determinado estilo de aprendizagem, definindo assim a necessidade ou não da criação de um novo EA para o aluno.

6 Conclusões

O módulo criado é um grande avanço na automatização da plataforma ADAPT pois possibilita ao ITS utilizar os dados produzidos pelo classificador no ciclo CBR e através de algoritmos genéticos, encontrar uma solução ótima para a criação de novos guiões, minimizando a distância entre os guiões criados e o perfil de aprendizagem encontrado pelo classificador.

Comparando com o modelo criado, o módulo produz bons resultados de *accuracy* na identificação de alterações nos perfis de aprendizagem de 12 alunos teste em 2 (75%), 5 (65%) e 10 semanas (65%).

Como trabalho futuro espera-se validar a metodologia recorrendo ao questionário ILS que valida os resultados do módulo pois determina as preferências atuais dos alunos. Este modelo não foi utilizado como teste por falta de dados de navegação reais e alunos que utilizem a plataforma (que como já foi dito, ainda não está totalmente em funcionamento, o que torna natural a sua falta de utilização).

CAPÍTULO IV – PLATAFORMA DE ANÁLISE DE DADOS

Este capítulo descreve o desenvolvimento de uma plataforma de análise dos dados contidos na base de dados do projeto ADAPT. O capítulo organiza-se em 4 sub capítulos: introdução; estado de arte; frameworks utilizadas; desenvolvimento do módulo e conclusões.

1 Introdução

A maioria dos sistemas de *e-Learning* não possui modelos de visualização de dados que facilitem a visualização do progresso dos alunos inscritos na plataforma. Este capítulo apresenta o desenvolvimento de um módulo destinado à exploração dos dados que os utilizadores produzem e daí retirar informação útil para os professores/tutores de modo a que estes identifiquem falhas, lacunas, comportamentos ou grupos que necessitem de ser modificados a fim de melhorar a experiência e o percurso académico dos alunos utilizadores da plataforma ADAPT. A motivação decorre, assim, da conveniência da criação de algoritmos de análise estatística e *data mining* que visam ajudar os professores a melhorar alguns OAs ou a descobrir padrões potencialmente interessantes.

2 Estado de arte

A aplicação de algoritmos de *data mining* aos dados criados pelos alunos de uma plataforma de *e-Learning* é uma área em desenvolvimento embora já exista imensa literatura sobre o assunto. O artigo (Castro et al, 2007) apresenta um *overview* sobre alguns dos algoritmos mais utilizados. São exemplo: redes neuronais, algoritmos genéticos, *clustering* e métodos de visualização, lógica difusa e agentes indutivos, entre outros.

A extração de dados é feita na sua maioria recorrendo a algoritmos de *clustering*, regras de associação (Valsamidis et al, 2011) e classificação.

A nível de *clustering* (Sael et al, 2013) (Zakrzewska et al, 2008), o algoritmo que normalmente é utilizado é o *K-means*, embora também sejam utilizados outros, tais como o EM ou *clustering* hierárquico aglomerativo.

Os algoritmos de extração de regras de associação são também bastante utilizados nesta área. O mais comum é o algoritmo *Apriori* (Romero et al, 2010).

Por fim, o artigo (Mihăescu et al, 2011) apresenta uma implementação que utiliza regressão linear de modo a prever a quantidade de conhecimento acumulado por cada aluno: "*This paper uses linear regression for modeling the quantity of accumulated knowledge in relationship with variables representing the performed activity.*"

A nível de técnicas de redução de dimensionalidade de dados a mais comum é o PCA (*Principal Component Analysis*) (Jolliffe, 2002).

3 Frameworks utilizadas no âmbito deste módulo

A *framework* base para o desenvolvimento deste módulo foi o WEKA (API do WEKA, 2014), uma *framework* de código aberto implementada em Java. Dentro do vasto número de algoritmos apresentados pelo weka podem distinguir-se 3 conjuntos:

- Classificação;
- *Clustering*;
- Associação;

Cada conjunto apresenta vários algoritmos de características diferentes.

Esta *framework* permite também criar processos de fluxo de dados para que se possam redirecionar os dados de um *dataset* por vários algoritmos diferentes, funcionalidade bastante útil, quando para extrair informação do *dataset* é necessária a utilização de pré-processamento de dados e passagem por dois ou mais algoritmos de *data mining*.

A API do weka para Java é também extremamente simples pelo que se torna fácil a sua utilização.

A geração de gráficos que pode ser vista durante este capítulo foi conseguida através da *framework* JFreeChart (JFreeChart, 2014), uma das *frameworks open source* de Java mais conhecidas.

4 Desenvolvimento da plataforma

Esta secção apresenta o desenvolvimento detalhado da plataforma.

O diagrama de casos de uso e sua respectiva descrição encontram-se especificados no anexo VI. O seu diagrama de actividades e correspondente especificação pode ser consultado no anexo VII.

4.1 Objetivo do Módulo

O objetivo do módulo consiste em responder às seguintes questões:

1. Identificação de grupos de utilizadores consoante o seu sexo, idade, estilo de aprendizagem predominante e conteúdos consultados;
2. Média das notas por estilo de aprendizagem, por disciplina;
3. Verificação de um possível impacto da idade e estilo de aprendizagem nos resultados alcançados;
 - a. Percentagem de visualização de artigos por estilo de aprendizagem: está de acordo com os que conduzem a melhor nota?
4. Percentagem de alunos cujo estilo de aprendizagem se alterou: quais as mudanças mais frequentes?
5. Média das notas e predominância de estilos de aprendizagem por cada disciplina de cada curso;
6. Distribuição dos alunos por estilo de aprendizagem;
7. Média das notas dos alunos por cada disciplina;

8. Predição das notas dos alunos que ainda não foram avaliados com base nas notas dos seus colegas e nos OAs que têm vindo a consultar;
9. Identificação de grupos de utilizadores consoante os seus parâmetros VARK, a média e desvio padrão dos parâmetros VARK dos OAs que consultou, número de OAs consultados e pontuação média.

4.2 Identificação de grupos de utilizadores

Para responder à questão 1 recorreu-se a técnicas de *clustering*.

4.2.1 Dataset

O *dataset*⁷ necessário para responder a esta pergunta foi extraído da base de dados recorrendo a *queries* SQL. Os grupos de utilizadores são identificados pelos atributos género, idade, estilo de aprendizagem predominante e percentagem de cada estilo de aprendizagem dos OAs que têm vindo a consultar.

O *dataset*, em formato CSV, foi obtido através dos campos identificados nas tabelas Utilizador (campos relativos ao id do utilizador, à sua data de nascimento e género), LearningStyle e Vark (preferências vark: *visual*, *aural*, *readwrite* e *kinesthetic*) apresentados na figura 4.1.

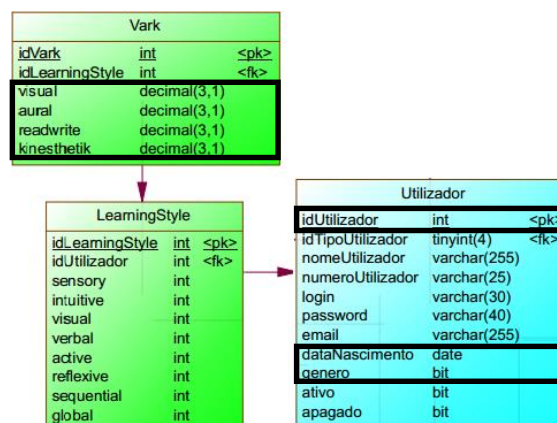


Figura 4.1 - Tabelas utilizadas para retirar os dados necessários para a pergunta 1.

Para calcular o número de OAs de cada estilo de aprendizagem que cada aluno tem vindo a consultar é utilizado o classificador de estilos de aprendizagem apresentado no Capítulo III, que nos fornece o estilo de aprendizagem de cada aluno, o facto de este estar ou não a alterar-se e a percentagem de OAs consultados para cada tipo de estilo de aprendizagem. Cruzando os dados da *query* acima referida com os dados obtidos pelo classificador de estilos de aprendizagem consegue criar-se o *dataset* pretendido (figura 4.2):

⁷ O *dataset* criado encontra-se especificado no anexo XII.

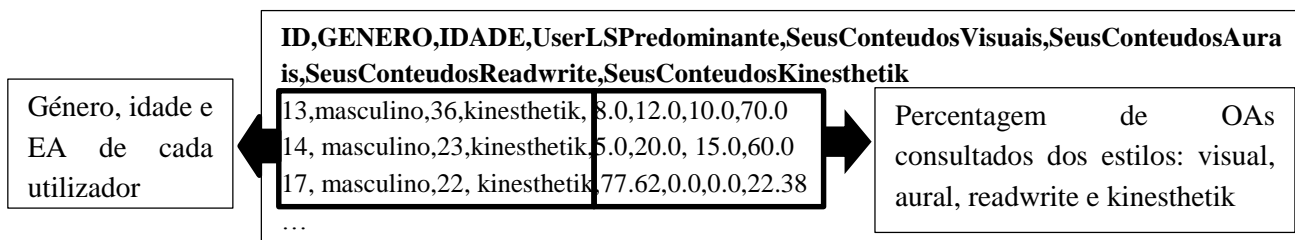


Figura 4.2- Dataset da query

Por exemplo, de acordo com a figura 4.2 o utilizador de id 13 é masculino, tem 36 anos, o seu estilo de aprendizagem predominante é *kinesthetic* e os OAs que mais consulta são do estilo *kinesthetic* (70%).

4.2.2 Aplicação do algoritmo de *Clustering*

Os algoritmos de *clustering* permitem verificar se um conjunto de dados está ou não relacionado e conseqüentemente de que forma devem ser agrupados. O algoritmo utilizado para verificar a existência de grupos neste *dataset* foi o *K-means* (K-means, 2014) (pelo facto de o *dataset* utilizar dados nominais), que cria *k clusters*, em que *k* é especificado na sua invocação. O *K-means* começa por criar *k* centroides que inicialmente são posicionados aleatoriamente. Ao longo de *X* iterações, o algoritmo recalcula a posição dos centroides (minimizando a distância entre os exemplos e o centroide mais próximo – a distância considerada neste caso foi a Euclideana) até estes estabilizarem ou terminar o número de iterações definido, criando assim *k* grupos ou *clusters* de dados.

Cálculo do valor de *k*:

Um dos problemas chave do *K-means* reside em como saber *a priori*, quantos *clusters* deverão ser criados ou por outras palavras, qual o valor de *k*. Normalmente este valor é estimado por um método simples mas bastante utilizado, que se apresenta na eq. (8):

$$K = \sqrt{\frac{N}{2}} \quad (8)$$


Assim, *k* toma o valor da raiz quadrada de *N* (o número de exemplos do *dataset*) a dividir por 2.

Aplicação do *K-means*:

O *K-means* foi implementado com recurso ao API do Weka que possibilita o seu ajuste consoante as necessidades. Usou-se *k* com valor fornecido pela estimativa acima referida e 1000 iterações. No entanto, como este módulo corre diretamente na plataforma, o utilizador

tem a possibilidade de modificar o valor de k de modo a ajustar as preferências do algoritmo. É sempre possível utilizar a estimativa, especificando o valor -1 na chamada ao módulo⁸.

Os resultados obtidos pelo *K-means* são guardados em ficheiro para mais tarde poderem ser utilizados na plataforma ADAPT. A figura 4.3 apresenta um *screenshot* do módulo em funcionamento.



```
Cluster: 0
  Utilizador: 30

Cluster: 1
  Utilizador: 17
  Utilizador: 18
  Utilizador: 19
  Utilizador: 20
  Utilizador: 21
  Utilizador: 24
  Utilizador: 26
  Utilizador: 27
  Utilizador: 28
```

Figura 4.3 - Resposta à pergunta 1.

De notar que os *clusters* têm o nome de *cluster k* porque na prática não podem ser associados a nenhum nome.

4.3 Média das notas por estilo de aprendizagem

A funcionalidade 2 pretende responder a qual a média das notas dos alunos por cada estilo de aprendizagem de OA consultado, por cada disciplina. Na prática consiste em criar uma estatística a partir de um *dataset* inicialmente gerado como se descreve de seguida.

4.3.1 Dataset

O *dataset* utilizado foi criado apenas com recurso a *queries* SQL que interligam as tabelas Utilizador (contém os dados relativos a cada utilizador), NavegacaoLog (contém os dados relativos a consultas a OAs), OA (contém os dados relativos aos conteúdos na base de dados. Em particular, utilizaram-se o seu id e parâmetros de estilo de aprendizagem), disciplina (contém os dados pertencentes às disciplinas existentes), TesteAvaliacao (contém os dados dos testes de avaliação. Esta tabela interage com as tabelas OA e Disciplina através da tabela DisciplinaPlano), UtilizadorTesteAvaliacao (contém a resposta e a pontuação de um dado aluno à pergunta de um teste de avaliação. Esta tabela interage com a tabela TesteAvaliacao através da tabela TesteAvaliacaoPergunta). A figura 4.4 apresenta as tabelas e respectivos campos utilizados na construção deste *dataset*.

⁸ O protótipo desenvolvido permite ao utilizador usar a estimativa ou não. Esta temática é abordada no Capítulo VI

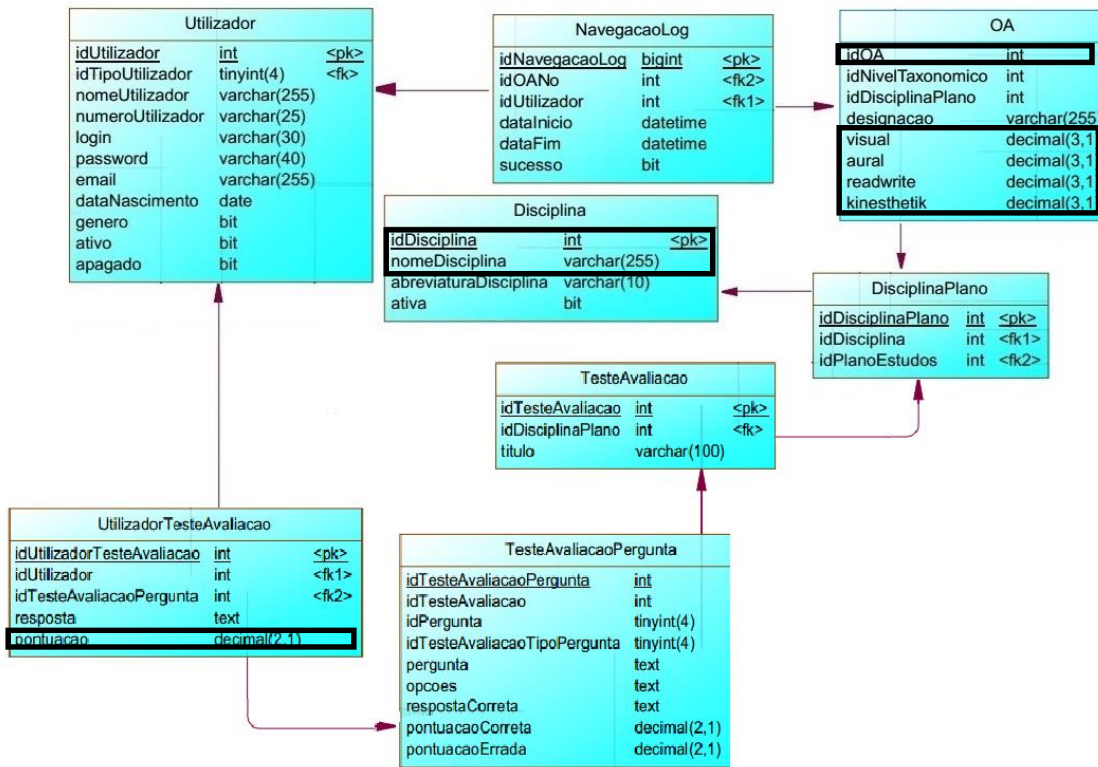


Figura 4.4 - Tabelas utilizadas para retirar os dados necessários à pergunta 2.

O dataset resultante é do tipo seguinte:

IdOA	IDDisciplina	DISCIPLINA	AURAL	VISUAL	READWRITE	KINESTHETIK	PONTUACAO
119	1	Sistemas Digitais	0,0	75,0	35,0	0,0	5,50000
169	1	Sistemas Digitais	0,0	50,0	90,0	50,0	5,42143
139	1	Sistemas Digitais	0,0	45,0	85,0	0,0	7,00000
100	1	Sistemas Digitais	0,0	80,0	10,0	10,0	4,00000
153	1	Sistemas Digitais	0,0	85,0	16,0	92,0	5,70000
...							

Figura 4.5 - Dataset correspondente à pergunta 2.

4.3.2 Algoritmo

O algoritmo utilizado tem como objetivo obter o estilo de aprendizagem predominante de cada OA (baseado na maior percentagem de cada estilo) e calcular a média das notas para cada um dos quatro estilos.

Foram também criados vários gráficos que apresentam a média das notas de cada estilo de aprendizagem para cada disciplina, fornecendo assim um *overview* mais imediato da distribuição dos dados pelas diversas disciplinas.

A figura 4.6 exemplifica um dos gráficos gerados a partir de dados experimentais:

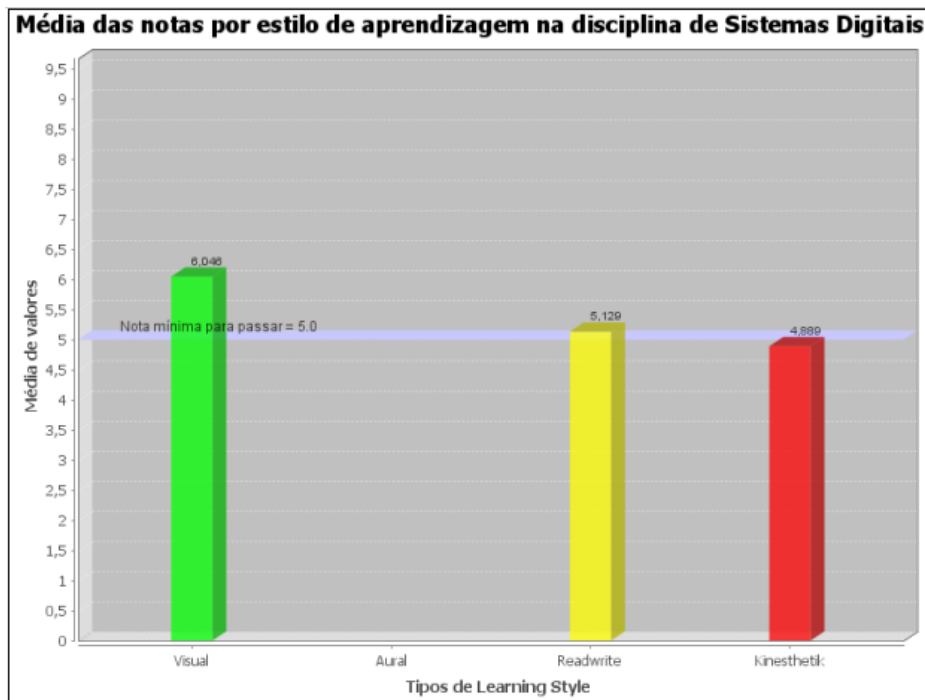


Figura 4.6 - Resultados da pergunta 2 para a disciplina de sistemas digitais.

As barras, representando a média das notas, vão variando a sua cor conforme o valor apresentado: dado que 5.0 é o valor mínimo para considerar uma nota positiva, a barra tomará a cor vermelha se a média for inferior a 5.0 (uma vez que o valor máximo que o campo “pontuação” na base de dados pode tomar é 10 pontos), amarela se for superior a 5 e inferior a 6 e verde se for superior a 6.

4.4 Extração de regras

Com esta funcionalidade pretendem extrair-se regras que associem idade, estilo de aprendizagem predominante e média das notas de cada aluno.

O algoritmo utilizado na geração de regras é o *Apriori*, um dos mais utilizados nesta área.

Considere-se o seguinte *dataset* que serve para efeitos explicativos do funcionamento do *Apriori*:

Tabela 4.1 - *Dataset* explicativo do algoritmo *Apriori*

id	Lista de Items
1	I1, I2, I5
2	I2, I4
3	I2, I3
4	I1, I2, I4
5	I1, I3
6	I2, I3
7	I1, I3
8	I1, I2, I3, I5
9	I1, I2, I3

O algoritmo *Apriori* consiste em Três passos distintos:

1. **Geração de candidatos** – Neste passo são gerados candidatos (com suporte mínimo) para k níveis através da análise de itemsets frequentes. O nível 1 (N_1) é gerado a partir da análise direta dos itemsets frequentes no *dataset*; o nível 2 (N_2) é gerado juntando os vários componentes de N_1 e analisando a frequência de items no *dataset* tendo em vista os candidatos gerados. A análise continua até $N_k = \emptyset$, isto é, não exista suporte mínimo para N_k .

E.g.: seja N_k o nível de geração de candidatos (i.e.: o número de items considerados por cada candidato) tem-se o seguinte conjunto de candidatos:

- a. Para $N_1 = \{\{I1\}, \{I2\}, \{I3\}, \{I4\}, \{I5\}\}$;
- b. Para $N_2 = \{\{I1,I2\}, \{I1,I3\}, \{I1,I5\}, \{I2,I3\}, \{I2,I4\}, \{I2,I5\}\}$;
- c. Para $N_3 = \{\{I1,I2,I3\}, \{I1,I2,I5\}\}$;

2. **Pruning** – Nos níveis superiores a 2 é utilizada uma propriedade do *Apriori* que define que os subsets de um itemset frequente são também frequentes ou seja, $\{I2, I3, I5\}$ subdivide-se nos candidatos $\{I2, I3\}$, $\{I2, I5\}$ que pertencem de facto a N_2 e $\{I3,I5\}$ que não é membro de N_2 e por isso é (foi) removido (pruned).

- Geração de regras** – Este passo compara o *dataset* com os candidatos escolhidos e gera N regras com uma confiança e suporte mínimos. Por exemplo, considerando uma confiança de 70%, para o *dataset* anterior, a regra R1: {I1,I2} -> I5 = 2/4 = 50% e por isso é rejeitada. A regra R2: {I1,I5} -> I2 = 2/2 = 100% por isso é selecionada.

4.4.1 Dataset

O *dataset*⁹ utilizado nesta funcionalidade foi criado a partir de *queries* SQL que relacionam os parâmetros do utilizador com as suas preferências VARK. Os atributos obtidos da base de dados pertencem às tabelas Vark, UtilizadorTesteAvaliacao, e Utilizador explicadas anteriormente e apresentam-se na figura 4.7:

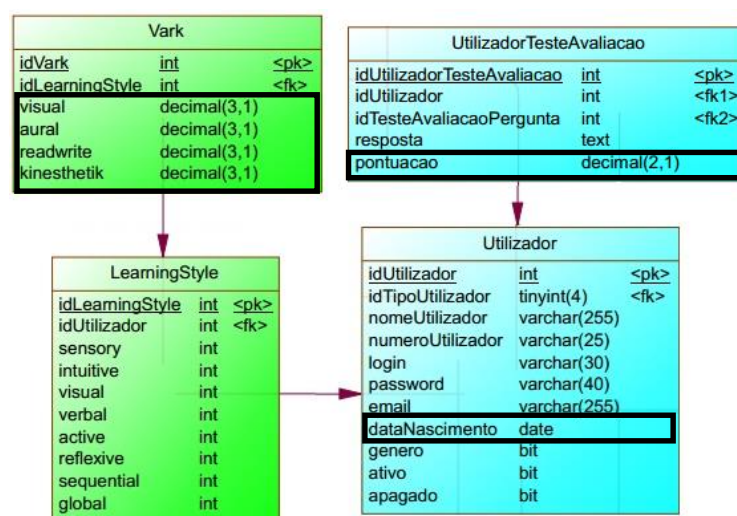


Figura 4.7 - Tabelas utilizadas para retirar os dados necessários para a pergunta 3.

Os resultados da *query* são processados de modo a que se consiga obter o estilo de aprendizagem predominante do aluno. Para aplicação do *Apriori* os dados numéricos têm de ser discretizados. O *dataset* criado (ficheiro CSV) consiste no seguinte¹⁰:

IDADE,LEARNING STYLE,MEDIA_NOTAS
36-40,kinesthetik,bom
36-40,kinesthetik,insatisfaz
17-21,kinesthetik,satisfaz
17-21,visual,insatisfaz
22-26,kinesthetik,satisfaz

Figura 4.8 - Dataset da pergunta 3.

⁹ O *dataset* utilizado encontra-se especificado no anexo XIII.

¹⁰ Uma vez que as notas dos alunos na BD apenas vão de 0 a 10 (aparente limitação imposta), as notas discretizadas dos alunos correspondem a: muito fraco ($x \leq 2$); fraco ($2 < x \leq 3$); insatisfaz ($3 < x \leq 4.9$); satisfaz ($4.9 < x \leq 7$); bom ($7 < x \leq 8.5$); muito bom ($8.5 < x \leq 10$).

Cada linha do *dataset* descreve um utilizador. Como se pode verificar, o primeiro aluno tem uma idade situada no intervalo dos 36 aos 40 anos, o seu estilo de aprendizagem predominante é *kinesthetic* e a sua média de notas é “bom”. Aplicação do algoritmo de geração de regras.

O algoritmo utilizado para a geração de regras – *Apriori* – foi implementado recorrendo à API do weka e gera regras a partir de um *dataset*. A ideia será extrair regras tais como, por exemplo: se idade 36-40 e estilo de aprendizagem *kinesthetic* => média de notas “bom”. O algoritmo assume que existe correlação entre os atributos e cria por defeito $\sqrt{\frac{N}{2}}$ regras (valor por defeito – “-1”) embora este número possa ser modificado pelo utilizador. A confiança mínima das regras, por omissão, é de 0.8 embora possa também ser alterada pelo utilizador.

Os resultados do algoritmo podem ser visualizados na plataforma ADAPT e consistem no seguinte:

```

Apriori
=====

Minimum support: 0.25 (5 instances)
Minimum metric <confidence>: 0.8
Significance level: 1
Number of cycles performed: 15

Generated sets of large itemsets:

Size of set of large itemsets L(1): 7

Large Itemsets L(1):
IDADE=17-21 8
IDADE=22-26 10
LEARNING STYLE=kinesthetik 6
LEARNING STYLE=visual 10
LEARNING STYLE=readwrite 5
MEDIA_NOTAS=insatisfaz 6
MEDIA_NOTAS=satisfaz 12

Size of set of large itemsets L(2): 5

Large Itemsets L(2):
IDADE=17-21 LEARNING STYLE=visual 5
IDADE=17-21 MEDIA_NOTAS=insatisfaz 5
IDADE=22-26 LEARNING STYLE=visual 5
IDADE=22-26 MEDIA_NOTAS=satisfaz 10
LEARNING STYLE=visual MEDIA_NOTAS=satisfaz 6

Size of set of large itemsets L(3): 1

Large Itemsets L(3):
IDADE=22-26 LEARNING STYLE=visual MEDIA_NOTAS=satisfaz 5

Best rules found:

1. IDADE=22-26 10 ==> MEDIA_NOTAS=satisfaz 10 conf:(1) lift:(1.75) lev:(0.2) [4] conv:(4.29)
2. IDADE=22-26 LEARNING STYLE=visual 5 ==> MEDIA_NOTAS=satisfaz 5 conf:(1) lift:(1.75) lev:(0.1) [2] conv:(2.14)
3. MEDIA NOTAS=satisfaz 12 ==> IDADE=22-26 10 conf:(0.83) lift:(1.75) lev:(0.2) [4] conv:(2.1)

```

Conf: Medida de confiança
Lift: Medida de lift
Lev: medida de leverage

Figura 4.9 – Resposta à pergunta 3.

4.5 Percentagem de visualização de OAs por estilo de aprendizagem

Esta funcionalidade pretende apurar qual a percentagem de visualização de OAs de cada estilo de aprendizagem e qual o estilo de aprendizagem associado a melhores notas.

4.5.1 Dataset

O *dataset* desta funcionalidade foi criado a partir das tabelas NavegacaoLog, OA, Disciplina e UtilizadorTesteAvaliacao, utilizando os campos id e parâmetros Vark do OA, início e fim de uma consulta ao OA, disciplina a que este pertence e pontuação associada ao utilizador que o consultou:

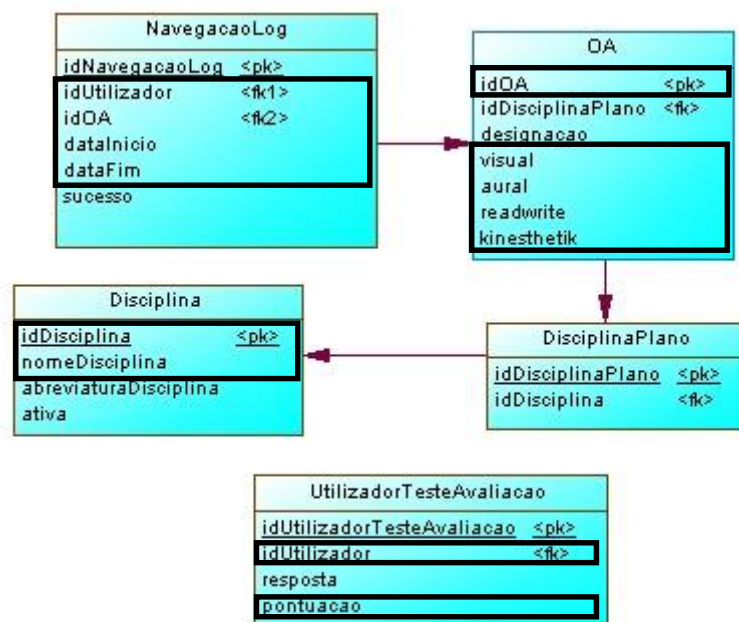


Figura 4.10 - Tabelas utilizadas para retirar os dados necessários para a pergunta 4.

Depois de obter estes dados o *dataset* criado tem o seguinte formato:

```

IDOA,idDisciplina,disciplina ,tempo, pontuacao
108,1,Sistemas Digitais,readwrite,14400,8.0
144,1,Sistemas Digitais,kinesthetic,33845,4.0
124,1,Sistemas Digitais,visual,6,5.0
160,1,Sistemas Digitais,visual,29282,9.0
...
  
```

Figura 4.11 - Dataset da pergunta 4.

4.5.2 Algoritmo

Após a criação do *dataset* o algoritmo contabiliza as ocorrências de cada estilo de aprendizagem para calcular a porcentagem de consulta de OAs por estilo. A média da pontuação de cada estilo (dado pela nota dos alunos que consultaram o artigo por mais de dois minutos) também foi tida em conta para que no resultado figure qual a porcentagem de consulta de OAs de cada estilo de aprendizagem e qual o estilo com OAs que regra geral levam a melhores notas. O algoritmo subdivide os resultados por disciplina, criando gráficos com a informação disponível para cada disciplina e também de uma forma geral.

Um exemplo de resultados obtidos por este algoritmo figura na imagem seguinte:



Figura 4.12 - Resultado da pergunta 4. Disciplina de sistemas digitais.

4.6 Percentagem de mudanças de estilo de aprendizagem

A pergunta 5 utiliza o classificador de estilos de aprendizagem para descobrir quais as mudanças de estilo de aprendizagem mais frequentes.

O *dataset* é fornecido pelo módulo de classificação de estilos de aprendizagem e consiste na informação das mudanças de estilo de aprendizagem ao longo de X semanas:

```
ID: 19 LEARNING STYLE: visual CHANGE: no CHANGE FOR WHAT: null VISUAL: 42.98386658699549 % AURAL: 0.0 %
READWRITE: 42.94015608307531 % KINESTHETIK: 14.07617650786753 %

ID: 20 LEARNING STYLE: kinesthetik CHANGE: yes CHANGE FOR WHAT: readwrite VISUAL: 28.568379360371814 %
AURAL: 11.278742826329777 % READWRITE: 60.1529099102569 % KINESTHETIK: 0.0 %

ID: 21 LEARNING STYLE: kinesthetik CHANGE: yes CHANGE FOR WHAT: visual VISUAL: 83.54431962514724 %
AURAL: 16.457325620422107 % READWRITE: 0.0 % KINESTHETIK: 0.0 %

ID: 22 LEARNING STYLE: visual CHANGE: no CHANGE FOR WHAT: null VISUAL: 62.419356027114645 % AURAL: 0.0
% READWRITE: 0.0 % KINESTHETIK: 37.581417624521066 %
```

Figura 4.13 - Resultado do módulo de classificador de estilos de aprendizagem.

4.6.1 Algoritmo

O algoritmo deste módulo consiste em contabilizar qual o número de mudanças de estilo de aprendizagem predominante para outro.

Os resultados são apresentados num gráfico que mostra a percentagem de mudanças de um estilo para outro:

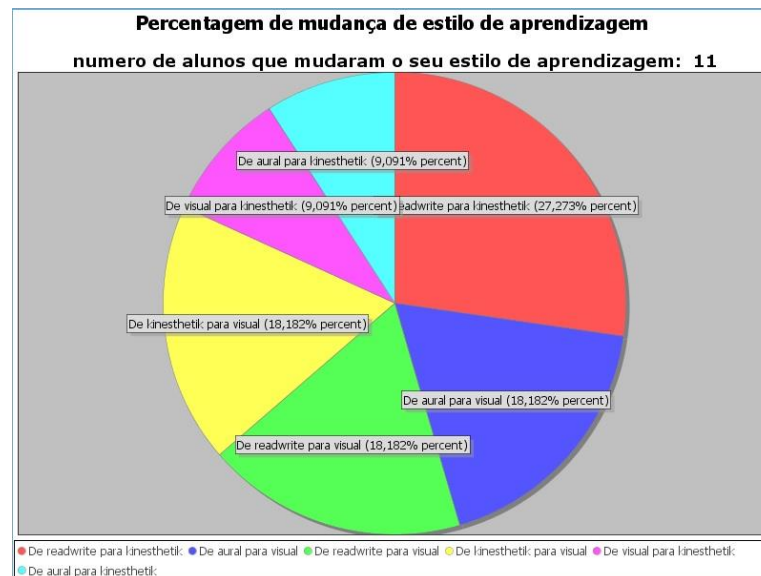


Figura 4.14 - Resposta à pergunta 5.

4.7 Média das notas e predominância de estilos por disciplina e curso

Esta funcionalidade permite determinar qual a pontuação média associada aos OAs e qual o estilo de aprendizagem predominante por disciplina. Esta funcionalidade permite também saber se cada disciplina é frequentada por alunos com boas notas ou não e qual o estilo de aprendizagem mais consultado.

4.7.1 Dataset

O *dataset* gerado para esta funcionalidade foi retirado da base de dados através de uma *query* que relaciona os OAs (os seus parâmetros VARK e pontuação associada), com a disciplina (id, nome e curso a que pertence) à qual pertencem. Das tabelas utilizadas as mais importantes são: OA, Disciplina, utilizadorTesteAvaliacao e curso, sendo que as restantes fazem apenas a interligação entre estas.

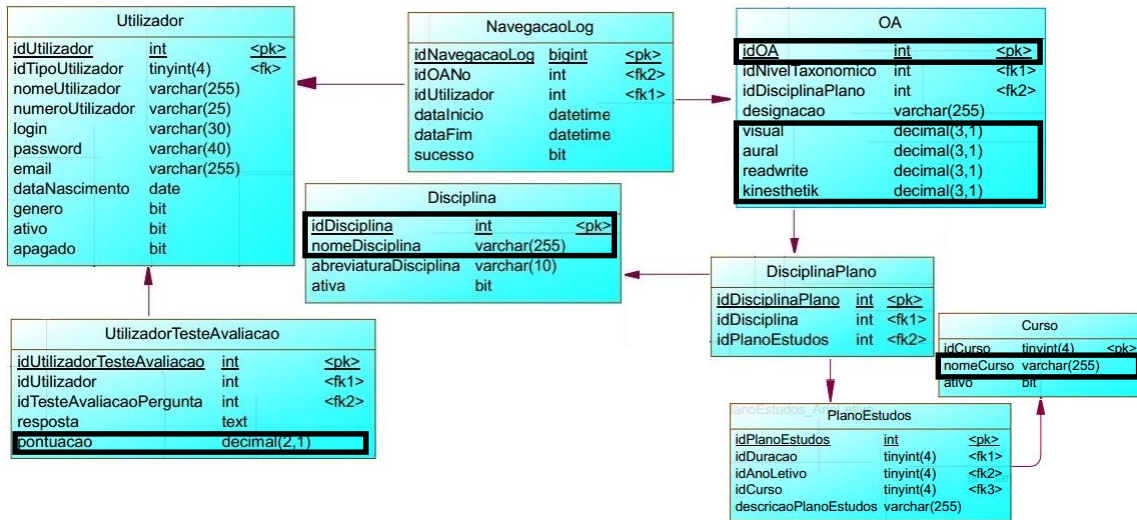


Figura 4.15 - Tabelas utilizadas para retirar os dados necessários para a pergunta 6.

Por fim, o *dataset* criado (em ficheiro csv) tem o seguinte formato:

SubjectID	SubjectName	CURSO	idOA	AURAL	VISUAL	READWRITE	KINESTHETIK	OASCORE
1	Sistemas Digitais	Engenharia Informática	133	0.0	60.0	65.0	0.0	4.50000
1	Sistemas Digitais	Engenharia Informática	148	0.0	85.0	10.0	90.0	6.66667
1	Sistemas Digitais	Engenharia Informática	168	0.0	75.0	0.0	10.0	5.91333
1	Sistemas Digitais	Engenharia Informática	116	0.0	90.0	10.0	70.0	4.33333
1	Sistemas Digitais	Engenharia Informática	134	0.0	65.0	50.0	60.0	5.00000

Figura 4.16 - *Dataset* da pergunta 6.

Cada linha do *dataset* corresponde a um OA, contendo também informação sobre os parâmetros VARK do OA, a sua pontuação associada, a disciplina da qual faz parte e o curso a que essa disciplina pertence. E.g.: a primeira linha diz respeito ao OA de id 133 com uma pontuação média de 4.5 valores que faz parte da disciplina de “Sistemas Digitais”, do curso de “Engenharia Informática” e tem como parâmetros VARK 0% *visual*, 60% *aural*, 65% *readwrite* e 0% *kinesthetic*¹¹.

4.7.2 Algoritmo

O algoritmo utilizado nesta funcionalidade apenas verifica qual o valor VARK predominante de cada OA (o seu EA) e calcula a média da pontuação dos OAs de cada disciplina. O algoritmo verifica também qual o estilo de aprendizagem que tem mais acessos fornecendo informação sobre o EA que mais contribuíram para a media da disciplina.

Note-se que se está a avaliar uma disciplina com base na nota associada aos seus OAs (dado por todos os utilizadores que o consultaram, independentemente se a sua nota se deve ou

¹¹ Cada preferência VARK pode tomar qualquer valor de 0 a 100%.

não à disciplina em questão). Neste caso faz sentido verificar a nota associada a um OA de modo global. Na disciplina em questão o OA pode não ter boa pontuação porque não foi consultado por ninguém, mas ser muito bom a nível geral porque foi consultado por alunos que estão inscritos noutras disciplinas – às quais tiveram boa nota – mas que ainda não foram avaliados à disciplina do OA em questão, o que faz do OA mau do ponto de vista da pontuação dos alunos associados à sua disciplina (uma vez que os alunos ainda não foram avaliados) mas bom, tendo em conta os alunos de uma forma global (uma vez que podem já ter tido boas notas a outras disciplinas).

Os resultados são apresentados no formato gráfico de barras. A figura 4.17 exemplifica este gráfico:

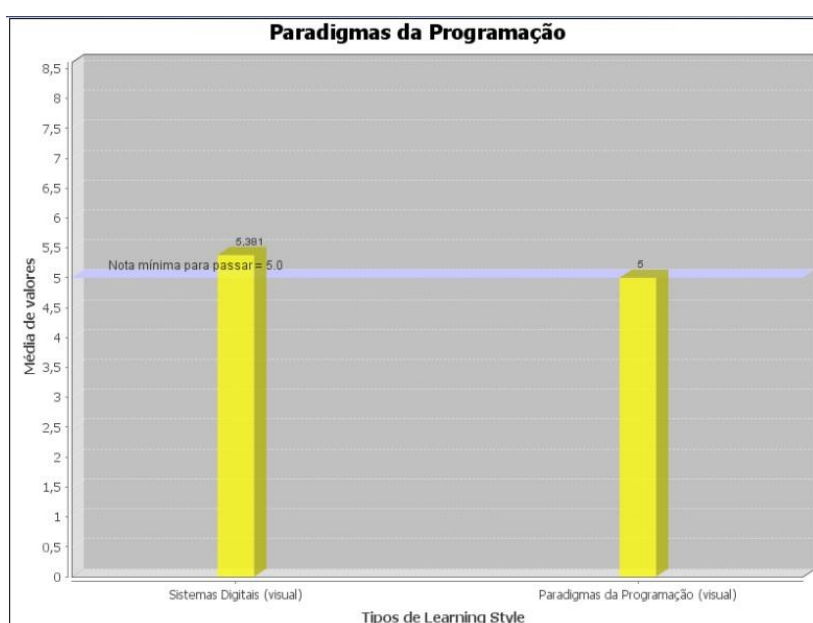


Figura 4.17 - Resposta à pergunta 6.

No caso representado, a disciplina de Sistemas Digitais tem uma média de 5,295 valores, o estilo de aprendizagem dos OAs mais consultados foi maioritariamente *visual*. Já a disciplina de Paradigmas da Programação apresenta uma média de 5.0 valores.

Uma vez mais, cada barra (que representa uma disciplina) pode tomar a cor vermelha, amarela ou verde consoante se encontra entres os valores 0-4.9; 5-5.9; 6-10.

4.8 Distribuição dos alunos por estilo de aprendizagem

Esta funcionalidade permite verificar qual a distribuição dos alunos por estilo de aprendizagem.

4.8.1 Dataset

O *dataset* utilizado para endereçar esta funcionalidade é uma vez mais baseado numa *query* SQL que fornece algumas informações sobre os utilizadores, nomeadamente o seu id e preferências *vark*. A figura 4.18 apresenta as tabelas utilizadas na construção deste *dataset*: As tabelas *Vark*, *Utilizador* e *LearningStyle*.

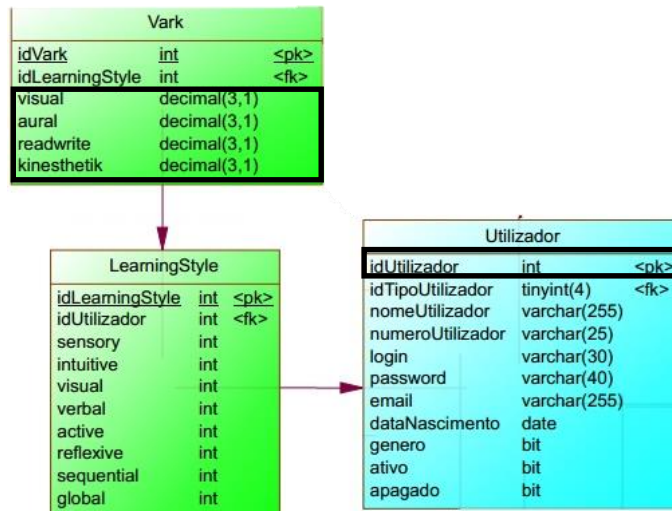


Figura 4.18 - Tabelas utilizadas para retirar os dados necessários para a pergunta 7.

O *dataset* criado através de SQL (em ficheiro csv) tem o seguinte aspeto:

```
idUSER,AURAL,VISUAL,READWRITE,KINESTHETIK
11,20.6,90.7,21.1,91.0
12,23.7,79.4,37.5,0.0
13,39.5,24.9,37.5,74.9
14,8.3,80.9,31.7,75.0
17,31.7,75.5,23.7,75.5
18,8.3,91.6,8.3,91.6
```

Figura 4.19 - *Dataset* da pergunta 7.

4.8.2 Algoritmo

O algoritmo correspondente a esta funcionalidade consiste na identificação do estilo de aprendizagem predominante de cada aluno e na contabilização dos alunos segundo cada estilo de aprendizagem, o que permite calcular a percentagem de alunos de cada estilo.

Os resultados (gráfico) são apresentados como mostra a figura 4.20:



Figura 4.20 - Resposta à pergunta 7.

Como se pode verificar, existem cerca de 40% de alunos *visual*, 24% de alunos *kinesthetic*, 20% de alunos *readwrite* e 16% de alunos *Aural*. O estilo *visual* é claramente predominante.

4.9 Notas (discretizadas) por OA, por disciplina

Esta funcionalidade apresenta as notas por OA discretizadas segundo cada disciplina i.e.: saber, por disciplina, o resultado qualitativo (muito fraco, fraco, insatisfaz, satisfaz, bom, muito bom) através das consultas efetuadas aos seus OAs, fornecendo assim uma visão sobre o número de OAs associados a cada nota.

4.9.1 *Dataset*

O *dataset*, tal como em todas as outras funcionalidades, foi criado através de *queries* SQL. As tabelas utilizadas foram: Utilizador (id do utilizador), Disciplina (id e nome da disciplina) e UtilizadorTesteAvaliacao (pontuação de um aluno numa disciplina). As restantes tabelas servem apenas para ligação de modo que se possa obter a pontuação por cada teste de cada disciplina obtida por cada aluno. A figura 4.21 mostra os campos e tabelas utilizadas.

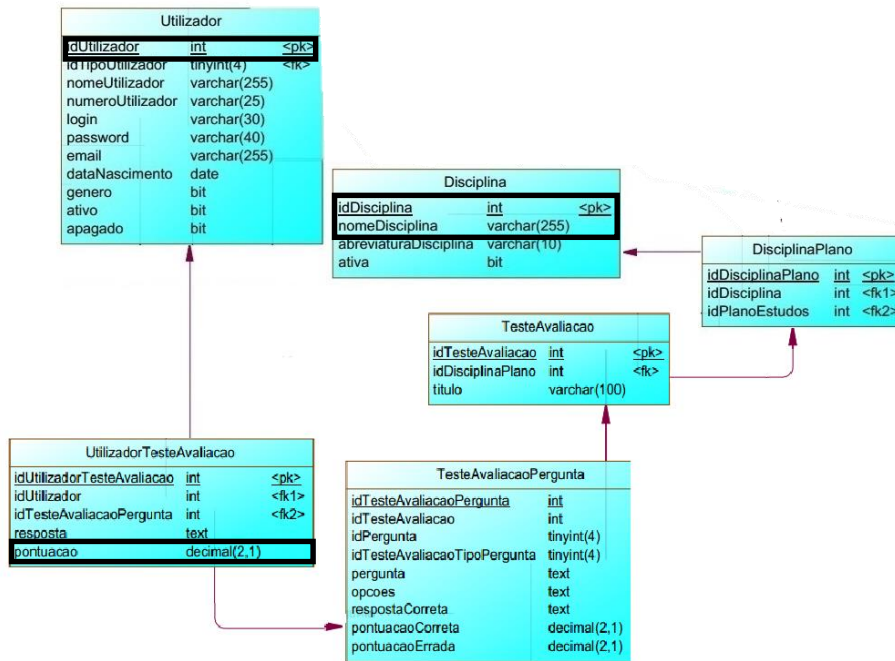


Figura 4.21 - Tabelas utilizadas para retirar os dados necessários para a pergunta 8.

Após a discretização dos dados fornecidos pela *query* SQL, o *dataset* obtido é o seguinte:

IdLO	idSubject	Subject	User score
119	1	Sistemas Digitais	satisfaz
144	1	Sistemas Digitais	insatisfaz
169	1	Sistemas Digitais	satisfaz
120	1	Sistemas Digitais	satisfaz
145	1	Sistemas Digitais	satisfaz
170	1	Sistemas Digitais	satisfaz
122	1	Sistemas Digitais	satisfaz
...			

Figura 4.22 - *Dataset* da pergunta 8.

Cada linha do *dataset* representa um OA. A primeira linha deste exemplo apresenta o OA de id 119, que pertence á disciplina de id 1 (Sistemas Digitais) e tem uma média de acessos de “satisfaz” (média dos alunos que o consultaram).

4.9.2 Algoritmo

O algoritmo consiste em contagens de notas por disciplina. São contabilizados os números de:

- Muito fraco ($x \leq 2$);
- Fraco ($2 < x \leq 3$);
- Insatisfaz ($3 < x \leq 4.9$);
- Satisfaz ($4.9 < x \leq 7$);
- Bom ($7 < x \leq 8.5$);

- Muito bom ($8.5 < x \leq 10$);

Os resultados consistem em gráficos de barras que representam o número de notas Muito fraco, Fraco, Insatisfaz, Satisfaz, Bom e Muito Bom, associadas aos OAs de cada disciplina. A figura 4.23 mostra um dos gráficos gerados para as disciplinas existentes:



Figura 4.23 - Resultados da pergunta 8: disciplina de sistemas digitais.

4.10 Predição de notas

A finalidade deste módulo consiste em analisar os dados existentes sobre os alunos já avaliados e prever a média dos alunos que ainda não foram avaliados com base nas suas notas, artigos consultados e estilo de aprendizagem. Esta funcionalidade foi implementada através uma rede neuronal treinada com os dados dos alunos avaliados tendo como label a predição de nota positiva ou negativa dos alunos.

4.10.1 Dataset

O *dataset* criado para o efeito de treino¹² da rede é composto pelos atributos estilo de aprendizagem predominante, um array de objetos booleanos que representam a consulta/não consulta de um determinado OA na plataforma, o número de OAs consultados e por fim a sua nota (label): positiva ou negativa. O *dataset* sobre o qual a rede será utilizada é composto pelos mesmos campos exceptuando a label final. A tabela seguinte apresenta o *dataset* de treino.

¹² O *dataset* de treino da rede encontra-se no especificado no anexo XIV

Tabela 4.2 - Dataset da pergunta 9.

idUser	EstiloAprendizagem	OA1	OA2	...	OAn	NumOasConsultados	Nota Positiva
1	Visual	true	false		true	10	false
2	Kinesthetic	false	false		true	16	true
...
N	Readwrite	true	true		true	13	false

A informação descrita foi extraída das tabelas Vark (EA do aluno), Utilizador (id e nome do utilizador), Disciplina (id e nome da disciplina), OA (id) e UtilizadorTesteAvaliacao (pontuação associada a um utilizador). As tabelas LearningStyle, NavegacaoLog, DisciplinaPlano, UtilizadorDisciplinaPlano, TesteAvaliacao, TesteAvaliacaoPergunta foram utilizadas apenas na interligação das anteriores. A figura 4.24 Mostra os campos e tabelas utilizados:

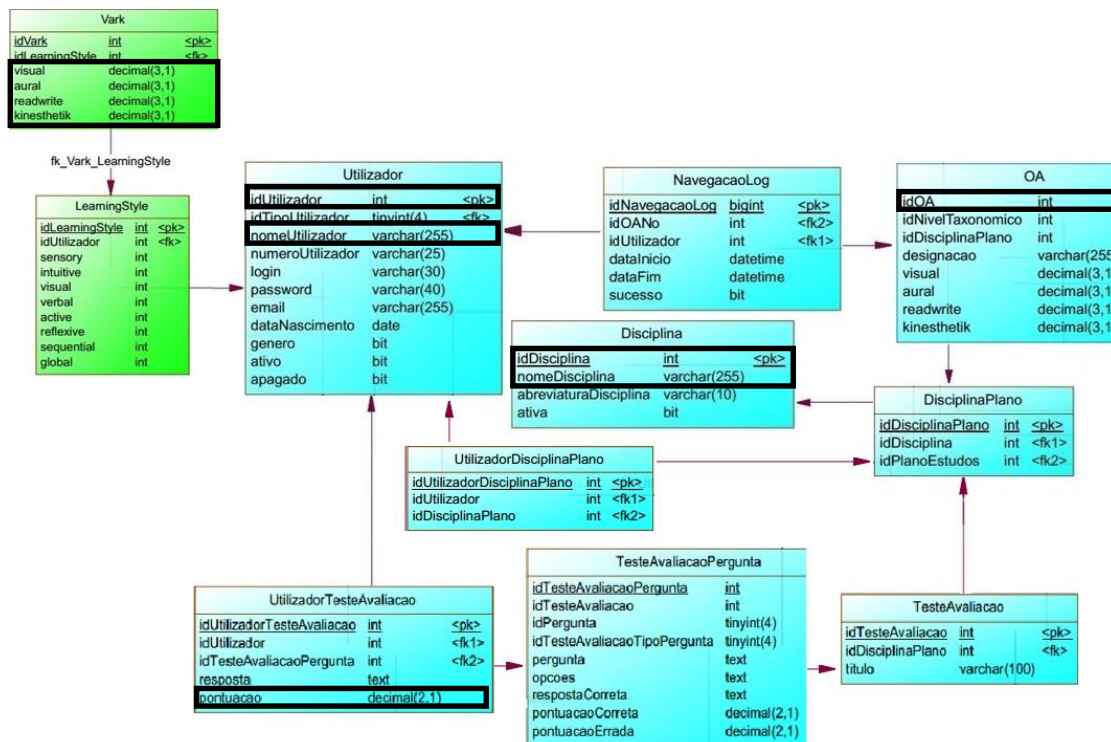


Figura 4.24 - Tabelas utilizadas para retirar os dados necessários para a pergunta 9.

Existe contudo um problema: o número de atributos do *dataset* de uma rede neuronal tem de ser igual tanto para o exemplo de treino (que contém os alunos já avaliados) como para o *dataset* sobre o qual a rede vai atuar (contém os alunos que ainda não foram avaliados). Uma vez que não se pode garantir que todos os utilizadores consultaram o mesmo número de OAs, têm de ser contabilizados todos os OAs da base de dados como atributos booleanos. Estes atributos representam o facto de um utilizador ter ou não consultado um dado OA. Ao utilizar-

se todos os OAs como atributo e uma vez que o número de OAs poderá ser modificado, está a criar-se um problema NP não escalável, i.e.: a rede neuronal que recebe este *dataset* de treino não será treinada em tempo útil o que constitui um problema porque terá de ser treinada periodicamente para ter em conta as mudanças no número de OAs existente.

Dado este elevado número de atributos, a solução passa por remover alguns. Para tal surgiram duas soluções:

- Redução, estabelecendo um número de visualizações mínimo;
- *Principal Component Analysis* (PCA) (Jolliffe, 2002);

A redução por número de visualizações consiste em eliminar um OA que tenha menos que X visualizações, em que X é estimado recorrendo à estimativa $X = \sqrt{\frac{N}{2}}$ em que N é o número de utilizadores do *dataset*.

A ideia principal do PCA é reduzir a dimensão de um *dataset* com um elevado número de variáveis interrelacionadas, mantendo o maior número possível de variação presente no *dataset* original. Isto é conseguido transformando as variáveis do *dataset* num novo *set* de variáveis (*principal components*) que não estão correlacionadas e que estão ordenadas de modo a que as primeiras retenham a maior variação presente no *dataset* original. Por outras palavras, esta técnica sumariza um *dataset* que contém um elevado número de variáveis (p) num conjunto menor de (k) variáveis compostas e derivadas a partir do conjunto original. A figura 4.25 ilustra este princípio:

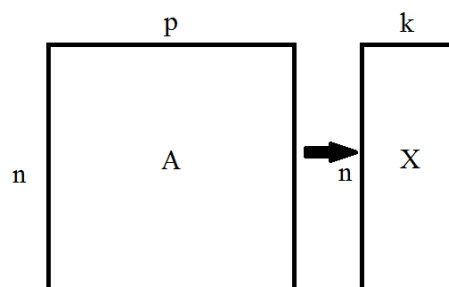


Figura 4.25 - *Principal Component Analysis*.

Uma vez definidos os algoritmos de eliminação de atributos, pode proceder-se à sua eliminação. Contudo, existe ainda um aspeto a ter em conta: como o exemplo de treino e o *dataset* no qual a rede será aplicada têm de ter igual número de atributos, o algoritmo de redução de dimensionalidade é aplicado a ambos os conjuntos de dados. Os dois *datasets* são depois guardados no formato “.arff” e servirão de *input* à rede neuronal.

4.10.2 Algoritmo

Criados os *datasets*, pode proceder-se à criação da rede neuronal. Notar que, uma vez que o número de OAs está constantemente a ser alterado, de cada vez que este módulo é invocado, a rede neuronal é de novo treinada e depois é aplicada sobre o *dataset* que contém os

alunos por avaliar. Como o processo é moroso, este módulo não pode correr diretamente na plataforma.

A rede neuronal, implementada com recurso ao operador MultilayerPerceptron (MultilayerPerceptron, 2014) do WEKA, é *feedforward* (Haykin, 1998) (o que significa que o primeiro passo é mapear as entradas da rede nas respetivas saídas) e utiliza *backpropagation* (Rojas, 1996) minimizando os erros através de *gradient descent* (algoritmo que procura minimizar o erro entre exemplo de treino e de teste fornecidos). A rede utiliza uma função de activação sigmoide (função baseada na forma “S”) que mapeia as entradas nas respetivas saídas através de *thresholds* determinados através dos pesos de cada *node*.

A rede é treinada com 100 iterações, através das quais o algoritmo *backpropagation* minimiza o erro dos pesos associados aos 9 *nodes* utilizados numa (única) camada intermédia. Utiliza um coeficiente de aprendizagem (*learning rate*) de 0.3 e um *momentum* de 0.2. Foram testadas várias combinações de opções e chegou-se à conclusão que estas eram as melhores para cada um dos parâmetros.

A precisão da rede é de 100% para *split* de 90% (separação do *dataset* 90% para treino e 10% para teste) utilizando um *dataset* composto por 20 utilizadores pertencentes a duas disciplinas (31 linhas) e de 77% para *split* de 66%. O resultado da rede neuronal em produção apresenta-se na figura 4.26:

O aluno Jorge Matos tem maior probabilidade de passar a Sistemas Digitais
O aluno José Dinis tem maior probabilidade de reprovar a Sistemas Digitais
O aluno Carla Santos tem maior probabilidade de reprovar a Sistemas Digitais

Figura 4.26 - Resposta à pergunta 9.

4.11 Identificação de grupos de utilizadores

Esta pergunta pretende verificar a existência de grupos de utilizadores num determinado período de tempo com base nos seguintes atributos:

- Id do utilizador;
- Parâmetros VARK utilizador;
- Média e desvio padrão dos parâmetros VARK dos OAs que consultou;
- Número de OAs consultados;
- Pontuação média (pontuação média que o aluno teve a cada disciplina);

A ideia base desta funcionalidade é verificar se existiu algum utilizador que mudasse de grupo em períodos temporais distintos.

4.11.1 *Dataset*

O *dataset* necessário foi criado recorrendo à informação das tabelas Vark (estilo de aprendizagem dos alunos), Utilizador (id que permitirá saber qual o utilizador), NavegacaoLog

(tempo de consulta de um OA), OA (id e EA do conteúdo) e UtilizadorTesteAvaliacao (média de um utilizador). A figura 4.27 ilustra os campos e tabelas da BD que foram utilizados.

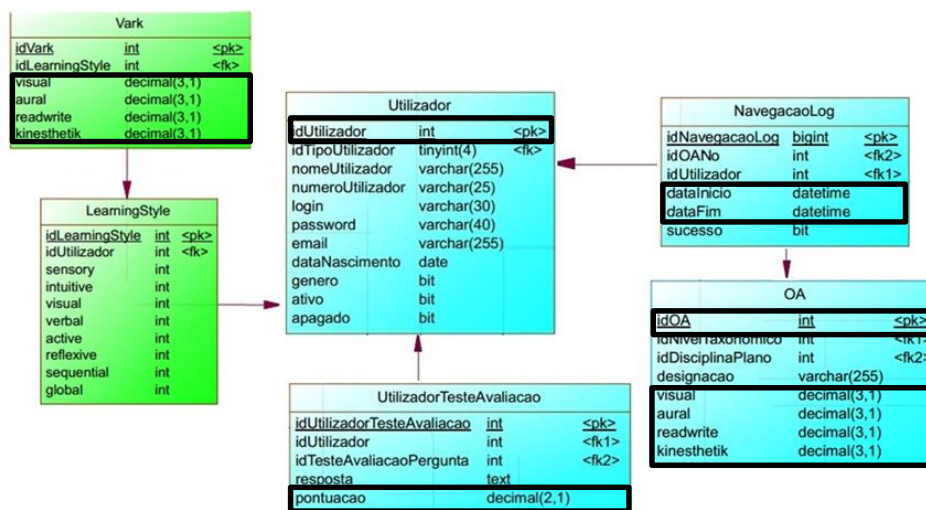


Figura 4.27 - Tabelas utilizadas para retirar os dados necessários para a pergunta 10.

A informação a retirar da base de dados está compreendida num período temporal definido por uma data inicial e uma data final. A figura 4.28 mostra o *dataset* obtido:

USER,USERSVISUAL,USERAURAL,USERREADWRITE,USERKINESTHETIK,OASMEDIAVISUAL,OASDESIVIOVISUAL,OASMEDIAAURAL,OASDESIVIOAURAL,OASMEDIAREADWRITE,OASDESVIOR EADWRITE,OASMEDIKINESTHETIK,OASDESVIOKINESTHETIK,USERIDSCONSULTADOS,USER PONTUACAO
18,91.6,8.3.8.3.91.6,82.50000,7.50000,0.00000,0.00000,42.50000,17.50000,0.00000,0.00000 ,2,6.95000
19,60.0,2.0.29.0,30.0,67.50000,30.65262,15.00000,33.54102,19.33333,20.26217,32.83333,3 9.62498,6,4.50000
20,4.0,18.0,57.0,82.0,55.00000,17.11307,0.00000,0.00000,52.85714,31.37983,10.71429,16.5 6773,7,7.00000
21,50.0,51.0,21.0,90.0,57.00000,31.79623,18.99000,38.04446,17.00000,20.63977,25.50000, 27.33587,10,7.00000

Figura 4.28 - Dataset da pergunta 10.

4.11.2 Algoritmo

Após criado o *dataset* o algoritmo escolhido para efetuar o *clustering* dos dados foi o X-*means* (Pelleg e Moore, 2000). Este algoritmo permite definir um intervalo para o número de *clusters* que terá em conta. O seu funcionamento é semelhante ao *K-means* mas com um passo intermédio que, de um modo geral, determina se deve ou não gerar mais algum centróide, isto é, mais um *cluster*. O algoritmo considera inicialmente o limite inferior (o número mínimo de *clusters* a utilizar) e vai adicionando centróides à medida das necessidades.

Essencialmente, o algoritmo é composto por dois passos que atuam até existir convergência (distribuição dos objetos pelos *clusters* existentes estar estabilizada) ou k maior que o limite máximo estipulado. Os passos consistem no seguinte:

- **Melhoramento de resultados** – Este passo consiste em correr o *K-means* de modo a chegar a uma convergência;
- **Melhoramento da estrutura** – Este passo determina se e onde os novos centroides devem aparecer a cada iteração do *K-means* (passo anterior). A estratégia adotada passa por dividir os centroides em dois: "The Improve-structure operation finds out if and where new centroids should appear. This is achieved by letting some centroids splitting in two" (Pelleg e Moore, 2000). Os centroides gerados (filhos) são movidos em direções opostas uma distância proporcional ao tamanho da região do seu "progenitor". De seguida, para cada região controlada pelos pais, é corrido o *K-means* com $K=2$ (2 filhos). Por fim, é verificado se de facto houve melhorias na estrutura ou não. Se não houve, os *clusters* que não fornecerem melhorias à estrutura são eliminados.

Este algoritmo foi utilizado com um intervalo de estimação para k de $[2,200]$ (mais que 200 grupos de utilizadores considera-se absurdo, pois implica a existência uma amostra de alunos elevadíssima).

A imagem seguinte diz respeito à identificação de grupos no período situado entre 23/10/2013 e 02/01/2014:

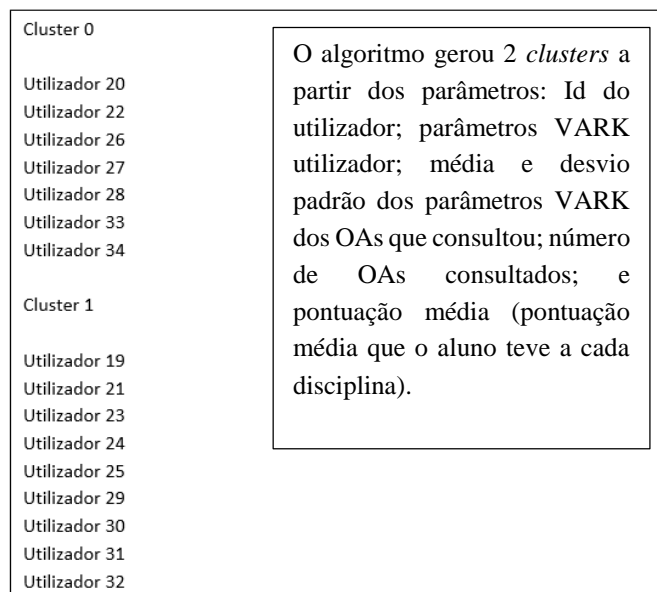


Figura 4.29 - *Clusters* formados na pergunta 10.

Neste caso o algoritmo apenas identificou dois grupos de utilizadores, distribuídos pelo *cluster* 0 e 1. Notar que a ideia base desta funcionalidade pressupõe a sua utilização em períodos diferentes, de modo a que o utilizador possa verificar se ao longo do tempo existiram mudanças de alunos entre grupos.

4.12 Métodos adicionais

Uma vez que ao utilizar técnicas de *data mining* apenas se utilizam os ids dos alunos e dos OAs, optou-se por se criarem dois métodos adicionais que fazem uma *query* à base de dados pelo id do OA ou do utilizador pretendido, mostrando ao utilizador da plataforma o nome correspondente ao id pesquisado.

4.13 Organização de ficheiros

Para facilitar a integração com o projeto ADAPT, optou-se por se criar uma estrutura de diretorias dentro de uma outra (diretoria mãe) denominada “*Results*”. Dentro da pasta “*Results*” existem duas diretorias:

- *QueryResults*: contém as diretorias relativas a cada *query* que contém os *datasets* a ser processados.
- *Outputs*: contém também as diretorias relativas a cada *query* que contém os *datasets* a ser processados. Cada diretoria dentro da diretoria *Outputs* contém os ficheiros numerados no intervalo [0,N] que correspondem aos resultados de cada um dos algoritmos descritos nas sub-secções anteriores.

4.14 Interação com o projeto ADAPT

A interação deste módulo com a plataforma ADAPT é feita de dois modos: 1) diretamente, porque as questões 1,3,5 e 10 requerem interação por parte dos utilizadores (i.e.: especificação de dados ou atributos) e por isso é necessário que corram embutidas na plataforma; 2) indiretamente porque as restantes perguntas não requerem interação com os utilizadores. A resposta a estas perguntas é criada periodicamente por uma aplicação que deposita os dados gerados numa diretoria acessível à plataforma de modo que esta os consiga obter, utilizar ou mostrar ao utilizador.

5 Conclusões

Este módulo possibilita aos tutores da plataforma identificar comportamentos, lacunas ou falhas existentes a nível dos seus conteúdos. O seu foco é dar a conhecer potencial informação não evidente mas potencialmente útil para os professores, uma vez que poderão aperfeiçoar ou criar novos conteúdos, dialogar com os alunos, possibilitando melhorar a experiência do utilizador da plataforma e conseqüentemente o seu percurso académico.

CAPÍTULO V – SISTEMA DE RECOMENDAÇÕES

Este capítulo explica o desenvolvimento do módulo de sistema de recomendações, organizando-se da seguinte forma: introdução; estado de arte; desenvolvimento do módulo; interação com o utilizador e plataforma; performance e conclusões.

1 Introdução

Num sistema de *e-Learning*, a sugestão/recomendação de novos artigos ao utilizador considera-se bastante importante dado que pode levá-lo a consultar conteúdos para além daqueles que uma simples pesquisa pode sugerir.

O módulo do motor de busca atua numa área de pesquisa limitada. Isto é, sugere ao aluno apenas artigos com base naquele que está de momento a consultar. Pelo contrário, este capítulo descreve a implementação de um módulo criado para fazer sugestões de uma maneira mais geral, não com base em palavras-chave ou conteúdos, mas sim nos artigos consultados pelos alunos mais próximos do aluno em consideração – i.e.: alunos com as mesmas disciplinas e estilo de aprendizagem predominante.

Este sistema tem em conta os melhores OAs (com o melhor compromisso entre a pontuação e tempo de visualização associados) e aplica algoritmos de extração de regras de associação. Os resultados, embora preliminares, mostram a viabilidade da metodologia (precisão de 86,49 %).

2 Estado de arte

Na última década os sistemas de recomendação têm sido amplamente utilizados em diversos sistemas, principalmente *on-line* em *sites* de *e-commerce*. Estes sistemas são normalmente utilizados para apresentar ao utilizador informação que favoreça os seus interesses¹³ com base em informação consultada por utilizadores semelhantes. Segundo o trabalho de (Itmazi et al, 2008), nos LMS, regra geral, recomendam-se aos alunos os OAs que lhes estão mais próximos ou seja, que lhes poderão suscitar maior interesse. Os autores deste trabalho descrevem 5 tipos de sistemas de recomendação:

- *Content-Based system (CBS)*: onde os OAs a recomendar são selecionados por terem uma relação direta entre o seu conteúdo e as preferências do aluno;
- *Collaborative Filtering Systems (CFS)*: recomendam OAs com base nas preferências e opiniões que têm em comum com outros utilizadores;
- *Demographic-Based Systems (DBS)*: utilizam conhecimento prévio sobre os alunos e as suas opiniões como base para efetuar recomendações;

¹³ Existem também sistemas de recomendação que fazem precisamente o contrário, isto é, são utilizados com intenção de esconder conteúdo indesejado.

- *Rule-Based Filtering (RBF)*: baseiam-se na filtragem de informação de acordo com um conjunto de regras que expressam essas políticas de filtragem. Estas regras podem estar relacionadas com o utilizador e/ou com o sistema;
- *Hybrid Recommender System (HBS)*: combinação de duas ou mais técnicas de entre as anteriores a fim de obter melhor performance;

Segundo os autores, o habitual é a implementação de HBS's pela sua capacidade de combinar várias técnicas e obter melhor performance. Um exemplo de um HBS é o trabalho (Vanitha et al, 2014). Este sistema de recomendações utiliza o algoritmo *Apriori* para extrair regras de associação, determina as regras mais importantes com recurso a ajuda de peritos e professores, e aplica-as aos alunos, gerando assim uma lista de OAs que lhes poderão interessar.

O trabalho (Khribi et al, 2009) apresenta uma outra abordagem mais genérica que efetua *clustering* das sessões dos utilizadores, aplica regras de associação sobre os *clusters* gerados e procura conteúdos semelhantes às preferências dos utilizadores. Por fim aplica técnicas de recomendação com base nos dados adquiridos.

Regra geral as técnicas mais utilizadas são regras de associação (nomeadamente o algoritmo *Apriori*) e *clustering* em conjunto com algoritmos que verificam a presença das preferências dos utilizadores nos resultados dos algoritmos de *data mining* referidos.

3 Desenvolvimento do sistema de recomendações

Esta secção está dividida em duas partes: a primeira descreve o *dataset* utilizado e como foi obtido; a segunda descreve, em si, a implementação deste módulo.

O diagrama de casos de uso e a sua descrição podem ser consultados no anexo XIII. O seu diagrama de actividades e respetiva especificação encontra-se especificado no anexo IX.

3.1 *Dataset* criado

Em cada sessão os dados a tratar são relativos ao aluno de momento autenticado na plataforma e aos restantes utilizadores. É pois necessário saber quais as disciplinas a que cada aluno está inscrito, quais os seus parâmetros VARK e quais os conteúdos (OAs) que consultou.

O módulo recebe inicialmente o ID do utilizador autenticado que servirá para através de uma *query SQL* obter os seus dados. De seguida é efetuada outra *query SQL* para obtenção dos mesmos dados relativamente aos restantes alunos inscritos na plataforma. As tabelas utilizadas são as tabelas Vark (necessárias para determinar o EA dos alunos); Utilizador (necessária para saber qual o utilizador que se está a ter em conta); NavegacaoLog que contém os dados relativos às consultas de conteúdos; UtilizadorTesteAvaliacao que contém os dados da média dos alunos e Disciplina que contém os dados de uma disciplina (e.g.: nome e id). As tabelas e campos descritos encontram-se detalhados na figura 5.1. As tabelas não descritas e presentes na figura servem apenas de interligação entre as restantes.

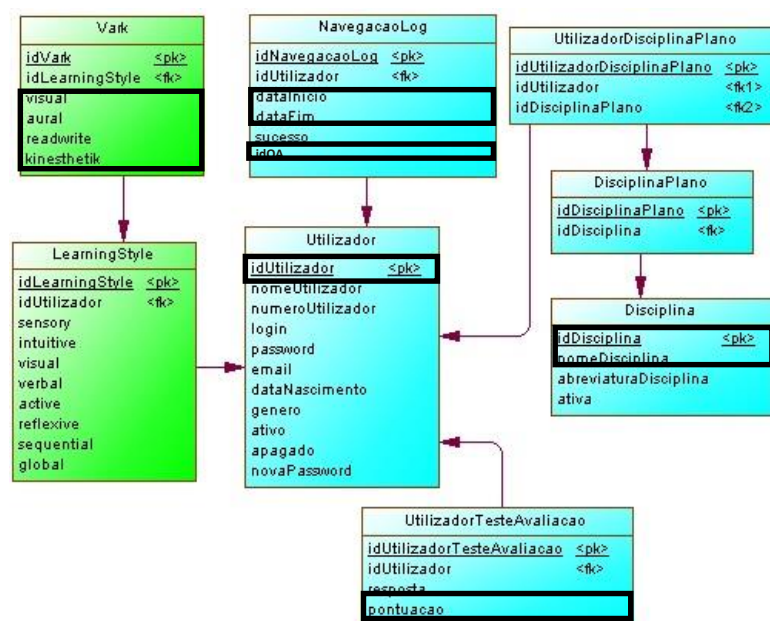


Figura 5.1 - Tabelas utilizadas para a geração do *dataset* do módulo de recomendações.

Os dados obtidos fornecem a maioria dos elementos necessários ao desenvolvimento deste módulo. Contudo, há que determinar ainda o estilo de aprendizagem de cada aluno, incluindo o do autenticado. Para tal, exatamente como nos módulos anteriores, atribui-se como estilo de aprendizagem predominante aquele que tiver maior valor dentro dos quatro parâmetros VARK.

Após o cálculo do estilo de aprendizagem dos alunos, reúnem-se todos os dados necessários ao funcionamento do módulo:

- Estilo de aprendizagem dos alunos;
- Artigos consultados;
- Disciplinas às quais estão inscritos.

Note-se que nem todos estes dados se destinam a figurar *dataset*. Com efeito, os parâmetros estilo de aprendizagem e disciplinas às quais os alunos estão inscritos servem apenas como filtros, permitindo assim registar no *dataset* apenas os artigos consultados por alunos do mesmo estilo de aprendizagem e que frequentam as mesmas disciplinas que o aluno autenticado.

Quanto ao formato do *dataset*, existem limitações relacionadas com a possibilidade de aplicação de um algoritmo de extração de regras de associação (secção seguinte) e que consiste no facto de ter de existir um limite fixo no número de artigos visualizados (i.e., no número de atributos do *dataset*). No módulo desenvolvido optou-se por fixar o limite para o número de atributos do *dataset*, como a média do número de OAs consultados a dividir por 2.

Os OAs consultados por cada aluno são ordenados segundo uma pontuação obtida pela eq. (8).

$$Pontuação = \frac{PontuaçãoArtigo + TempoVisualização}{PontuaçãoArtigo * TempoVisualização} \quad (8)$$

Uma vez ordenados (por ordem decrescente) os OAs pelo valor obtido na fórmula anterior – em que *PontuaçãoArtigo* representa a média das notas associada a todos os alunos que consultaram o OA e *TempoVisualização* o tempo médio de consulta do artigo em causa – limita-se então o número de OAs consultados a metade do número médio de OAs consultados por todos os utilizadores. Um exemplo do *dataset* assim criado apresenta-se na figura 5.2:

```

ConsultedLO0, ConsultedLO1, ConsultedLO2
LO 147, LO 180, LO 185
LO 115, LO 146, LO 179
LO 122, LO 130, LO 178
LO 100, LO 103, LO 108

```

Figura 5.2 - *Dataset* do módulo de recomendação.

Cada linha do *dataset* corresponde aos OAs (LO provém de *Learning Object* em Inglês) consultados por cada utilizador de perfil semelhante ao utilizador atualmente autenticado.

3.2 Algoritmo

Após a criação do *Dataset* no passo anterior, o algoritmo corre em dois passos:

1. Aplicar um algoritmo de extração de regras de associação sobre o *dataset* de modo a obter regras úteis sobre os OAs consultados;
2. Verificar se o utilizador autenticado se enquadra em alguma ou algumas dessas regras.

Extração de regras

Para extrair regras de associação utilizou-se o algoritmo *Apriori* (implementação do Weka). Este algoritmo foi parametrado para gerar no máximo 50 regras (o que se pensa ser um número aceitável, pois a geração de mais regras tenderia a criar regras com pouco interesse), com um mínimo de 75% de confiança e com dependência de atributos (quem consulta um OA também consulta outro, os atributos estão relacionados). O algoritmo gera algumas regras como as que são apresentadas na figura 5.3.


```

Best rules found:

1. ConsultedLO0=LO 147 1 ==> ConsultedLO1=LO 180 1 conf:(1) lift:(4) lev:(0.19) [0] conv:(0.75)
2. ConsultedLO1=LO 180 1 ==> ConsultedLO0=LO 147 1 conf:(1) lift:(4) lev:(0.19) [0] conv:(0.75)
3. ConsultedLO0=LO 147 1 ==> ConsultedLO2=LO 185 1 conf:(1) lift:(4) lev:(0.19) [0] conv:(0.75)
4. ConsultedLO2=LO 185 1 ==> ConsultedLO0=LO 147 1 conf:(1) lift:(4) lev:(0.19) [0] conv:(0.75)
5. ConsultedLO0=LO 115 1 ==> ConsultedLO1=LO 146 1 conf:(1) lift:(4) lev:(0.19) [0] conv:(0.75)
6. ConsultedLO1=LO 146 1 ==> ConsultedLO0=LO 115 1 conf:(1) lift:(4) lev:(0.19) [0] conv:(0.75)
7. ConsultedLO0=LO 115 1 ==> ConsultedLO2=LO 179 1 conf:(1) lift:(4) lev:(0.19) [0] conv:(0.75)
8. ConsultedLO2=LO 179 1 ==> ConsultedLO0=LO 115 1 conf:(1) lift:(4) lev:(0.19) [0] conv:(0.75)
9. ConsultedLO0=LO 122 1 ==> ConsultedLO1=LO 130 1 conf:(1) lift:(4) lev:(0.19) [0] conv:(0.75)
10. ConsultedLO1=LO 130 1 ==> ConsultedLO0=LO 122 1 conf:(1) lift:(4) lev:(0.19) [0] conv:(0.75)
11. ConsultedLO0=LO 122 1 ==> ConsultedLO2=LO 178 1 conf:(1) lift:(4) lev:(0.19) [0] conv:(0.75)
12. ConsultedLO2=LO 178 1 ==> ConsultedLO0=LO 122 1 conf:(1) lift:(4) lev:(0.19) [0] conv:(0.75)
13. ConsultedLO0=LO 100 1 ==> ConsultedLO1=LO 103 1 conf:(1) lift:(4) lev:(0.19) [0] conv:(0.75)
14. ConsultedLO1=LO 103 1 ==> ConsultedLO0=LO 100 1 conf:(1) lift:(4) lev:(0.19) [0] conv:(0.75)
15. ConsultedLO0=LO 100 1 ==> ConsultedLO2=LO 108 1 conf:(1) lift:(4) lev:(0.19) [0] conv:(0.75)
16. ConsultedLO2=LO 108 1 ==> ConsultedLO0=LO 100 1 conf:(1) lift:(4) lev:(0.19) [0] conv:(0.75)
17. ConsultedLO1=LO 180 1 ==> ConsultedLO2=LO 185 1 conf:(1) lift:(4) lev:(0.19) [0] conv:(0.75)
18. ConsultedLO2=LO 185 1 ==> ConsultedLO1=LO 180 1 conf:(1) lift:(4) lev:(0.19) [0] conv:(0.75)
19. ConsultedLO1=LO 146 1 ==> ConsultedLO2=LO 179 1 conf:(1) lift:(4) lev:(0.19) [0] conv:(0.75)
20. ConsultedLO2=LO 179 1 ==> ConsultedLO1=LO 146 1 conf:(1) lift:(4) lev:(0.19) [0] conv:(0.75)

```

Figura 5.3 - Regras geradas pelo algoritmo *Apriori*.

Como se pode verificar pela regra 1, por exemplo, quem consulta o OA 147 também consulta o 180 com uma confiança de 100%. A regra 10 afirma que quem consulta o OA 130 também consulta o 122 com uma confiança de 100%.

De notar que as regras aqui apresentadas são apenas do tipo: se consultou X, também consultou Y, mas nem sempre assim é, pois podem existir regras do género: quem consultou W e X também consultou Y e Z.

Verificar se o utilizador autenticado se encaixa em alguma dessas regras

Neste passo verifica-se se o utilizador se enquadra nalguma das regras, i.e.: se algum dos OAs que consultou está presente na lista de OAs que figuram no antecedente das regras geradas. Ou seja, são os OAs que figuram no conseqüente que poderão ser importantes para o utilizador autenticado. Por fim, o módulo devolve a lista com os IDs e nomes dos OAs.

4 Interação com o utilizador e plataforma

Este módulo não interage diretamente com o utilizador, sendo-lhe apenas mostrado o resultado final da sua execução que se traduz numa lista de *links* para os artigos sugeridos.

O módulo é invocado pela plataforma especificando os parâmetros da base de dados em uso, isto é, o ID do utilizador autenticado e o ID do conteúdo em visualização (que é utilizado apenas para prevenir que também apareça na lista de OAs recomendados). O módulo é invocado quando uma página do sistema é carregada, isto é, quando um conteúdo do guião do aluno é carregado, apresentando conteúdos alternativos.

5 Performance

Para calcular a performance deste módulo criou-se um *dataset*¹⁴ composto por 12 alunos. Estes alunos distribuem-se de acordo com os seus parâmetros de estilo de aprendizagem e disciplinas nas quais estão inscritos, conforme representado na tabela 5.1:

Tabela 5.1 - Distribuição dos alunos no *dataset*

Aluno	Disciplina(s) à(s) qual(is) o aluno está inscrito	Estilo de aprendizagem predominante
1	Sistemas digitais	<i>visual</i>
2	Paradigmas da programação	
3	Sistemas digitais, Paradigmas da programação	
4	Sistemas digitais	<i>aural</i>
5	Paradigmas da programação	
6	Sistemas digitais, Paradigmas da programação	
7	Sistemas digitais	<i>readwrite</i>
8	Paradigmas da programação	
9	Sistemas digitais, Paradigmas da programação	
10	Sistemas digitais	<i>kinesthetic</i>
11	Paradigmas da programação	
12	Sistemas digitais, Paradigmas da programação	

Para os 12 alunos foi criado um modelo que especifica os resultados esperados para uma recomendação "ideal":

- Para um aluno inscrito a "Sistemas Digitais" e com um estilo de aprendizagem *aural* os conteúdos a recomendar seriam os consultados por alunos com o mesmo estilo de aprendizagem e da mesma disciplina;
- Para um aluno inscrito a "Sistemas Digitais" e "Paradigmas da Programação" e com um estilo de aprendizagem *visual* os Conteúdos a recomendar seriam os consultados por alunos com o mesmo estilo de aprendizagem e as mesmas disciplinas;
- O modelo não prevê qualquer tipo de intersecções entre as consultas efetuadas por alunos pertencentes a várias disciplinas.

O módulo foi testado a fim de verificar se as recomendações por ele efetuadas estão ou não em concordância com o desejado, calculando assim a sua precisão.

A precisão do módulo quando comparado com o modelo criado atinge os 86,49 %. Os 100% não são atingidos porque o sistema prevê a intersecção de disciplinas (e o modelo não) E.g.: os alunos inscritos a "Sistemas Digitais" e "Paradigmas da Programação" contribuem para a geração de recomendações dos alunos pertencentes a "Sistemas digitais" e alunos pertencentes a "Paradigmas da programação". Aparentemente esta intersecção pode parecer um erro mas a

¹⁴ O *dataset* encontra-se especificado no anexo XVI.

verdade é que a literatura admite que no caso dos sistemas de sugestão, nem tudo pode ser “a preto e branco”, isto é, muitas vezes é necessário recomendar algo que não esteja totalmente relacionado com o utilizador, criando-se assim alguma variabilidade e novidade.

Uma vez mais, a dificuldade na obtenção de dados e de curadores do modelo criaram barreiras difíceis de transpôr mas que ainda assim se tentaram colmatar da melhor maneira.

Um *screenshot* do módulo em funcionamento pode ser visto na figura 5.4, mas o capítulo seguinte, totalmente dedicado a protótipos desenvolvidos no âmbito deste projeto, clarifica melhor a interface com o utilizador.

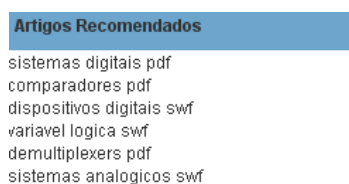


Figura 5.4 - *Screenshot* do módulo de recomendação em operação.

Uma vez que o módulo não tem interação com o utilizador a sua interface é bastante simples, consistindo apenas numa área com os vários OAs recomendados.

6 Conclusões

A sugestão/recomendação de novos artigos ao utilizador é uma mais-valia do ponto de vista académico, uma vez que poderá alargar os seus horizontes, tendo assim algum impacto no seu desenvolvimento e resultados finais. Os resultados, embora preliminares, mostram a viabilidade da metodologia – O módulo atinge uma precisão de 86,49 %.

Num futuro próximo espera-se de facto juntar uma equipa que, em consenso, crie um modelo de acordo com as suas ideologias sobre quais as melhores recomendações para um aluno. Por fim, espera também testar-se a metodologia com um volume considerável de dados reais.

CAPÍTULO VI – PROTÓTIPOS DESENVOLVIDOS

Este capítulo descreve o desenvolvimento de alguns protótipos e encontra-se organizado da seguinte forma: introdução; motor de busca e sistema de recomendações; e plataforma de análise da base de dados do projeto.

1 Introdução

No âmbito desta tese foram desenvolvidos alguns protótipos que são descritos neste capítulo. Dos protótipos desenvolvidos destacam-se dois grupos: o motor de busca e sistema de recomendações e a plataforma de análise de dados do projeto. O motor de busca e sistema de recomendações é composto por dois protótipos:

- Construção de índices – Aplicação que permite a criação periódica dos índices necessários ao motor de busca;
- Visualização de resultados – Aplicação que permite visualizar os resultados obtidos tanto para o motor de busca como para o sistema de recomendações;

A plataforma de análise da base de dados é composta também por dois protótipos:

- Criação de dados – Permite efetuar a criação periódica dos dados estatísticos necessários à plataforma de análise da base de dados;
- Visualização de resultados – Permite visualizar os resultados obtidos por esta plataforma;

Todos os protótipos foram desenvolvidos com recurso à *toolkit* SWING (Swing Framework, 2014) de JAVA e são descritos neste capítulo.

2 Motor de busca e sistema de recomendações

Como já foi explicado, um motor de busca é composto maioritariamente por dois módulos distintos: a indexação e a pesquisa. A indexação é algo a ser feito periodicamente pois todos os OAs inseridos têm de ser indexados antes de se tornarem pesquisáveis pelo motor de busca. Assim, este ponto apresenta dois protótipos: o primeiro permite definir os parâmetros necessários à indexação de OAs por parte do módulo do motor de busca; o segundo permite a visualização dos resultados de uma pesquisa efetuada pelo motor de busca e de uma recomendação levada a cabo pelo módulo de recomendações.

2.1 Construção de índices

A construção de índices é uma aplicação simples em interface SWING que permite ao utilizador definir os parâmetros para a indexação dos OAs presentes na base de dados. Dos parâmetros a definir é obrigatório inserir:

- Base de dados a utilizar;

- Utilizador da base de dados;
- Caminho da diretoria onde serão criados os índices (o caminho terá de ser válido);
- Periodicidade (apenas uma vez ou periodicamente);
- Periodicidade em horas (apenas no caso de na opção anterior ter sido escolhido a função de periodicidade).

A imagem seguinte mostra a interface de parametrização da indexação com que o utilizador se depara:

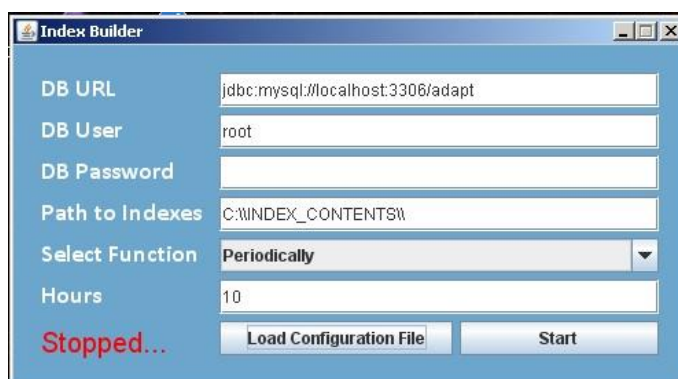


Figura 6.1 - Interface de parametrização da indexação.

A interface permite ainda o carregamento de dados por ficheiro. Para tal o utilizador terá de pressionar o botão “Load Configuration File” e escolher um ficheiro com os dados de cada parâmetro na ordem estipulada, conforme mostrado na figura 6.1.

Por fim o utilizador pressiona o botão “Start” e a aplicação lança uma linha de execução que neste caso fará a indexação de 10 em 10 horas. A interface tem o aspeto representado na figura 6.2:

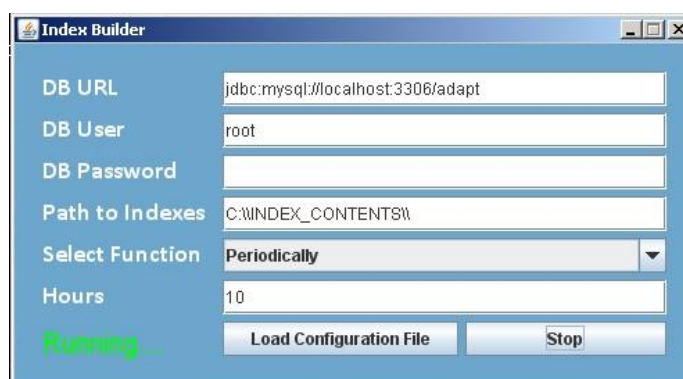


Figura 6.2 - Interface de parametrização da indexação: Aplicação a correr.

Os índices são criados na diretoria “C:\INDEX_CONTENTS”. É criada uma diretoria “LuceneIndexes” que contém os índices fornecidos pelo *Lucene* e um ficheiro de texto “ContentScore.txt” que contém os respetivos parâmetros pontuação, tempo de visualização e estilo de aprendizagem de cada OA.

Este protótipo pode ainda ser lançado na versão linha de comandos que liberta o utilizador da interface gráfica, focando-se apenas na criação de dados. Para isso o utilizador terá de incluir os parâmetros “-F” (significando que os parâmetros necessários ao funcionamento do módulo serão carregados a partir de um ficheiro) e o respetivo ficheiro.

2.2 Visualização de resultados

Uma vez que tanto o motor de busca como o sistema de recomendações não têm qualquer tipo de interação com o utilizador a não ser a apresentação dos resultados obtidos, neste protótipo optou-se por criar um botão que simula um evento “*pageLoad*” da plataforma ADAPT enviando o ID do utilizador e o ID e designação do OA correntemente em visualização para os módulos, mostrando os resultados obtidos. A designação do OA correntemente em visualização serve como frase de pesquisa para o motor de busca, o ID do utilizador e do OA correntemente em visualização servem também os propósitos de ambos os módulos.

A interface deste protótipo pode ser vista na figura 6.3, embora a interface dos módulos nela contidos (motor de busca e sistema de recomendações) sejam apresentados em maior detalhe nas figuras 6.4 e 6.5:



Figura 6.3 - Visualização de resultados do protótipo do motor de busca e sistema de recomendações.

Da esquerda para a direita, os retângulos a cor preta representam os resultados do módulo do motor de busca, o conteúdo correntemente em visualização pelo aluno e os resultados obtidos pelo sistema de recomendações. O retângulo a cor vermelha assinala o botão que despoleta a ação dos dois módulos.

Após pressionar o botão delimitado pela linha a vermelha, o resultado do módulo do motor de busca é para este caso, o representado na figura 6.4:

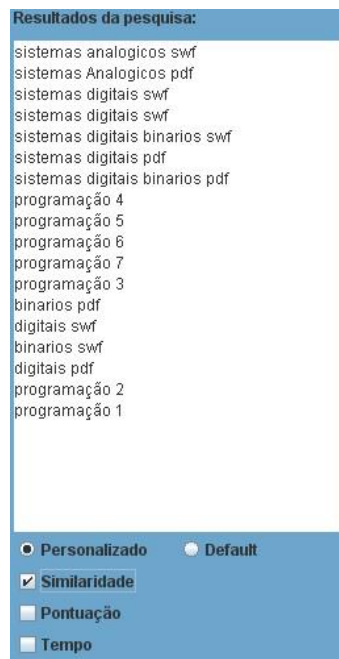


Figura 6.4 - Resultados do módulo do motor de busca.

Os resultados podem ainda ser ordenados pela fórmula original de ordenação ou pela conjugação dos parâmetros de similaridade de cosseno, Pontuação ou Tempo de visualização dos OAs. O resultado do sistema de sugestões é representado na figura 6.5:

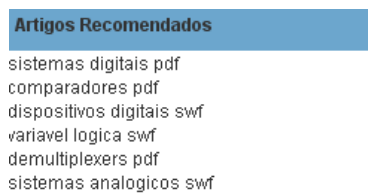


Figura 6.5 - Resultados do sistema de recomendações

Por fim, o protótipo fica com o seguinte aspeto:

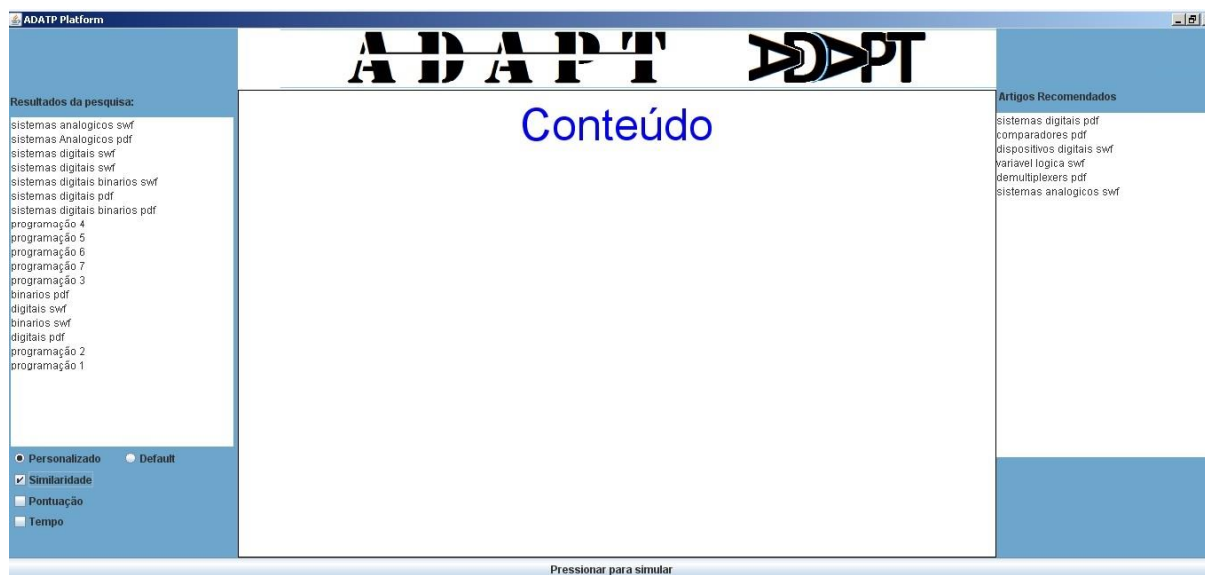


Figura 6.6 - Aspeto do protótipo depois de receber os resultados dos módulos de pesquisa e recomendação.

3 Plataforma de análise da base de dados do projeto

A plataforma de análise utiliza dois módulos distintos: o módulo de geração de dados e o módulo de visualização. Esta separação foi feita porque a geração de dados leva algum tempo a efetuar, uma vez que na sua maioria os dados gerados são imagens e a sua renderização leva algum tempo, obrigando o utilizador a ter de esperar pelos resultados. Deste modo optou-se pela criação de um módulo de geração de dados que cria periodicamente os elementos necessários à plataforma e um módulo de visualização que mostra os resultados ao utilizador.

3.1 Criação de dados

A aplicação que possibilita a criação de dados é em tudo muito semelhante à aplicação de indexação dos OAs previamente descrita no protótipo de indexação de conteúdos do motor de busca. A diferença está nos módulos chamados: esta aplicação chama o módulo de criação de dados da plataforma que efetua várias estatísticas e gera os respetivos gráficos.

A seguinte imagem mostra a interface da aplicação:

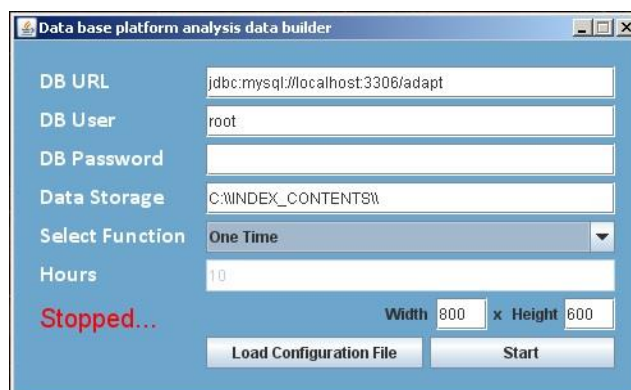


Figura 6.7 - Interface de parametrização da geração de dados.

Como se pode verificar, a interface é semelhante à da indexação dos conteúdos para o motor de busca. Contudo, os dados criados são diferentes: é criada uma diretoria “Results” no caminho especificado no campo “Data Storage”. Na diretoria “Results” existem duas diretorias, uma denominada “QueryResults”, necessária à criação da outra diretoria “Outputs” que contém os dados relativos às perguntas que a plataforma vem a responder. Os parâmetros de “Width” e “Height” permitem especificar o tamanho dos gráficos a criar.

Este protótipo pode ainda ser lançado na versão linha de comandos que liberta o utilizador da interface gráfica, focando-se apenas na criação de dados. Para isso o utilizador terá de incluir os parâmetros “-F” (significando que os parâmetros necessários ao funcionamento módulo serão carregados a partir de um ficheiro) e o respetivo ficheiro.

3.2 Visualização de dados

A plataforma de visualização de dados conta com uma interface gráfica que permite aos utilizadores (tutores/professores) visualizar os dados de quatro categorias específicas e relacionadas entre si:

- Alunos;
- Disciplinas;
- Conteúdos;
- Estilo de aprendizagem;

Estas categorias estão relacionadas entre si e na plataforma são divididas em duas categorias distintas:

- Dados estatísticos;
- *Data mining*;

A figura 6.8 mostra a separação existente:



Figura 6.8 - Plataforma de análise de dados - Separação entre dados estatísticos e *data mining*

Uma vez que cada uma destas *tabs* (categoria de dados) é subdividida em várias outras, são descritas em primeiro lugar as *tabs* relativas à apresentação de dados estáticos e posteriormente as *tabs* relativas a *data mining*.

3.2.1 Dados Estatísticos

Esta *tab* é subdividida nas seguintes subcategorias:

- Utilizadores;
- Disciplinas;
- Geral;

As duas primeiras não implicam interação com o utilizador e apresentam gráficos relativos a:

- Distribuição dos alunos por estilo de aprendizagem;
- Média das notas por estilo de aprendizagem por cada disciplina;
- Percentagem de utilização de artigos por estilo de aprendizagem por cada disciplina;
- Notas discretizadas por disciplina – i.e.: número de “muito fraco”, “fraco”, “insatisfaz”, “satisfaz”, “bom” e “muito bom”, existentes por disciplina;

A última *tab* envolve alguma interação com o utilizador e apresenta dados mais gerais tais como:

- A média das notas e predominância de estilo de aprendizagem por disciplina por curso – i.e.: média e estilo de aprendizagem predominante de cada disciplina por cada curso existente na plataforma;
- Percentagem de mudanças de estilo de aprendizagem ao longo de um período temporal definido pelo utilizador:
 - Neste ponto, o utilizador escolhe a data inicial e final que delimitam o período a avaliar e a plataforma gera um gráfico com as mudanças efetuadas. As figuras 6.9 e 6.10 são ilustrativas;

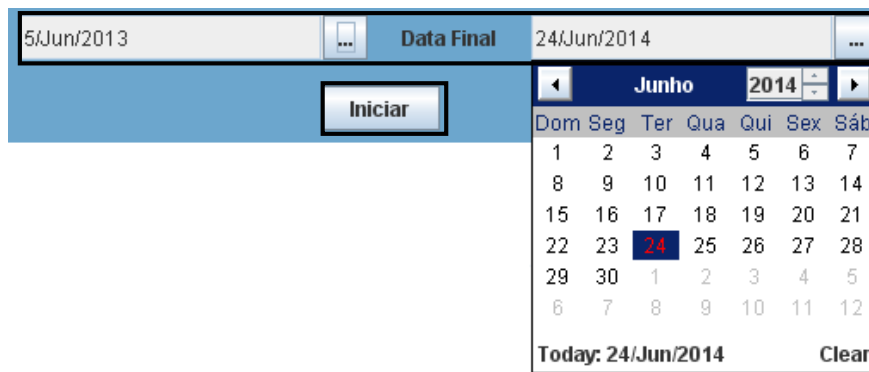


Figura 6.9 - Definição do período temporal para calcular a percentagem de mudanças de estilo de aprendizagem ao longo de um período temporal definido pelo utilizador.

Os cálculos são efetuados logo após o utilizador pressionar o botão iniciar. É gerado um gráfico e é apresentado logo acima do painel da figura anterior:

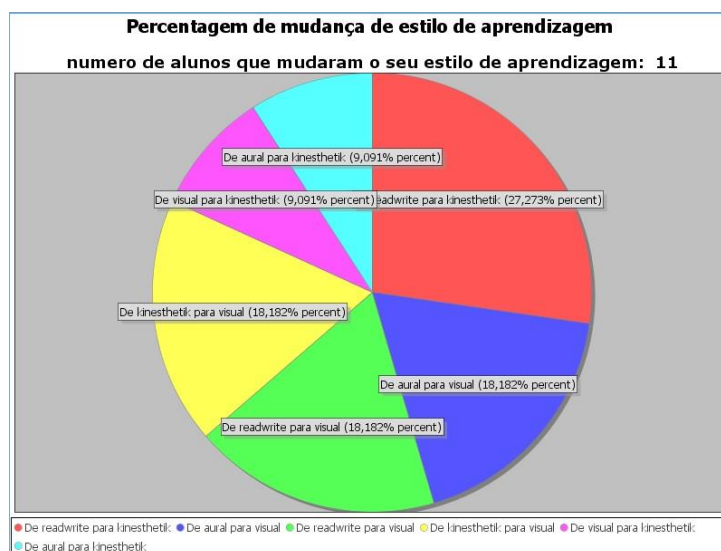


Figura 6.10 - Percentagem de mudanças de estilo de aprendizagem ao longo de um período temporal definido pelo utilizador - Gráfico gerado

3.2.2 Data mining

Esta categoria de algoritmos exige mais interação por parte do utilizador. Conforme referido, a plataforma aplica algoritmos de *data mining* para extrair conclusões que de outro modo não seriam evidentes e que poderão ter utilidade para os professores/tutores. A plataforma divide as técnicas de *data mining* em três categorias:

- *Clustering* – Permite aplicar *clustering* para verificar a existência de grupos de interesse;
- Predição – Permite prever as notas dos alunos que ainda não foram avaliados com base nos artigos consultados e com base em experiências anteriores por outros alunos da plataforma;
- Regras – Permite extrair regras de associação;

A figura 6.11 mostra (delimitado a preto) as *tabs*/subcategorias agrupadas pela categoria de data mining:



Figura 6.11 - Plataforma de análise de dados – Subcategorias da categoria *data mining*.

3.2.2.1 Clustering

A subcategoria de *clustering* subdivide-se ainda em duas categorias:

- *Cluster 1* – Utiliza o algoritmo *K-means* para gerar grupos baseados no género, idade, estilo de aprendizagem predominante e a percentagem de cada estilo de aprendizagem que os alunos têm vindo a consultar;
- *Cluster 2* - Utiliza o algoritmo para gerar grupos baseados nos seguintes parâmetros:
 - Parâmetros VARK utilizador;
 - Média e desvio padrão dos parâmetros VARK dos OAs que consultou;
 - Número de OAs consultados;
 - Pontuação média (pontuação média que o aluno teve a cada disciplina);

Esta funcionalidade fornece ao utilizador a possibilidade de gerar *clusters* em períodos temporais lado a lado permitindo-lhe avaliar se algum utilizador mudou de *cluster*.

3.2.2.1.1 Cluster 1 – *Cluster* de utilizadores segundo os seus parâmetros

Para gerar grupos utilizando o *K-means*, o utilizador apenas terá de escolher o valor de *k* (número de grupos que o algoritmo deve gerar) e pressionar no botão iniciar. A figura em baixo apresenta a utilização da plataforma e os respetivos resultados obtidos para $K = \sqrt{\frac{N}{2}}$ *clusters* ($K = -1$), em que *N* é o número de instâncias do *dataset*:

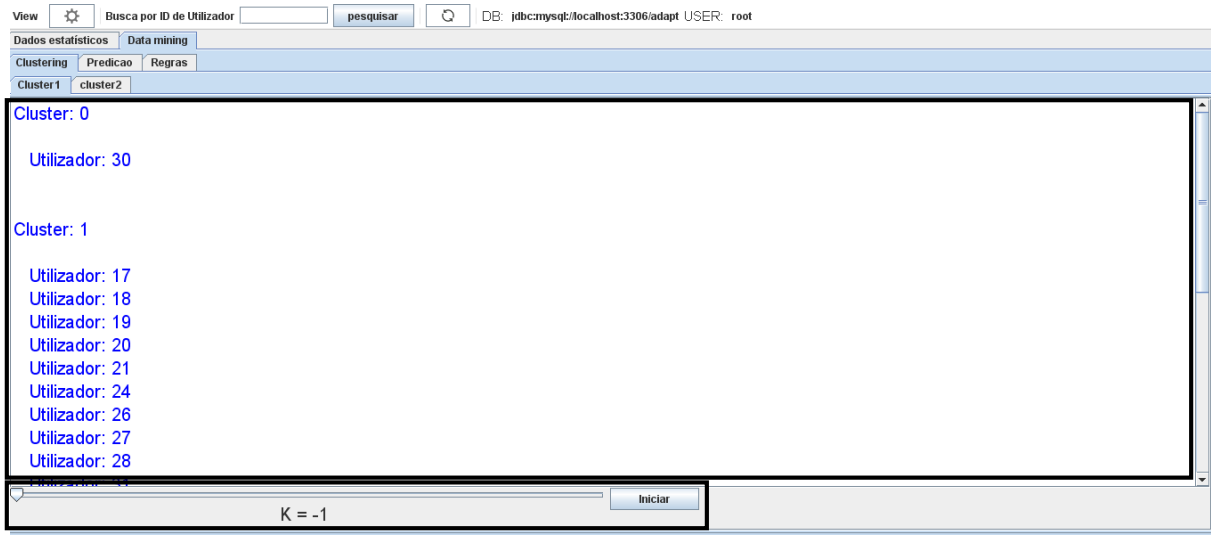


Figura 6.12 - Clustering – *K-means*

A apresentação de dados é feita em modo textual, não muito diferente da apresentação por defeito efetuada pelo WEKA ou rapidminer (Rapidminer, 2014).

3.2.2.1.2 Cluster 2 – *Cluster* de utilizadores segundo os parâmetros dos conteúdos consultados

Como foi explicado anteriormente, esta funcionalidade permite que o utilizador crie *clusters* em períodos temporais visíveis lado a lado, permitindo-lhe avaliar se algum utilizador mudou de *cluster* ao longo do tempo. Deste modo, o utilizador deve escolher em primeiro lugar os períodos temporais nos quais deseja fazer *clustering* e pressionar o botão iniciar de “ambos os ecrãs”: o correspondente ao primeiro período temporal no qual deseja verificar a existência de grupos e o correspondente ao segundo período temporal.

Esta funcionalidade utiliza o algoritmo *X-means* para gerar grupos e apresenta os dados ao utilizador como se mostra na figura 6.13.

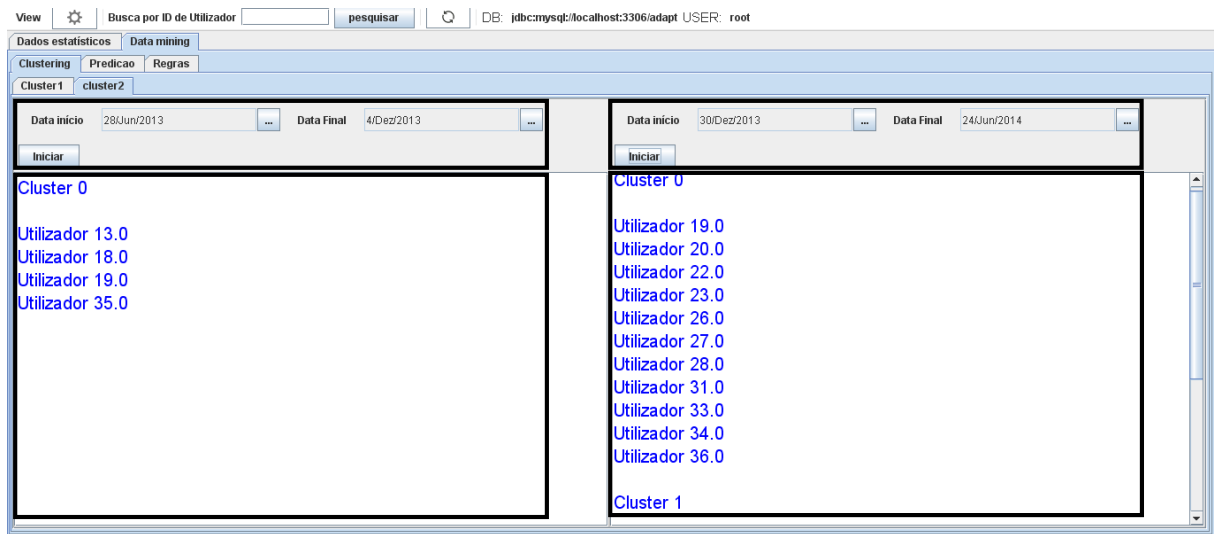


Figura 6.13 - Clustering por períodos temporais – X-means.

3.2.2.2 Predição

Esta subcategoria fornece ao utilizador a possibilidade de prever se um aluno não avaliado terá mais possibilidade de ter média positiva ou negativa a uma disciplina, com base em dois fatores:

- O seu comportamento na plataforma;
- Comportamento e média dos seus colegas já avaliados.

Para visualizar os dados, o utilizador terá apenas de pressionar o botão iniciar que despoleta a funcionalidade 9 e retorna os resultados como se pode verificar na figura 6.14:

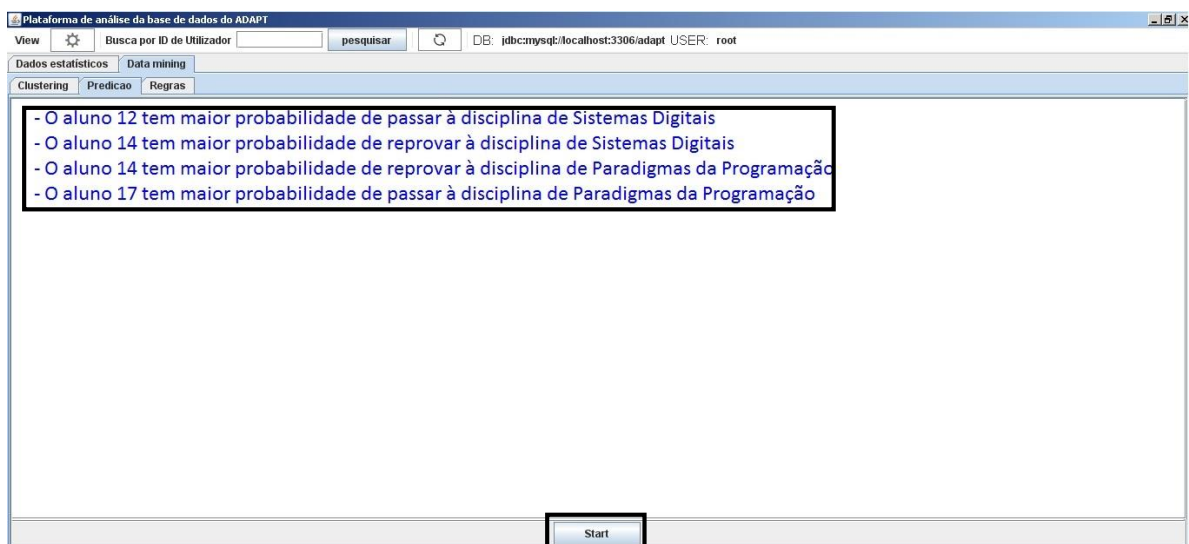


Figura 6.14 - Predição da média dos alunos não avaliados.

3.2.2.3 Regras

Esta subcategoria permite ao utilizador extrair regras sobre os alunos com base nos seguintes parâmetros:

- Idade;
- Estilo de aprendizagem;
- Média de notas;

A interação do utilizador com a aplicação prevê também que este especifique o grau de confiança das regras extraídas e o número de regras a extrair.

A figura 6.15 apresenta os resultados para $\sqrt{\frac{N}{2}}$ regras, em que N é o número de instâncias do *dataset*:

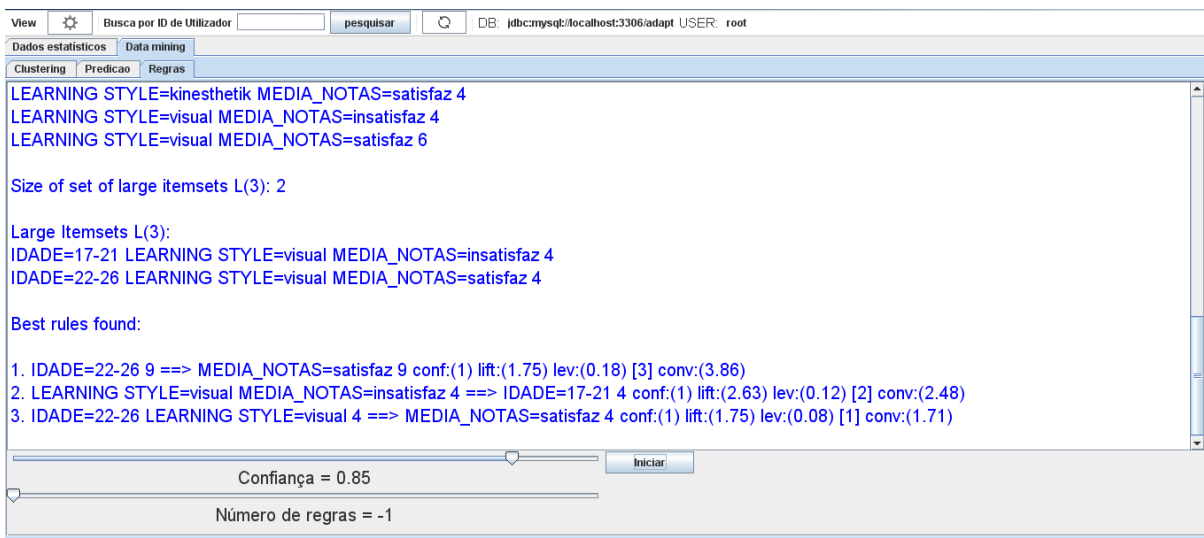


Figura 6.15 - Extração de regras de associação.

3.2.3 Outras funcionalidades

Existem ainda algumas funcionalidades adicionais fornecidas pela plataforma: a atualização de todos os dados em simultâneo, a possibilidade da definição dos parâmetros de acesso à base de dados e a possibilidade de pesquisar pelo nome do aluno segundo o seu ID.

A atualização de dados é possível graças ao botão “atualizar” (↻), existente na barra superior da plataforma que atualiza todos os dados visíveis na plataforma.

A definição dos parâmetros da base de dados é possível pressionando o botão “definições” (⚙) que abre uma janela como a apresentada na figura 6.16.

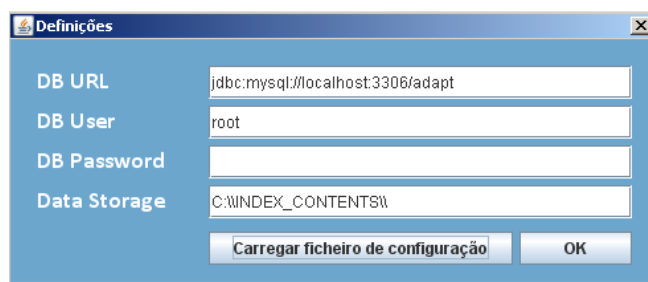


Figura 6.16 - Definições da plataforma de visualização de dados.

Esta janela permite ao utilizador especificar a URL da base de dados, o nome de utilizador e palavra-chave para aceder à base de dados e o caminho de criação da diretoria onde os dados devem ser guardados. A especificação destes parâmetros pode ser feita carregando um ficheiro ao pressionar o botão “Carregar ficheiro de configuração”. O ficheiro a carregar deve conter os parâmetros segundo a ordem especificada na caixa.

A pesquisa pelo ID do utilizador é efetuada inserindo na caixa de texto existente na barra superior da aplicação o ID do utilizador e pressionar o botão pesquisar. Esta funcionalidade existe, porque trabalhando com data mining, na maioria das vezes utilizam-se IDs e não nomes:

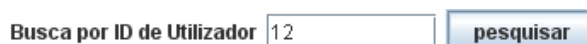


Figura 6.17 - Pesquisa por id de utilizador.

O utilizador verá o nome do utilizador numa caixa de texto tal como mostra a figura 6.18:



Figura 6.18 - Resultado da pesquisa pelo ID do utilizador

CAPÍTULO VII – CONCLUSÕES E TRABALHO FUTURO

O desenvolvimento dos quatro módulos descritos nesta dissertação era imperativo para um funcionamento mais completo da plataforma ADAPT. Os módulos desenvolvidos pertencem na sua grande parte à área de *link-mining* e fornecem à plataforma interatividade e automatização – duas características desejadas nos sistemas de *e-Learning*. Através da exploração de *links* e sugestão de conteúdos alternativos, a plataforma está capacitada para acompanhar o processo de formação dos alunos, ajustando-se às suas necessidades e sugerindo conteúdos alternativos, dois objetivos do projeto ADAPT e características desejadas num LMS – cada aluno tem o seu próprio estilo de aprendizagem que pode mudar ao longo do tempo. Um bom sistema terá de acompanhar essa mudança e paralelamente, fornecer ao aluno conteúdos alternativos (variabilidade).

Embora os módulos não tenham ainda sido completamente integrados na plataforma, as experiências efetuadas e os protótipos desenvolvidos demonstram que têm grande potencial: o motor de busca apresenta uma *accuracy* de 27%, precisão de 73,14% e *recall* de 40,50% na procura dos OAs mais adequados a cada aluno. Na ordenação de resultados o módulo apresenta uma distância de *Hamming* ao modelo criado de 108, a sua média é de 3,6 e o desvio padrão é de 4,30; o classificador de estilos de aprendizagem tem uma *accuracy* de 75%; o sistema de recomendações uma precisão de 86,49% e a rede neuronal implementada na predição de notas integrada na plataforma de análise de dados atinge respectivamente 100% para *split* de 90% utilizando um *dataset* composto por 20 utilizadores (linhas) e de 77% para *split* de 66%.

Contudo, há que referir que os módulos foram desenvolvidos com recurso apenas a dados experimentais, dificultando as experiências desenvolvidas, dado que trabalhar em *data mining* requer grandes quantidades de dados. Esta lacuna deve-se ao facto de existirem poucos registos de dados, indispensáveis ao desenvolvimento dos módulos dada a pouca utilização da plataforma ADAPT até este momento. A falta de curadores para os modelos de teste criados foi também um entrave.

Como trabalho futuro espera-se que no decorrer dos próximos anos letivos se obtenha um volume de dados desejável para testar e validar os módulos desenvolvidos com dados reais. Espera-se também melhorar a performance dos algoritmos implementados que podem ser melhorados se testados com dados reais.

Ao longo do tempo espera-se também melhorar a plataforma (quer a nível de interação quer a nível de funcionalidades), que está aberta a críticas e sugestões construtivas e fundamentadas.

REFERÊNCIAS BIBLIOGRÁFICAS

- ADAPT (2014). <http://gecad.isep.ipp.pt:8181/>. Plataforma de *e-Learning* ADAPT (página internet oficial), Portugal.
- Adobe Connect (2014): <http://www.adobe.com/pt/products/adobeconnect.html>
- APACHE FOUNDATION (2012). <http://www.apache.org/licenses/>. Apache software foundation (página internet oficial).
- API do WEKA: <http://weka.sourceforge.net/doc.stable/>
- Bailin, A. e Levin, L. Introduction: Intelligent Computer Assisted Language Instruction. 1989 *Computers and the Humanities*, 23, 3-11
- Bates, A. (2005). *Technology, e-Learning and Distance Education* London: Routledge (em Inglês).
- Blackboard Collaborate (2014): <http://www.blackboard.com/platforms/collaborate/products/blackboard-collaborate.aspx>
- Castro, F., Vellido, A., Nebot, A. e Mugica, F.. Applying Data Mining Techniques to *e-Learning*. *Computational Intelligence Volume 62*, 2007, pp 183-221
- Chambers, J. e Sprecher, J.. *Computer-Assisted Instruction: Its Use in the Classroom*. Englewood Cliffs, New Jersey: Prentice-Hall Inc. 1983
- Cisco WebEx (2014): <http://www.webex.com/index-sky-3.html>
- Coffman, J, e Weaver, A. C. (2011). Learning to Rank Results in Relational Keyword Search. *CIKM '11 Proceedings of the 20th ACM international conference on Information and knowledge management*, Glasgow, Scotland, UK.
- Corbett, A.T. e Anderson, J. R.. LISP Intelligent Tutoring System Research in Skill Acquisition. In Larkin, J. & Chabay, R. (Eds.) *Computer assisted instruction and intelligent tutoring systems: shared goals and complementary approaches* (pp.73-110) Englewood Cliffs, New Jersey: Prentice-Hall Inc. 1992
- Desire2Learn (2013): <http://www.desire2learn.com/>
- e-Learning* (2014). <http://en.wikipedia.org/wiki/E-Learning>.
- Estatísticas *e-Learning* (2011-2014): <http://e-Learningindustry.com/important-statistics-about-the-e-Learning-market-for-2013-infographic>
- FCT (2014). <http://www.fct.pt/>. Fundação para a Ciência e a Tecnologia (página internet oficial), Portugal.

- FEDER (2011). <http://www.qca.pt/fundos/feder.asp>. Fundo Europeu de Desenvolvimento Regional (página internet oficial), Portugal.
- Felder, R.M., Learning and Teaching Styles. *Engineering Education, Journal of Engineering Education*, Vol. 78, No. 7, pp. 674-681, 1988
- Ferraz, A., P., C., M. e Belhot, R., V.. Taxonomia de Bloom: revisão teórica e apresentação das adequações do instrumento para definição de objetivos instrucionais. *Gest. Prod.* vol.17 no.2 São Carlos 2010.
- Ford, L. A New Intelligent Tutoring System. *2008 British Journal of Educational Technology*, 39(2), 311-318
- FORMULA LUCENE (2012). http://lucene.apache.org/core/4_0_0/core/org/apache/lucene/search/similarities/TFIDFSimilarity.html. Lucene (página internet oficial).
- GECAD (2011). <http://www.gecad.isep.ipp.pt/>. Grupo de Investigação em Engenharia do Conhecimento e Apoio à Decisão. (página internet oficial), Porto.
- Getoor, L. e Diehl, C. P.. (2005). Link Mining: A Survey. *SIGKDD Explorations Volume 7 Issue 2*.
- Gomes, A., Learning Styles. *An E-Learning Tool*, International Conference on Engineering Education – ICEE 2007, Coimbra, Portugal, 2000
- Graf, S., Kinshuk e Tzu-Chien Liu (2008). Identifying Learning Styles in Learning Management Systems by Using Indications from Students' Behavior. *Advanced Learning Technologies, 2008. ICALT '08. 8th IEEE International Conference*
- Harb, H. M., Khalifa, A. R., Ishkewy, H. M. (2009). Personal Search Engine Based on User Interests and Modified Page Rank. *Computer Engineering & Systems, 2009. ICCES 2009. Cairo, Egito*.
- Haykin S. (1998). *Neural Networks: A Comprehensive Foundation*. USA (em Inglês)
- Hrastinski, S.. Synchronous *e-Learning* vs. asynchronous *e-Learning*. *EDUCAUSE Quarterly*, vol. 31, no. 4 (October–December 2008)
- ISEC (2014). <http://www.wisec.pt/>. Instituto Superior de Engenharia de Coimbra (página internet oficial), Coimbra.
- ISEP (2014). <http://www.isep.pt/>. Instituto Superior de Engenharia do Porto (página internet oficial), Porto.
-

- Itmazi, J. e Megías, M.. Using recommendation systems in Course Management Systems to recommend Learning objects. The international Arab Journal of Information Technology, Vol. 5, No.3, Julho 2008.
- JFreeChart (2012): <http://www.jfree.org/jfreechart/>.
- Jolliffe, I.T.. Principal Component Analysis. 2nd edition 2002. ISBN 0-387-95442-2.
- Kawakami, M., Yasuda, H. e Sasaki, R. , Sanya. Development of an *E-Learning* Content-Making System for Information Security (ELSEC) and its Application to Anti-phishing Education. e-Education, e-Business, e-Management, and *e-Learning*, 2010. IC4E '10. International Conference, Sanya.
- Khribi, M. K., Jemni, M. e Nasraoui, O.. Automatic Recommendations for *E-Learning* Personalization Based on Web Usage Mining Techniques and Information Retrieval. Educational Technology & Society, 12 (4), 30–42. 2009
- K-means (2014): http://en.wikipedia.org/wiki/K-means_clustering
- Kolb,A., Kolb,D.A., Learning styles and learning spaces: Enhancing experiential learning in higher education, Academy of Management Learning and Education, 4(2), 193-212, 2005
- Larkin, J. e Chabay, R. (Eds.). Computer Assisted Instruction and Intelligent Tutoring Systems: Shared Goals and Complementary Approaches. Hillsdale, New Jersey: Lawrence Erlbaum Associates. 1992
- LUCENE (2012). <http://lucene.apache.org/>. Framework Lucene Apache (página internet oficial).
- LUCENE (2012). <http://lucene.apache.org/>. Framework Lucene Apache (página internet oficial).
- LUCENE API (2013).http://lucene.apache.org/core/3_0_3/api/all/org/apache/lucene/. Lucene core API (página internet oficial).
- Marques, V., M, Pereira, C., Gomes, A., Reis, C., Faria, L., Martins, C., Pires, E., E., S., Pedagogy Embedding in a Learning Management System - The ADAPT Project. Revista OPDES - Pedagogia no Ensino Superior, Edição ESEC - Escola Superior de Educação de Coimbra, Dezembro de 2010
- McCandless, M., Hatcher, E. e Gospodnetić, O (2010). Lucene in Action (em Inglês).
- Mihăescu, M. C.. Classification of Learners Using Linear Regression. Computer Science and Information Systems (FedCSIS), 2011 Federated Conference em 18-21 Sept. 2011
- Modelo de Gregorc (2014): <http://gregorc.com/books.html#id5>

- Modelo de Honey e Mumford (2014)
<http://www2.le.ac.uk/departments/gradschool/training/eresources/teaching/theories/honey-mumford>
- Moodle: <https://moodle.org/?lang=pt>
- MultilayerPerceptron: <http://www.dtreg.com/mlfn.htm>
- Nirmalani, C.. Handbook of Research on Educational Communications and Technology (2014). (em Inglês).
- Nowicki, K. e Snijders, T. A. B.. (1997). Estimation and prediction for stochastic blockstructures. *Journal of Classification* Volume 14 Issue 1
- Ozmutlu, H. C., Cavdur, F. e Ozmutlu, S.. (2004). Neural Network Applications for Automatic New Topic Identification on Excite Web Search Engine Data Logs. *Proceedings of the American Society for Information Science and Technology*. Volume 41, Issue 1, pages 310–316.
- Pages 727-734. 2000"
- Pelleg, D. e Moore, A.. X-means: Extending K-means with Efficient Estimation of the Number of Clusters. *ICML '00 Em Proceedings of the Seventeenth International Conference on Machine Learning*
- Preethi, N. e Devi, T. (2013). New Integrated Case and Relation Based (CARE) Page Rank Algorithm. *Computer Communication and Informatics (ICCCI), 2013 International Conference*.
- Rapidminer (2014): <http://rapidminer.com/>.
- Rojas, R.. *The Backpropagation Algorithm*. 1996
- Romero, C., Romero, J. R., Luna, J. M. e Ventura, S.. Mining Rare Association Rules from *e-Learning* Data. *Proceeding of The Third International Conference of Education Data Mining, Pittsburgh, USA, 171-180*. 2010
- Romero, C., Ventura, S., Espejo, P.G. e Hervás, C. (2008). Data Mining Algorithms to Classify Students. *Proc. of the 1st Int. Conf. on Educational Data Mining (EDM'08)*
- Rosas, C. C., Ponce, G. A. e Lopez-Morteo, G. A.. (2011). A Specialized Search Assistant for Learning Objects. *ACM Journal* Volume 5 Issue 4.
- Sael, N., Marzak, A., e Behja, H.. Multilevel clustering and association rule mining for learners' profiles analysis. *International Journal of Computer Science Issues (IJCSI)*. May 2013, Vol. 10 Issue 3
-

- Sakai (2014): <http://sakaiproject.org/>
- SCORM: <http://scorm.com/pt/scorm-explicou/>.
- Shute, V. J. e Psocka, J.. Intelligent Tutoring Systems: Past, Present, and Future. Human resources directorate manpower and personnel research division. pp. 2-52. 1994.
- Silverstein, C., Marais, H., Henzinger, M. e Moricz, M.. (1999). Analysis of a very large web search engine *query* log. ACM SIGIR Volume 33 Issue 1, Fall 1999
- Stash,N., Cristea,A., De Bra,P., Explicit Intelligence in Adaptive Hypermedia: Generic Adaptation Languages for Learning Preferences and Styles. Workshop CIAH2005, Combining Intelligent and Adaptive Hypermedia Methods/Techniques em Web Based Education Systems, em conjunto com HT'05, pp., 75-84, 2005
- Swing Framework (2014): <http://docs.oracle.com/javase/7/docs/api/javafx/swing/package-summary.html>
- UTAD (2014). <http://www.utad.pt/>. Universidade de Trás-os-Montes e Alto Douro (página internet oficial), Vila Real.
- Valsamidis, S. , Kontogiannis, S., Kazanidis, I., Karakos, A.. *E-Learning* Platform Usage Analysis. Interdisciplinary Journal of *E-Learning* and Learning Objects Volume 7, 2011
- Vanitha, V. e Shanthi, R.. Hybrid recommender System for *e-Learning* applications. NCI2TM: 2014.
- VARK (2014): A Guide to Learning Styles, <http://www.vark-learn.com/english/index.asp>
- Venkatesan, K., Nelaturu, S., Vullamparthi, A.J. e Rao, S.. Hybrid ontology based e - Learning expert system for children with Autism. Information and Communication Technology (ICoICT), 2013, Bandung
- Wei, C., Fang, Z., Zhang, Y. e Jin, Y.. A Personalized CAI Courseware System. CiSE 2009. Wuhan
- Woolley, D. R. (12 February 2013). "PLATO: The Emergence of Online Community". <http://thinkofit.com/plato/dwplato.htm>.
- Zakrzewska, D.. Cluster Analysis for Users' Modeling in Intelligent *E-Learning* Systems. Computer Science Volume 5027, 2008, pp 209-214
- Zorrilla, M., García, D. e Álvarez, E. (2010). A Decision Support System to improve *e-Learning* Environments. Proceedings of the 2010 EDBT/ICDT Workshops

ANEXOS

Anexo I – Diagrama da Base de dados;

Anexo II – Diagrama de casos de uso do módulo do motor de busca;

Anexo III – Diagrama de actividades do módulo do motor de busca;

Anexo IV – Diagrama de casos de uso do módulo do classificador de estilos de aprendizagem;

Anexo V – Diagrama de actividades do módulo do classificador de estilos de aprendizagem;

Anexo VI – Diagrama de casos de uso da plataforma de análise de dados;

Anexo VII – Diagrama de actividades da plataforma de análise de dados;

Anexo VIII – Diagrama de casos de uso do módulo do sistema de recomendações;

Anexo IX – Diagrama de actividades do módulo do sistema de recomendações;

Anexo X – *Dataset* utilizado nos testes de performance do motor de busca;

Anexo XI – *Dataset* utilizado nos testes de performance do classificador de estilos de aprendizagem;

Anexo XII – *Dataset* utilizado para efetuar o clustering de utilizadores (funcionalidade 1 da plataforma de análise de dados);

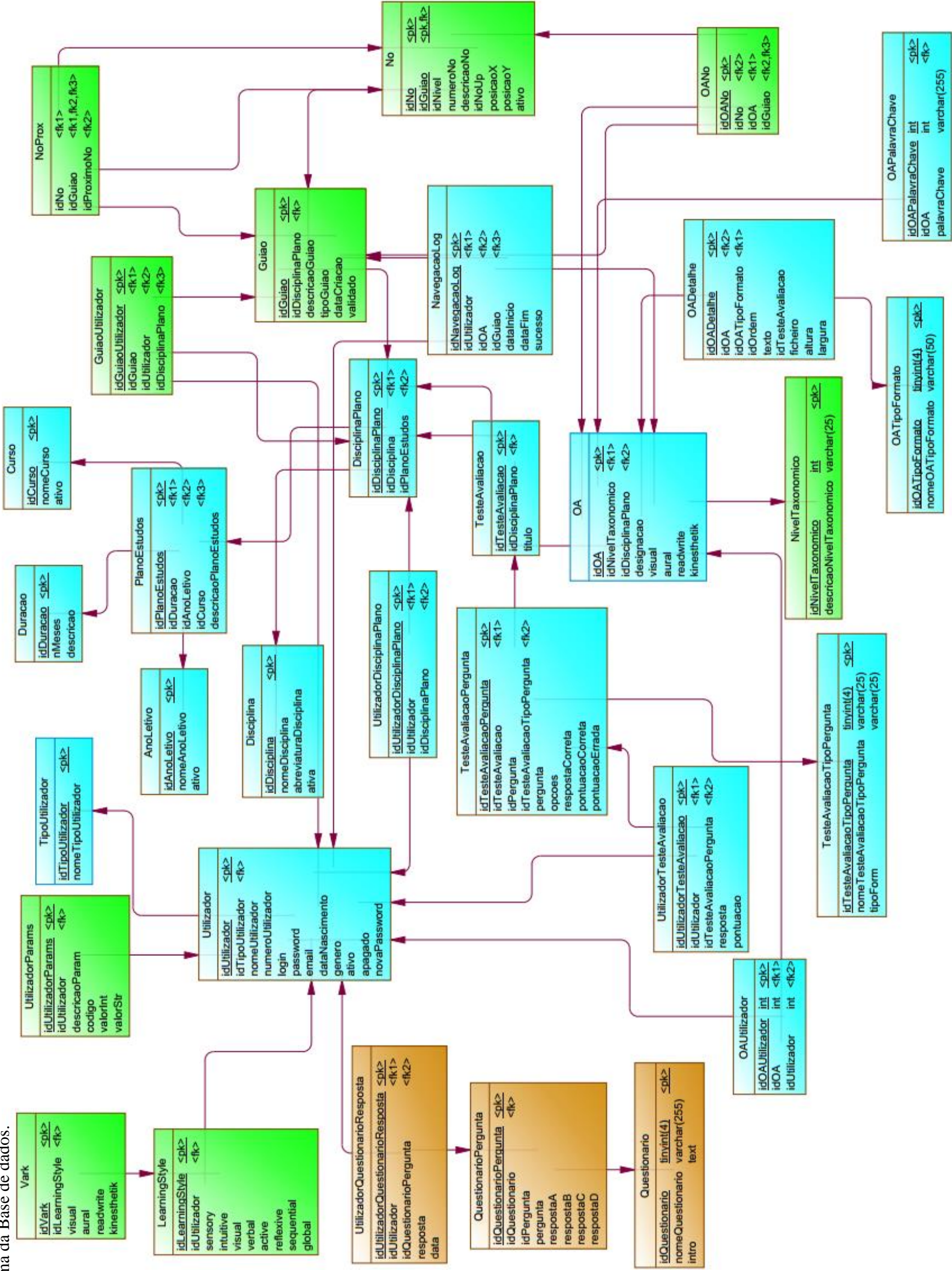
Anexo XIII – *Dataset* utilizado para gerar regras de associação sobre os dados dos utilizadores (feature 3 da plataforma de análise de dados);

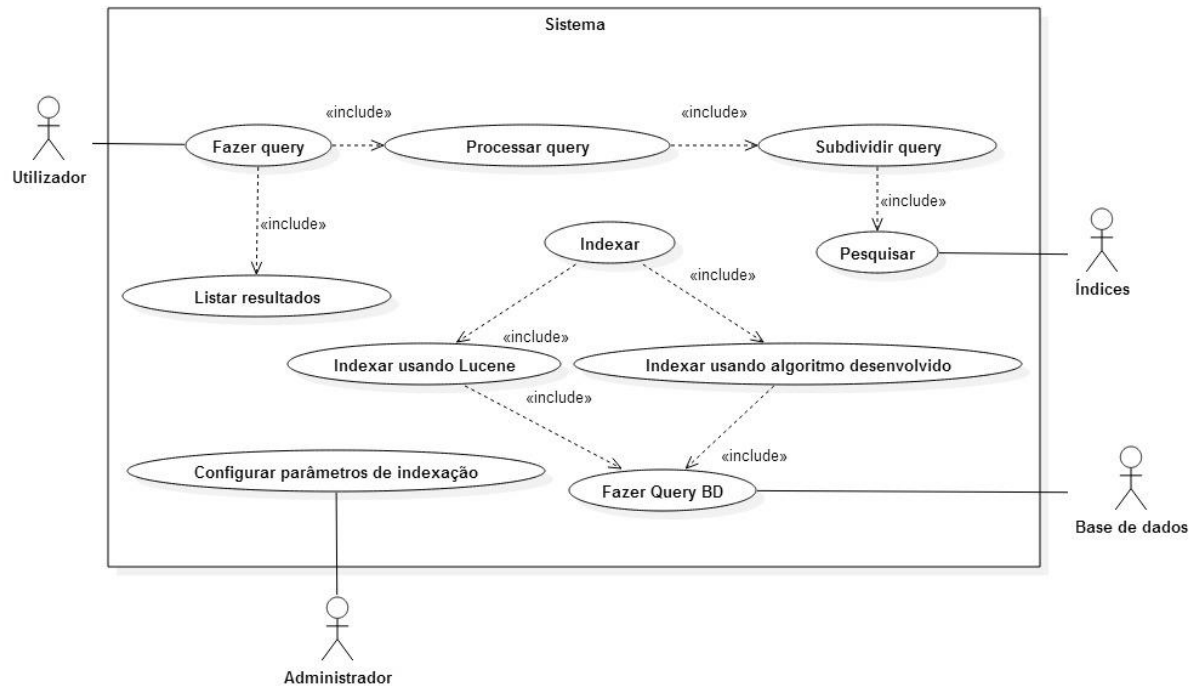
Anexo XIV – *Dataset* utilizado para treinar a rede neuronal que permite predizer as notas dos alunos (funcionalidade 9 da plataforma de análise de dados);

Anexo XV: *Dataset* utilizado para efetuar o clustering sobre os dados dos utilizadores e suas consultas (funcionalidade 10 da plataforma de análise de dados).

Anexo XVI – *Dataset* utilizado no sistema de recomendações;

Anexo I: Diagrama da Base de dados.



Anexo II: Diagrama de casos de uso do módulo do motor de busca.

Actores	Utilizador	Utilizador da plataforma. Este apenas faz a <i>query</i> (ainda que de um modo abstracto porque na prática não é ele que a faz mas sim o sistema baseando-se no título do artigo que este está de momento a consultar).
	Base de dados	Base de dados MySQL do projeto ADAPT que contém informação relativa aos conteúdos a pesquisar.
	Índices	Os índices são criados numa pasta local e utilizados para pesquisar.
	Administra dor	Administra a plataforma e configura os parâmetros de indexação para que esta possa ser efetuada periodicamente.
Casos de uso	Fazer <i>query</i>	É feita pelo utilizador tendo em conta o conteúdo que o utilizador está de momento a visualizar
	Listar resultados	Consiste na listagem de conteúdos e é incluído pelo caso de uso fazer pesquisa.
	Processar <i>query</i>	É incluído pelo caso de uso "Fazer <i>query</i> " e inclui ao caso de uso "subdivisão da <i>query</i> ". Este caso de uso limpa e prepara a <i>query</i> para ser subdividida.
	Subdivir da <i>query</i>	É incluída pelo processamento da <i>query</i> e na prática cria vários subsets de <i>queries</i> . Inclui o caso de uso "Pesquisa".
	Pesquisar	É incluída pelo caso de uso "subdivisão da <i>query</i> " e pesquisará no ator "índices" pela informação requerida.
	Indexar	Este caso de uso deriva nos casos de uso "indexação Lucene" e "indexação própria".
	Indexar usando Lucene	Este caso de uso é incluído pelo caso de uso "indexação" e indexa os parâmetros relativos ao Lucene. i.e.: o título e descrição de um conteúdo.

indexação usando algoritmo desenvolvido	Este caso de uso é incluído pelo caso de uso "indexação" e indexa os parâmetros relativos ao EA, à pontuação e tempo de navegação do OA.
Fazer <i>Query</i> BD	Este caso de uso é incluído pelos casos de uso "indexação Lucene" e "indexação própria" e interage directamente com o actor BD de modo a que possa inquirir a BD pelos dados necessários aos dois casos de uso relativos à indexação
Configurar parâmetros de indexação	Este caso de uso permite ao administrados configurar a periodicidade da indexação e onde vão ser guardados os índices do motor de busca.

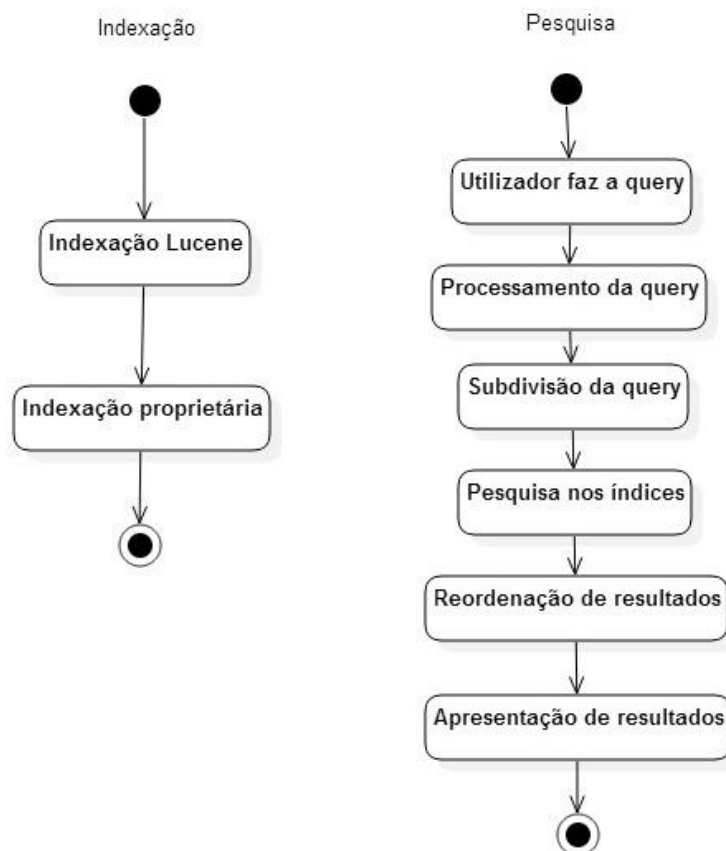
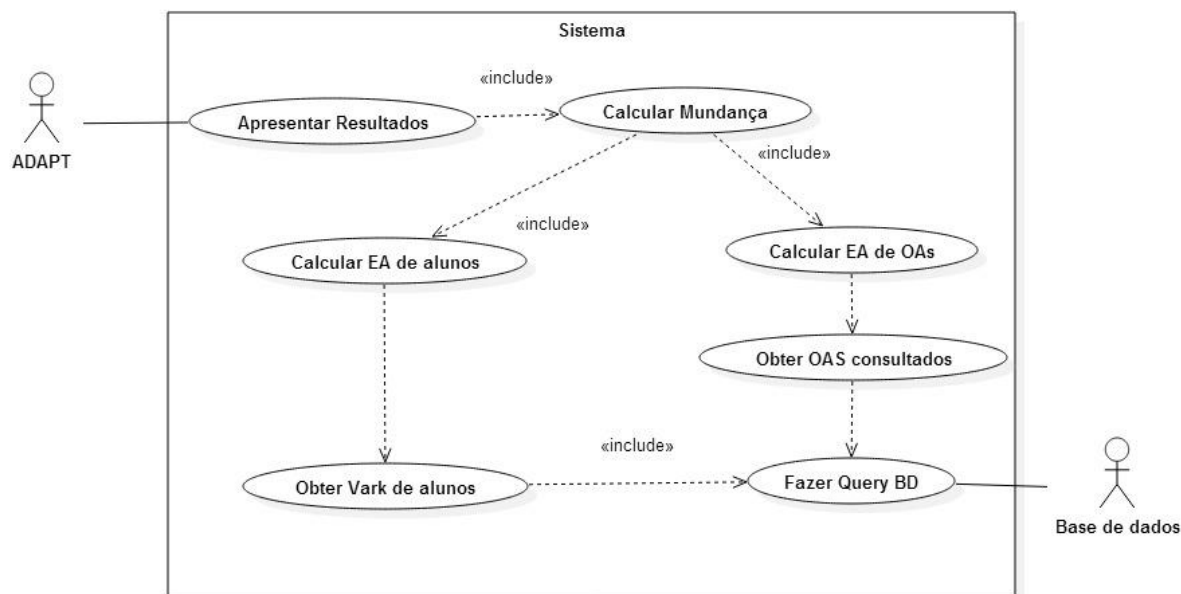
Anexo III: Diagrama de actividades do módulo do motor de busca.

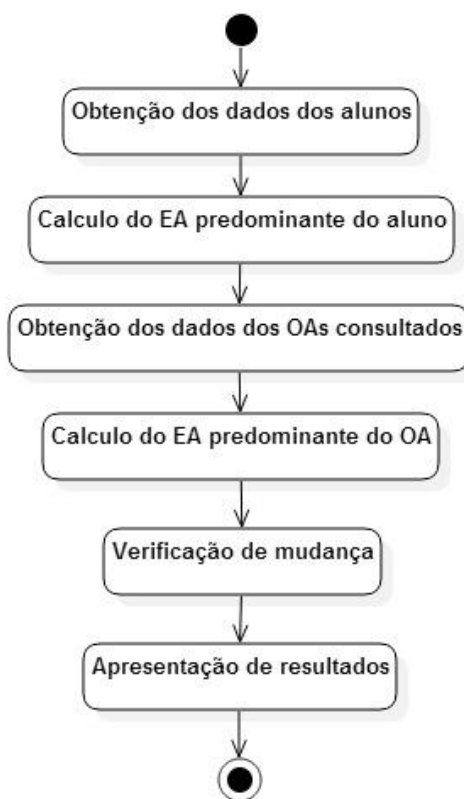
Diagrama A	Indexação Lucene	É efectuada a indexação do título e descrição dos conteúdos pelo Lucene.
	Indexação proprietária	É efectuada a indexação dos parâmetros relativos ao EA, à pontuação e tempo de navegação do OA.
Diagrama B	Utilizador faz a <i>query</i>	O utilizador faz uma <i>query</i> aos índices.
	Processamento da <i>query</i>	A <i>query</i> é processada, ou seja são removidos todas as palavras sem significado relevante, caracteres especiais etc.
	Subdivisão da <i>query</i>	A <i>query</i> é subdividida em vários subsets de <i>queries</i> .
	Pesquisa nos índices	É efetuada uma pesquisa nos índices.
	Apresentação de resultados	Os resultados são por fim apresentados.

Anexo IV: Diagrama de casos de uso do módulo do classificador de estilos de aprendizagem.



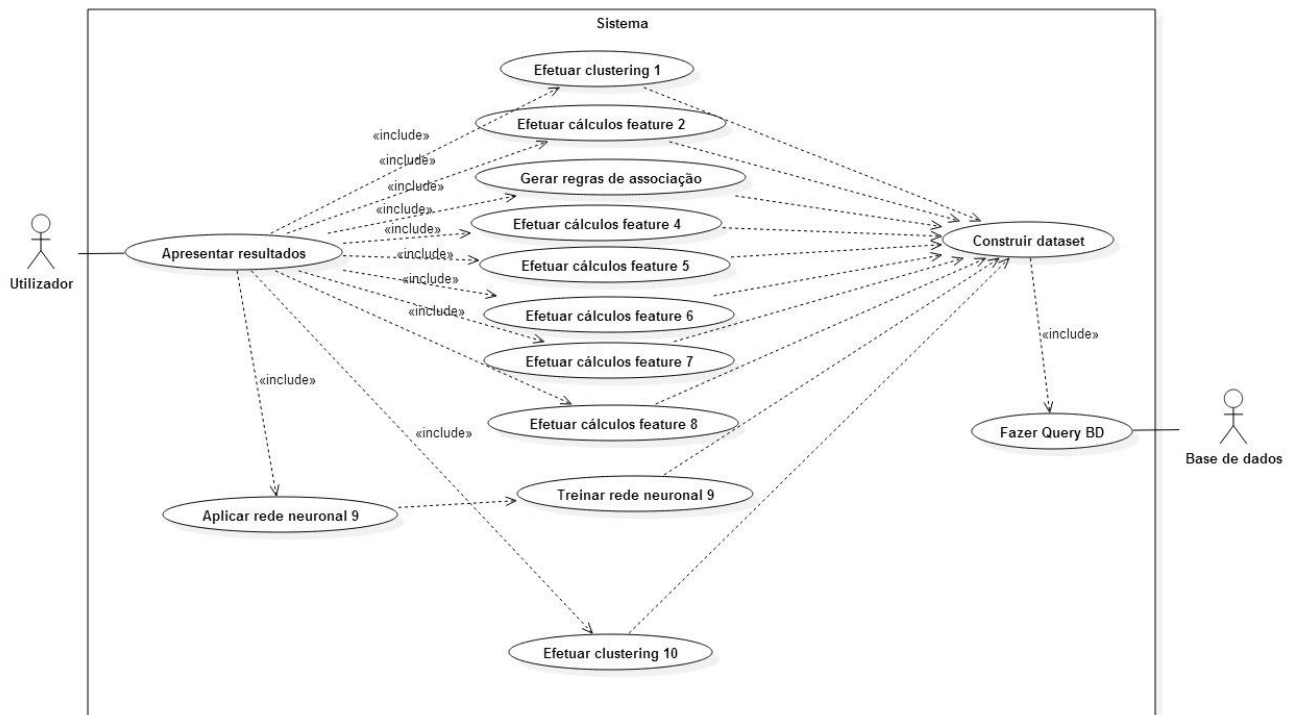
Actores	ADAPT	Chama e recebe os resultados do módulo.
	Base de dados	A base de dados MySQL do projeto é de onde vêm os dados utilizados pelo caso de uso <i>Query</i> para calcular as mudanças de estilo de aprendizagem
Casos de uso	Obter OAs consultados	Obtém as consultas que os alunos da plataforma têm vindo a fazer ao longo de X semanas.
	Obter Vark de alunos	Obtém as preferências Vark de cada aluno.
	Fazer <i>Query</i> BD	Faz as pesquisas à base de dados do projeto.
	Calcular EA de OAs	Com base nos parâmetros Vark de cada OA calcula o seu estilo de aprendizagem predominante.
	Calcular EA de alunos	Com base nos parâmetros Vark de cada aluno calcula o seu estilo de aprendizagem predominante.
	Calcular Mundaça	Comparando o EA do aluno com as contagens de OAs diferentes que têm vindo a consultar o sistema calcula se existe ou não mudança.
	Apresentar resultados	Apresenta os resultados ao módulo.

Anexo V: Diagrama de actividades do módulo do classificador de estilos de aprendizagem.



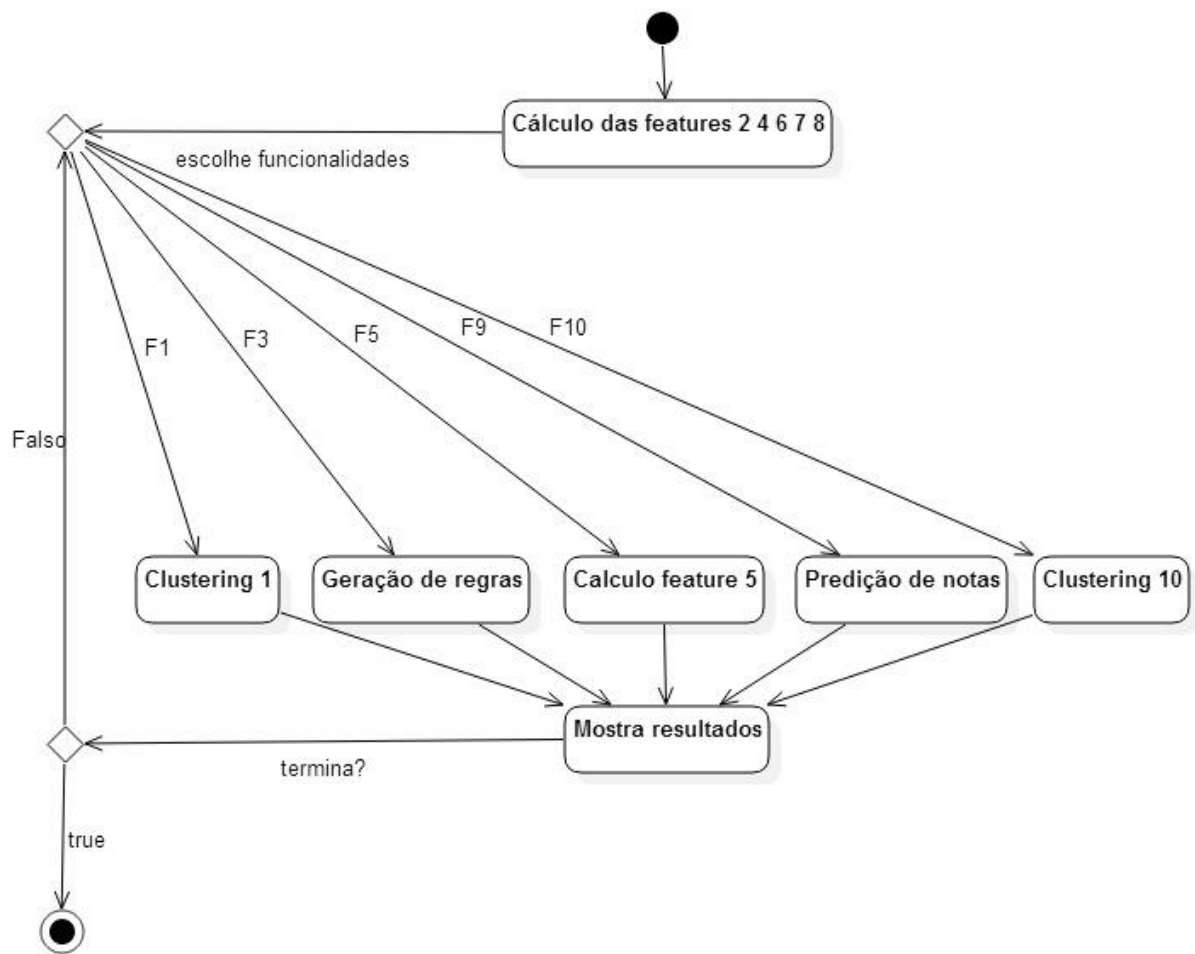
Obtenção dos dados dos alunos	Obtenção dos parâmetros vark dos alunos.
Cálculo do EA predominante do aluno	Cálculo do EA predominante do aluno.
Obtenção dos dados dos OAs consultados	Obtenção dos OAs consultados e dos seus parâmetros Vark.
Cálculo do EA predominante do OA	Cálculo do EA predominante do AO.
Verificação de mudança	Verifica se o estilo do aluno tem vindo a modificar consoante os conteúdos que tem consultado.
Apresentação de resultados	Apresentação dos resultados do módulo ao ADAPT.

Anexo VI: Diagrama de casos de uso da plataforma de análise de dados.

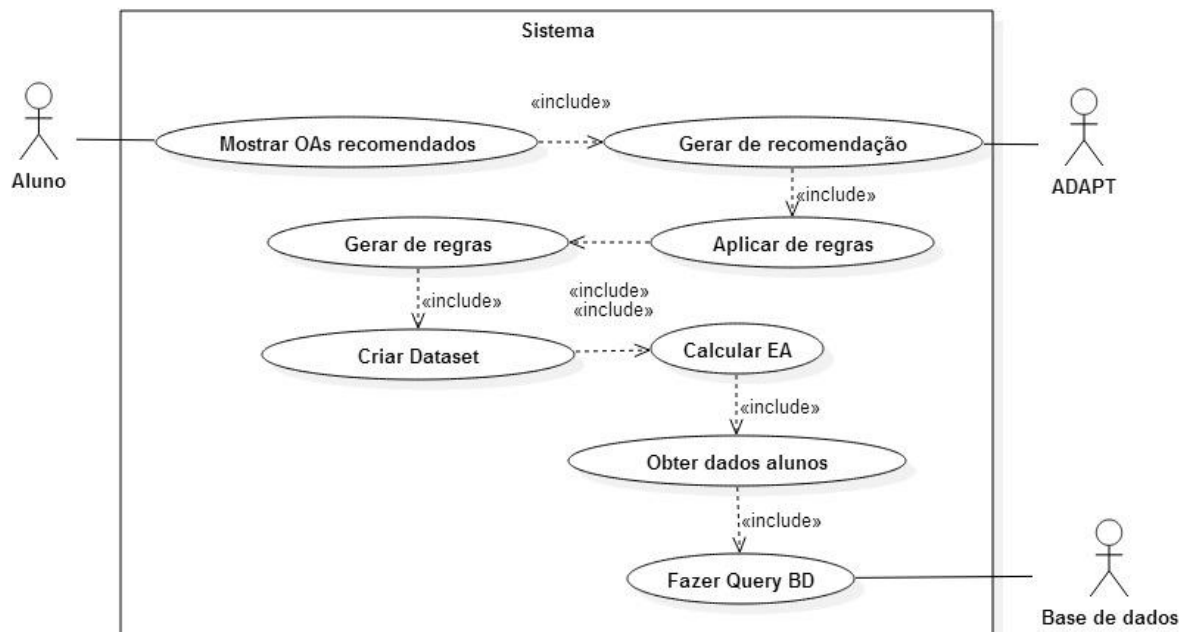


Atores	Utilizador	Chama e recebe os resultados do módulo.
	Base de dados	Base de dados mySQL do projeto. É aqui que se encontram os dados necessários às features deste módulo.
Casos de uso	Fazer Query BD	Pesquisa à BD do projeto pelos dados necessários às features implementadas.
	Construir dataset	Construção dos datasets de todas as features.
	Efetuar clustering 1	Clustering sobre os dados dos alunos (feature 1).
	Efetuar cálculos feature 2	Cálculo da média das notas por estilo de aprendizagem .
	Gerar regras de associação	Extração de regras através dos dados produzidos pela utilização da plataforma pelos alunos (feature 3).
	Efetuar cálculos feature 4	Cálculo da percentagem de visualização de OAs por estilo de aprendizagem.
	Efetuar cálculos feature 5	Cálculo da percentagem de mudanças de estilo de aprendizagem.
	Efetuar cálculos feature 6	Cálculo da média das notas e predominância de estilos por disciplina e curso.

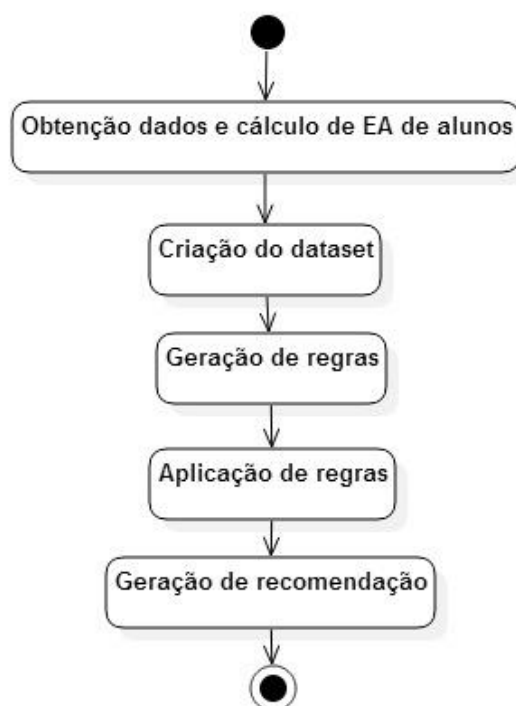
Efetuar cálculos feature 7	Cálculo da distribuição dos alunos por estilo de aprendizagem.
Efetuar cálculos feature 8	Cálculo das notas (discretizadas) por OA, por disciplina.
Treinar rede neuronal 9	Treino da rede neuronal necessária para predizer as notas dos alunos com base nas suas consultas e consultas efetuadas pelos seus colegas avaliados. (Feature 9).
Aplicar rede neuronal 9	Aplicação da rede neuronal a fim de predizer as notas dos alunos (Feature 9).
Efetuar clustering 10	Clustering de dados sobre os dados dos alunos e as consultas que efetuou (feature 10).
Apresentar resultados	Apresentação de resultados ao utilizador.

Anexo VII: Diagrama de actividades da plataforma de análise de dados-

Cálculo das features 2, 4, 6, 7, 8	Estas features são corridas automaticamente carregadas.
Clustering 1	Clustering sobre os dados dos alunos (feature 1).
Geração de regras	Extração de regras através dos dados produzidos pela utilização da plataforma pelos alunos (feature 3).
Calculo feature 5	Cálculo da percentagem de mudanças de estilo de aprendizagem.
Predição de notas	Aplicação da rede neuronal a fim de prever as notas dos alunos (Feature 9).
Clustering 10	Clustering de dados sobre os dados dos alunos e as consultas que efetuou (feature 10).
Mostra resultados	Apresentação dos resultados ao utilizador.

Anexo VIII: Diagrama de casos de uso do módulo do sistema de recomendações

Actores	Utilizador	Interage com o módulo apenas visualizando os seus resultados.
	Base de dados	A base de dados mySQL do projeto é de onde vêm os dados utilizados pelo caso de uso <i>Query</i> para calcular as mudanças de estilo de aprendizagem.
	Aluno	Apenas visualiza a recomendação que lhe é proposta.
casos de uso	Obter dados alunos	Este caso de uso obtém os dados dos alunos, especificamente os seus parâmetros <i>vark</i> e disciplinas às quais estão inscritos.
	Fazer <i>Query</i> BD	Obtenção de dados directa da Base de dados myQSL do projecto.
	Calcular EA	Este caso de uso calcula os estilos de aprendizagem dos alunos.
	Criar <i>Dataset</i>	Este caso de uso verifica quais os utilizadores que têm o mesmo estilo de aprendizagem e estão inscritos às mesmas disciplinas que o aluno. Com base nisto, cria um <i>dataset</i> que contém consultas de alunos inscritos às mesmas disciplinas e com o mesmo estilo de aprendizagem.
	Gerar de regras	Geração de regras de associação (<i>Apriori</i>) com base no <i>dataset</i> criado.
	Aplicar de regras	Neste caso de uso é verificado se as consultas do aluno "encaixam" nas regras geradas. Neste caso de uso encontram-se OAs que tiveram interesse para alunos que consultaram os mesmos OAs que o aluno a recomendar.
	Gerar de recomendação	É gerada uma lista de OAs a apresentar ao aluno e é apresentada.
	Mostrar OAs recomendados	É mostrada a listagem de OAs a recomendar.

Anexo IX: Diagrama de actividades do módulo do sistema de recomendações

Obtenção dados e cálculo de EA de alunos	Este caso de uso obtém os dados dos alunos, especificamente os seus parâmetros <i>var</i> e disciplinas às quais estão inscritos. O seu EA é depois calculado.
Criação do <i>dataset</i>	Este caso de uso verifica quais os utilizadores que têm o mesmo estilo de aprendizagem e estão inscritos às mesmas disciplinas que o aluno. Com base nisto, cria um <i>dataset</i> que contém consultas de alunos inscritos às mesmas disciplinas e com o mesmo estilo de aprendizagem.
Geração de regras	Geração de regras de associação (<i>Apriori</i>) com base no <i>dataset</i> criado.
Aplicação de regras	Neste caso de uso é verificado se as consultas do aluno "encaixam" nas regras geradas. Neste caso de uso encontram-se OAs que tiveram interesse para alunos que consultaram os mesmos OAs que o aluno a recomendar.
Geração de recomendação	É gerada uma lista de OAs a apresentar ao aluno e é apresentada.

Anexo X: Dataset utilizado nos testes de performance do motor de busca.

Id OA	Título	Descrição	Palavra-chave
100	sistemas analogicos swf	sistemas analogicos sao os sistemas antecessores aos sistemas digitais . este tipo de sistemas baseia-se num sinal	analogicos
101	sistemas Analogicos pdf	sistemas analogicos sao os sistemas antecessores aos sistemas digitais . este tipo de sistemas baseia-se num sinal	analogicos
102	sistemas digitais swf	Os sistemas digitais sao hoje em dia o tipo de sistemas utilizado em praticamente tudo . tornam-se simples de implementar	digitais
103	sistemas digitais pdf	Os sistemas digitais sao hoje em dia o tipo de sistemas utilizado em praticamente tudo . tornam-se simples de implementar	digitais
104	sistemas digitais binarios swf	Os sistemas digitais sao hoje em dia o tipo de sistemas utilizado em praticamente tudo. tornam-se simples de implementar,	digitais
105	sistemas digitais binarios pdf	Os sistemas digitais sao hoje em dia o tipo de sistemas utilizado em praticamente tudo. tornam-se simples de implementar,	digitais
107	logica positiva e logica negativa pdf	A logica positiva ou negativa foi algo inventado para interpretar os valores dos circuitos integrados digitais . A sua	logica
106	logica positiva e logica negativa swf	A logica positiva ou negativa foi algo inventado para interpretar os valores dos circuitos integrados digitais . A sua	logica
109	dispositivos digitais pdf	tipos de circuitos integrados existentes	componentes
110	circuitos combinacionais cswf	circuitos capazes de combinar varias variaveis num so output . and , or , xor sao exemplos de circuitos combinacionais	componentes
108	dispositivos digitais swf	tipos de circuitos integrados existentes	componentes
111	circuitos combinacionais pdf	circuitos capazes de combinar varias variaveis num so output . and , or , xor sao exemplos de circuitos combinacionais	componentes
112	variavel logica swf	variaveis que tomam normalmente o valor de 0 ou 1 e que sao necessarias como input de circuitos combinacionais ou outros	logica
113	variavel logica pdf	variaveis que tomam normalmente o valor de 0 ou 1 e que sao necessarias como input de circuitos combinacionais ou outros	logica

114	funcao logica swf	funcao ou equacao algebrica - algebra de boole - que associa varias variaveis logicas com varios estados a fim de	logica
115	funcao logica pdf	funcao ou equacao algebrica - algebra de boole - que associa varias variaveis logicas com varios estados a fim de	logica
116	funcao logica negacao ou inversao swf	funcao ou equacao algebrica - algebra de boole - que associa varias variaveis logicas com varios estados a fim de	logica
117	funcao logica negacao ou inversao pdf	funcao ou equacao algebrica - algebra de boole - que associa varias variaveis logicas com varios estados a fim de	logica
118	funcao logica intersecao ou produto logico swf	funcao ou equacao algebrica - algebra de boole - que associa varias variaveis logicas com varios estados a fim de	logica
119	funcao logica intersecao ou produto logico pdf	funcao ou equacao algebrica - algebra de boole - que associa varias variaveis logicas com varios estados a fim de	logica
120	funcao logica reuniao ou soma logica swf	funcao ou equacao algebrica - algebra de boole - que associa varias variaveis logicas com varios estados a fim de	logica
121	funcao logica reuniao ou soma logica pdf	funcao ou equacao algebrica - algebra de boole - que associa varias variaveis logicas com varios estados a fim de	logica
122	expressoes logicas swf	equacoes algebricas - algebra de boole - que associa varias variaveis logicas com varios estados a fim de produzir um	logica
123	expressoes logicas pdf	equacoes algebricas - algebra de boole - que associa varias variaveis logicas com varios estados a fim de produzir um	logica
124	expressoes logicas expressoes equivalentes e expressoes	equacoes algebricas - algebra de boole - que associa varias variaveis logicas com varios estados a fim de produzir um	logica

	complementares swf		
125	expressões lógicas complementares pdf	equações algébricas - álgebra de boole - que associa várias variáveis lógicas com vários estados a fim de produzir um	lógica
126	expressões lógicas complementares swf	equações algébricas - álgebra de boole - que associa várias variáveis lógicas com vários estados a fim de produzir um	lógica
127	expressões lógicas complementares pdf	equações algébricas - álgebra de boole - que associa várias variáveis lógicas com vários estados a fim de produzir um	lógica
128	expressões lógicas soma de produtos e produto de somas swf	equações algébricas - álgebra de boole - que associa várias variáveis lógicas com vários estados a fim de produzir um	lógica
129	expressões lógicas soma de produtos e produto de somas pdf	equações algébricas - álgebra de boole - que associa várias variáveis lógicas com vários estados a fim de produzir um	lógica
130	expressões lógicas fcsp e fcps swf	equações algébricas - álgebra de boole - que associa várias variáveis lógicas com vários estados a fim de produzir um	lógica
131	expressões lógicas fcsp e fcps pdf	equações algébricas - álgebra de boole - que associa várias variáveis lógicas com vários estados a fim de produzir um	lógica
132	expressões lógicas fcsp e fcps swf	equações algébricas - álgebra de boole - que associa várias variáveis lógicas com vários estados a fim de produzir um	lógica
133	expressões lógicas fcsp e fcps pdf	equações algébricas - álgebra de boole - que associa várias variáveis lógicas com vários estados a fim de produzir um	lógica

134	postulados swf	postulados	logica
135	postulados pdf	postulados	logica
137	simplificac ao de funcoes pelos teoremas pdf	Os teoremas facilitam a resolucao das expressoes algebricas	logica
136	simplificac ao de funcoes pelos teoremas swf	Os teoremas facilitam a resolucao das expressoes algebricas	logica
139	tabelas de verdade fcsp e fcps swf pdf	Representacao do output da funcao logica	mapas de karnaugh
138	tabelas de verdade fcsp e fcps swf	Representacao do output da funcao logica	mapas de karnaugh
140	mapas de karnaugh swf	Representacao do output da funcao logica	mapas de karnaugh
141	mapas de karnaugh pdf	Representacao do output da funcao logica	mapas de karnaugh
142	mapas de karnaugh tabelas de verdade swf	Representacao do output da funcao logica	mapas de karnaugh
143	mapas de karnaugh tabelas de verdade pdf	Representacao do output da funcao logica	mapas de karnaugh
144	mapas de karnaugh grupos de adjacencia swf	Representacao do output da funcao logica	mapas de karnaugh
145	mapas de karnaugh grupos de	Representacao do output da funcao logica	mapas de karnaugh

	adjacencia pdf		
146	mapas de karnaugh fmsp swf	Representacao do output da funcao logica	mapas de karnaugh
147	mapas de karnaugh fmsp pdf	Representacao do output da funcao logica	mapas de karnaugh
148	multiplexers swf	Dispositivo que codifica as informacoes de duas ou mais fontes de dados num único canal . Sao utilizados em situações onde o custo de implementacao de canais separados para cada fonte de dados e maior que o custo e a inconveniencia de utilizar as funcoes de multiplexacao/demultiplexacao	componentes
149	multiplexers pdf	Dispositivo que codifica as informacoes de duas ou mais fontes de dados num único canal . Sao utilizados em situações onde o custo de implementacao de canais separados para cada fonte de dados e maior que o custo e a inconveniencia de utilizar as funcoes de multiplexacao/demultiplexacao	componentes
151	descodificadores swf	circuito combinacional que tem o papel inverso do codificador (eletronica), isto e , converte um código binario de entrada (natural , BCD ,...) de N bits de entrada em M linha de saída (em que N pode ser qualquer inteiro e M é um inteiro menor ou igual a $2N$), de modo que cada linha de saida sera ativada por uma unica combinacao das possiveis de entrada .	componentes
152	descodificadores pdf	circuito combinacional que tem o papel inverso do codificador (eletronica), isto e , converte um código binario de entrada (natural , BCD ,...) de N bits de entrada em M linha de saída (em que N pode ser qualquer inteiro e M é um inteiro menor ou igual a $2N$), de modo que cada linha de saida sera ativada por uma unica combinacao das possiveis de entrada .	componentes
153	codificadores swf	O codificador e um circuito logico que , como o próprio nome o diz , codifica um sinal que se encontra em uma forma para outra forma , usando um tipo de codigo .	componentes
154	codificadores pdf	O codificador e um circuito logico que , como o próprio nome o diz , codifica um sinal que se encontra em uma forma para outra forma , usando um tipo de codigo .	componentes
155	demultiplexers swf	Dispositivo que executa a operacao inversa do multiplexer , isto e , distribui informacoes de uma unica entrada para uma das diversas saidas . Um demultiplexer e o contrario de um multiplexer .	componentes
156	demultiplexers pdf	Dispositivo que executa a operacao inversa do multiplexer , isto e , distribui informacoes de uma unica entrada para uma das diversas saidas . Um demultiplexer e o contrario de um multiplexer .	componentes
157	comparadores swf	Os circuitos comparadores sao muito utilizados em eletronica por permitirem a construcao de diversos circuitos eletronicos , como protecoes de fontes , conversores Analogico digital e outros .	componentes

158	comparadores pdf	Os circuitos comparadores são muito utilizados em eletrônica por permitirem a construção de diversos circuitos eletrônicos, como proteções de fontes, conversores Analógico digital e outros.	componentes
159	somadores swf	adição Binária e a operação mais simples de Circuito Aritmético e a de adição. A adição binária é executada da mesma forma que a decimal inclusive no "Vai - Um" (Carry Out, de saída, e Carry In, de entrada). Quando somamos dois números binário começamos pela coluna menos significativa (que representa unidade entre centena, dezena, etc.).	componentes
160	somadores pdf	adição Binária e a operação mais simples de Circuito Aritmético e a de adição. A adição binária é executada da mesma forma que a decimal inclusive no "Vai - Um" (Carry Out, de saída, e Carry In, de entrada). Quando somamos dois números binário começamos pela coluna menos significativa (que representa unidade entre centena, dezena, etc.).	componentes
161	mapas karnaugh	output da função lógica	mapas de karnaugh
162	tabelas de verdade	Representação do saída da função lógica	mapas de karnaugh
163	mapas de karnaugh e tabelas de verdade	Representação do output	mapas de karnaugh
164	mapas de karnaugh grupos de adjacência	Representação variáveis	mapas de karnaugh
165	grupos de adjacência pdf	grupos de adjacência dos mapas de karnaugh	mapas de karnaugh
166	mapas de karnaugh fmsf pdf	Representação do output da função lógica	mapas de karnaugh
167	sistemas digitais swf	Os sistemas digitais são mais rápidos e simples de implementar	digitais
168	digitais pdf	Os sistemas digitais são hoje em dia o tipo de sistemas utilizado em praticamente tudo	digitais
169	binários swf	Tipo de sistemas mais utilizado: digitais	digitais
170	binários pdf	Os sistemas digitais binários pdf	digitais
171	digitais swf	Os sistemas digitais binários swf	digitais
172	programação 1	Os sistemas digitais são hoje em dia o tipo de sistemas utilizado em praticamente tudo. tornam-se simples de implementar	digitais
173	programação 2	Os sistemas digitais são representados por saídas binárias 1 ou 0	digitais

174	programação 3	video sobre Os sistemas digitais: Sistemas cuja saída é 0 ou 1	digitais
175	programação 4	introdução sobre sistemas digitais	digitais
176	programação 5	sistemas digitais iniciantes	digitais
177	programação 6	sistemas digitais intermediários	digitais
178	programação 7	sistemas digitais avançados	digitais

Anexo XI: *Dataset* utilizado nos testes de performance do classificador de estilos de aprendizagem.

IDUSER	VISUAL	AURAL	READWRITE	KINESTHETIC	TempoPermanecidoNoLink	BLOOM
13	80.0	0.0	10.0	10.0	7205	1
12	80.0	0.0	10.0	10.0	14401	1
15	80.0	0.0	10.0	10.0	7200	3
15	80.0	0.0	10.0	10.0	3605	3
16	80.0	0.0	10.0	10.0	130	3
17	80.0	0.0	10.0	10.0	1080	3
12	75.0	0.0	25.0	0.0	188	4
18	75.0	0.0	25.0	0.0	1808	4
19	75.0	0.0	25.0	0.0	4988	4
15	75.0	0.0	25.0	0.0	1688	4
19	75.0	0.0	25.0	0.0	1920	4
15	70.0	0.0	50.0	15.0	1993	5
14	60.0	0.0	40.0	0.0	3600	6
16	60.0	0.0	40.0	0.0	720	6
20	50.0	0.0	50.0	0.0	1080	5
18	80.0	0.0	10.0	90.0	2503	1
19	80.0	0.0	10.0	90.0	360	1
21	60.0	0.0	40.0	0.0	29400	5
20	80.0	0.0	60.0	20.0	30602	4
12	60.0	0.0	70.0	10.0	14400	2
12	60.0	0.0	70.0	30.0	272	5
30	55.0	0.0	60.0	0.0	30780	4
36	90.0	0.0	10.0	70.0	30780	5
35	75.0	0.0	35.0	0.0	1680	2
21	90.0	0.0	10.0	70.0	30780	3
35	80.0	0.0	30.0	75.0	30602	2
34	90.0	0.0	10.0	0.0	29040	2
32	90.0	0.0	10.0	0.0	30780	2
26	90.0	0.0	10.0	0.0	29702	2
13	60.0	0.0	50.0	65.0	25201	1
31	60.0	0.0	50.0	0.0	29100	5
35	60.0	0.0	50.0	60.0	29222	6
32	60.0	0.0	50.0	60.0	29400	2
19	60.0	0.0	65.0	0.0	30602	1
20	45.0	0.0	85.0	0.0	30780	1
36	80.0	0.0	15.0	85.0	40082	2
21	75.0	0.0	65.0	0.0	1260	5
23	75.0	0.0	65.0	0.0	1980	5

35	75.0	0.0	65.0	0.0	633	5
22	75.0	0.0	65.0	0.0	30602	5
21	80.0	0.0	15.0	80.0	1980	6
24	80.0	0.0	15.0	80.0	1260	6
31	80.0	0.0	15.0	80.0	453	6
28	80.0	0.0	15.0	80.0	30780	6
23	80.0	0.0	15.0	80.0	30602	6
23	60.0	0.0	50.0	0.0	1080	3
26	60.0	0.0	50.0	0.0	2160	3
28	60.0	0.0	50.0	0.0	2228	3
34	60.0	0.0	50.0	0.0	1408	3
36	60.0	0.0	50.0	0.0	633	3
21	60.0	0.0	50.0	0.0	29222	3
24	60.0	0.0	50.0	0.0	30602	3
25	80.0	0.0	15.0	85.0	1260	4
29	80.0	0.0	15.0	85.0	1800	4
19	80.0	0.0	15.0	85.0	30780	4
20	30.0	0.0	80.0	0.0	453	3
32	30.0	0.0	80.0	0.0	1408	3
30	80.0	0.0	15.0	85.0	633	2
33	80.0	0.0	15.0	85.0	1350	2
33	80.0	0.0	15.0	85.0	30302	2
28	25.0	0.0	85.0	0.0	30302	1
22	85.0	0.0	10.0	90.0	30602	2
36	80.0	0.0	50.0	0.0	30602	3
36	85.0	0.0	16.0	92.0	29040	6
19	85.0	0.0	16.0	92.0	29400	6
23	85.0	0.0	16.0	92.0	30302	6
12	85.0	0.0	10.0	0.0	210	1
12	85.0	0.0	10.0	0.0	3651	1
18	80.0	0.0	10.0	85.0	7201	4
34	80.0	0.0	10.0	85.0	30602	4
18	85.0	0.0	20.0	0.0	7211	3
18	85.0	0.0	20.0	0.0	349209	3
12	85.0	0.0	20.0	0.0	561	3
25	80.0	0.0	20.0	85.0	29702	5
22	85.0	0.0	20.0	0.0	29282	6
32	75.0	0.0	65.0	0.0	30602	3
33	60.0	0.0	50.0	0.0	29222	1
19	30.0	0.0	80.0	0.0	30602	5
31	80.0	0.0	10.0	10.0	8553	3
32	80.0	0.0	10.0	10.0	12573	3

20	80.0	0.0	10.0	10.0	9753	3
21	80.0	0.0	10.0	10.0	12453	3
20	80.0	0.0	10.0	10.0	8553	3
17	80.0	0.0	10.0	10.0	186	3
22	75.0	0.0	0.0	10.0	12573	3
23	75.0	0.0	0.0	10.0	9753	3
24	75.0	0.0	0.0	10.0	12453	3
25	75.0	0.0	0.0	10.0	8553	3
18	75.0	0.0	0.0	10.0	3780	3
23	75.0	0.0	0.0	10.0	480	3
26	50.0	0.0	90.0	50.0	17553	3
28	50.0	0.0	90.0	50.0	17673	3
29	50.0	0.0	90.0	50.0	12453	3
15	50.0	0.0	90.0	50.0	12453	3
19	50.0	0.0	90.0	50.0	7200	3
19	50.0	0.0	90.0	50.0	1200	3
17	90.0	0.0	60.0	0.0	8553	5
19	90.0	0.0	60.0	0.0	8553	5
18	90.0	0.0	60.0	0.0	3600	5
16	90.0	0.0	60.0	0.0	3780	5
15	90.0	0.0	60.0	0.0	7200	5
16	90.0	0.0	60.0	0.0	4080	5
16	90.0	0.0	60.0	0.0	1800	5
16	90.0	0.0	60.0	0.0	900	5
25	15.0	0.0	90.0	50.0	4200	1
26	15.0	0.0	90.0	50.0	7800	1
28	15.0	0.0	90.0	50.0	1986	1
34	15.0	0.0	90.0	50.0	660	1
24	50.0	50.0	0.0	0.0	900	4
23	50.0	50.0	0.0	0.0	4500	4
25	50.0	50.0	0.0	0.0	5700	4
22	50.0	50.0	0.0	0.0	5100	4
26	50.0	50.0	0.0	0.0	1380	4
30	50.0	50.0	0.0	0.0	29702	4
32	50.0	50.0	0.0	0.0	29282	4
31	50.0	50.0	0.0	0.0	29702	4
23	50.0	0.0	30.0	20.0	5700	6
36	50.0	0.0	30.0	20.0	4500	6
25	50.0	0.0	30.0	20.0	4740	6
26	50.0	0.0	30.0	20.0	4500	6
26	50.0	0.0	30.0	20.0	1986	6
33	50.0	0.0	30.0	20.0	1080	6

19	50.0	0.0	30.0	20.0	29282	6
24	50.0	0.0	30.0	20.0	29702	6
17	30.0	0.0	20.0	50.0	4500	3
34	30.0	0.0	20.0	50.0	4740	3
31	30.0	0.0	20.0	50.0	4500	3
22	30.0	0.0	20.0	50.0	5100	3
33	30.0	0.0	20.0	50.0	4560	3
24	30.0	0.0	20.0	50.0	4740	3
25	30.0	0.0	20.0	50.0	5400	3
26	30.0	0.0	20.0	50.0	4080	3
25	30.0	0.0	20.0	50.0	1206	3
32	30.0	0.0	20.0	50.0	780	3
31	30.0	0.0	20.0	50.0	29222	3
23	30.0	0.0	20.0	50.0	30302	3
16	0.0	99.9	0.0	0.0	10806	5
16	0.0	99.9	0.0	0.0	7200	5
32	0.0	99.9	0.0	0.0	8700	5
16	0.0	99.9	0.0	0.0	5100	5
21	0.0	99.9	0.0	0.0	420	5
23	0.0	99.9	0.0	0.0	2106	5
30	0.0	99.9	0.0	0.0	36360	5
26	0.0	99.9	0.0	0.0	36482	5
29	0.0	99.9	0.0	0.0	36482	5
20	0.0	99.9	0.0	0.0	29282	5
21	0.0	99.9	0.0	0.0	35222	5
27	0.0	99.9	0.0	0.0	31572482	5
27	0.0	99.9	0.0	0.0	31572482	5
15	75.0	0.0	0.0	25.0	2106	6
22	75.0	0.0	0.0	25.0	8406	6
24	75.0	0.0	0.0	25.0	5706	6
22	75.0	0.0	0.0	25.0	1986	6
34	75.0	0.0	0.0	25.0	31860	6
35	75.0	0.0	0.0	25.0	29282	6
23	35.0	0.0	30.0	35.0	1206	5
25	35.0	0.0	30.0	35.0	1746	5
26	35.0	0.0	30.0	35.0	1206	5
31	35.0	0.0	30.0	35.0	1206	5
33	35.0	0.0	30.0	35.0	2106	5
32	35.0	0.0	30.0	35.0	1206	5
34	35.0	0.0	30.0	35.0	2106	5
24	35.0	0.0	30.0	35.0	726	5
21	35.0	0.0	30.0	35.0	29400	5

21	35.0	0.0	30.0	35.0	29102	5
35	0.0	90.0	0.0	10.0	1986	4
36	0.0	90.0	0.0	10.0	1986	4
32	0.0	90.0	0.0	10.0	1986	4
31	0.0	90.0	0.0	10.0	2106	4
25	0.0	90.0	0.0	10.0	726	4
24	0.0	90.0	0.0	10.0	1206	4
29	0.0	90.0	0.0	10.0	1206	4
21	0.0	90.0	0.0	10.0	1206	4
30	0.0	90.0	0.0	10.0	37860	4
24	0.0	90.0	0.0	10.0	29702	4
33	0.0	90.0	0.0	10.0	29282	4
32	0.0	90.0	0.0	10.0	29282	4
32	0.0	90.0	0.0	10.0	36482	4
13	80.0	0.0	10.0	10.0	7205	1
12	80.0	0.0	10.0	10.0	14401	1
15	80.0	0.0	10.0	10.0	7200	3
15	80.0	0.0	10.0	10.0	3605	3
16	80.0	0.0	10.0	10.0	130	3
17	80.0	0.0	10.0	10.0	1080	3
12	75.0	0.0	25.0	0.0	188	4
18	75.0	0.0	25.0	0.0	1808	4
19	75.0	0.0	25.0	0.0	4988	4
15	75.0	0.0	25.0	0.0	1688	4
19	75.0	0.0	25.0	0.0	1920	4
15	70.0	0.0	50.0	15.0	1993	5
14	60.0	0.0	40.0	0.0	3600	6
16	60.0	0.0	40.0	0.0	720	6
20	50.0	0.0	50.0	0.0	1080	5
18	80.0	0.0	10.0	90.0	2503	1
19	80.0	0.0	10.0	90.0	360	1
21	60.0	0.0	40.0	0.0	29400	5
20	80.0	0.0	60.0	20.0	30602	4
12	60.0	0.0	70.0	10.0	14400	2
12	60.0	0.0	70.0	30.0	272	5
30	55.0	0.0	60.0	0.0	30780	4
36	90.0	0.0	10.0	70.0	30780	5
35	75.0	0.0	35.0	0.0	1680	2
21	90.0	0.0	10.0	70.0	30780	3
35	80.0	0.0	30.0	75.0	30602	2
34	90.0	0.0	10.0	0.0	29040	2
32	90.0	0.0	10.0	0.0	30780	2

26	90.0	0.0	10.0	0.0	29702	2
13	60.0	0.0	50.0	65.0	25201	1
31	60.0	0.0	50.0	0.0	29100	5
35	60.0	0.0	50.0	60.0	29222	6
32	60.0	0.0	50.0	60.0	29400	2
19	60.0	0.0	65.0	0.0	30602	1
20	45.0	0.0	85.0	0.0	30780	1
36	80.0	0.0	15.0	85.0	40082	2
21	75.0	0.0	65.0	0.0	1260	5
23	75.0	0.0	65.0	0.0	1980	5
35	75.0	0.0	65.0	0.0	633	5
22	75.0	0.0	65.0	0.0	30602	5
21	80.0	0.0	15.0	80.0	1980	6
24	80.0	0.0	15.0	80.0	1260	6
31	80.0	0.0	15.0	80.0	453	6
28	80.0	0.0	15.0	80.0	30780	6
23	80.0	0.0	15.0	80.0	30602	6
23	60.0	0.0	50.0	0.0	1080	3
26	60.0	0.0	50.0	0.0	2160	3
28	60.0	0.0	50.0	0.0	2228	3
34	60.0	0.0	50.0	0.0	1408	3
36	60.0	0.0	50.0	0.0	633	3
21	60.0	0.0	50.0	0.0	29222	3
24	60.0	0.0	50.0	0.0	30602	3
25	80.0	0.0	15.0	85.0	1260	4
29	80.0	0.0	15.0	85.0	1800	4
19	80.0	0.0	15.0	85.0	30780	4
20	30.0	0.0	80.0	0.0	453	3
32	30.0	0.0	80.0	0.0	1408	3
30	80.0	0.0	15.0	85.0	633	2
33	80.0	0.0	15.0	85.0	1350	2
33	80.0	0.0	15.0	85.0	30302	2
28	25.0	0.0	85.0	0.0	30302	1
22	85.0	0.0	10.0	90.0	30602	2
36	80.0	0.0	50.0	0.0	30602	3
36	85.0	0.0	16.0	92.0	29040	6
19	85.0	0.0	16.0	92.0	29400	6
23	85.0	0.0	16.0	92.0	30302	6
12	85.0	0.0	10.0	0.0	210	1
12	85.0	0.0	10.0	0.0	3651	1
18	80.0	0.0	10.0	85.0	7201	4
34	80.0	0.0	10.0	85.0	30602	4

18	85.0	0.0	20.0	0.0	7211	3
18	85.0	0.0	20.0	0.0	349209	3
12	85.0	0.0	20.0	0.0	561	3
25	80.0	0.0	20.0	85.0	29702	5
22	85.0	0.0	20.0	0.0	29282	6
32	75.0	0.0	65.0	0.0	30602	3
33	60.0	0.0	50.0	0.0	29222	1
19	30.0	0.0	80.0	0.0	30602	5
31	80.0	0.0	10.0	10.0	8553	3
32	80.0	0.0	10.0	10.0	12573	3
20	80.0	0.0	10.0	10.0	9753	3
21	80.0	0.0	10.0	10.0	12453	3
20	80.0	0.0	10.0	10.0	8553	3
17	80.0	0.0	10.0	10.0	186	3
22	75.0	0.0	0.0	10.0	12573	3
23	75.0	0.0	0.0	10.0	9753	3
24	75.0	0.0	0.0	10.0	12453	3
25	75.0	0.0	0.0	10.0	8553	3
18	75.0	0.0	0.0	10.0	3780	3
23	75.0	0.0	0.0	10.0	480	3
26	50.0	0.0	90.0	50.0	17553	3
28	50.0	0.0	90.0	50.0	17673	3
29	50.0	0.0	90.0	50.0	12453	3
15	50.0	0.0	90.0	50.0	12453	3
19	50.0	0.0	90.0	50.0	7200	3
19	50.0	0.0	90.0	50.0	1200	3
17	90.0	0.0	60.0	0.0	8553	5
19	90.0	0.0	60.0	0.0	8553	5
18	90.0	0.0	60.0	0.0	3600	5
16	90.0	0.0	60.0	0.0	3780	5
15	90.0	0.0	60.0	0.0	7200	5
16	90.0	0.0	60.0	0.0	4080	5
16	90.0	0.0	60.0	0.0	1800	5
16	90.0	0.0	60.0	0.0	900	5
25	15.0	0.0	90.0	50.0	4200	1
26	15.0	0.0	90.0	50.0	7800	1
28	15.0	0.0	90.0	50.0	1986	1
34	15.0	0.0	90.0	50.0	660	1
24	50.0	50.0	0.0	0.0	900	4
23	50.0	50.0	0.0	0.0	4500	4
25	50.0	50.0	0.0	0.0	5700	4
22	50.0	50.0	0.0	0.0	5100	4

26	50.0	50.0	0.0	0.0	1380	4
30	50.0	50.0	0.0	0.0	29702	4
32	50.0	50.0	0.0	0.0	29282	4
31	50.0	50.0	0.0	0.0	29702	4
23	50.0	0.0	30.0	20.0	5700	6
36	50.0	0.0	30.0	20.0	4500	6
25	50.0	0.0	30.0	20.0	4740	6
26	50.0	0.0	30.0	20.0	4500	6
26	50.0	0.0	30.0	20.0	1986	6
33	50.0	0.0	30.0	20.0	1080	6
19	50.0	0.0	30.0	20.0	29282	6
24	50.0	0.0	30.0	20.0	29702	6
17	30.0	0.0	20.0	50.0	4500	3
34	30.0	0.0	20.0	50.0	4740	3
31	30.0	0.0	20.0	50.0	4500	3
22	30.0	0.0	20.0	50.0	5100	3
33	30.0	0.0	20.0	50.0	4560	3
24	30.0	0.0	20.0	50.0	4740	3
25	30.0	0.0	20.0	50.0	5400	3
26	30.0	0.0	20.0	50.0	4080	3
25	30.0	0.0	20.0	50.0	1206	3
32	30.0	0.0	20.0	50.0	780	3
31	30.0	0.0	20.0	50.0	29222	3
23	30.0	0.0	20.0	50.0	30302	3
16	0.0	99.9	0.0	0.0	10806	5
16	0.0	99.9	0.0	0.0	7200	5
32	0.0	99.9	0.0	0.0	8700	5
16	0.0	99.9	0.0	0.0	5100	5
21	0.0	99.9	0.0	0.0	420	5
23	0.0	99.9	0.0	0.0	2106	5
30	0.0	99.9	0.0	0.0	36360	5
26	0.0	99.9	0.0	0.0	36482	5
29	0.0	99.9	0.0	0.0	36482	5
20	0.0	99.9	0.0	0.0	29282	5
21	0.0	99.9	0.0	0.0	35222	5
27	0.0	99.9	0.0	0.0	31572482	5
27	0.0	99.9	0.0	0.0	31572482	5
15	75.0	0.0	0.0	25.0	2106	6
22	75.0	0.0	0.0	25.0	8406	6
24	75.0	0.0	0.0	25.0	5706	6
22	75.0	0.0	0.0	25.0	1986	6
34	75.0	0.0	0.0	25.0	31860	6

35	75.0	0.0	0.0	25.0	29282	6
23	35.0	0.0	30.0	35.0	1206	5
25	35.0	0.0	30.0	35.0	1746	5
26	35.0	0.0	30.0	35.0	1206	5
31	35.0	0.0	30.0	35.0	1206	5
33	35.0	0.0	30.0	35.0	2106	5
32	35.0	0.0	30.0	35.0	1206	5
34	35.0	0.0	30.0	35.0	2106	5
24	35.0	0.0	30.0	35.0	726	5
21	35.0	0.0	30.0	35.0	29400	5
21	35.0	0.0	30.0	35.0	29102	5
35	0.0	90.0	0.0	10.0	1986	4
36	0.0	90.0	0.0	10.0	1986	4
32	0.0	90.0	0.0	10.0	1986	4
31	0.0	90.0	0.0	10.0	2106	4
25	0.0	90.0	0.0	10.0	726	4
24	0.0	90.0	0.0	10.0	1206	4
29	0.0	90.0	0.0	10.0	1206	4
21	0.0	90.0	0.0	10.0	1206	4
30	0.0	90.0	0.0	10.0	37860	4
24	0.0	90.0	0.0	10.0	29702	4
33	0.0	90.0	0.0	10.0	29282	4
32	0.0	90.0	0.0	10.0	29282	4
32	0.0	90.0	0.0	10.0	36482	4
13	80.0	0.0	10.0	10.0	7205	1
12	80.0	0.0	10.0	10.0	14401	1
15	80.0	0.0	10.0	10.0	7200	3
15	80.0	0.0	10.0	10.0	3605	3
16	80.0	0.0	10.0	10.0	130	3
17	80.0	0.0	10.0	10.0	1080	3
12	75.0	0.0	25.0	0.0	188	4
18	75.0	0.0	25.0	0.0	1808	4
19	75.0	0.0	25.0	0.0	4988	4
15	75.0	0.0	25.0	0.0	1688	4
19	75.0	0.0	25.0	0.0	1920	4
15	70.0	0.0	50.0	15.0	1993	5
14	60.0	0.0	40.0	0.0	3600	6
16	60.0	0.0	40.0	0.0	720	6
20	50.0	0.0	50.0	0.0	1080	5
18	80.0	0.0	10.0	90.0	2503	1
19	80.0	0.0	10.0	90.0	360	1
21	60.0	0.0	40.0	0.0	29400	5

20	80.0	0.0	60.0	20.0	30602	4
12	60.0	0.0	70.0	10.0	14400	2
12	60.0	0.0	70.0	30.0	272	5
30	55.0	0.0	60.0	0.0	30780	4
36	90.0	0.0	10.0	70.0	30780	5
35	75.0	0.0	35.0	0.0	1680	2
21	90.0	0.0	10.0	70.0	30780	3
35	80.0	0.0	30.0	75.0	30602	2
34	90.0	0.0	10.0	0.0	29040	2
32	90.0	0.0	10.0	0.0	30780	2
26	90.0	0.0	10.0	0.0	29702	2
13	60.0	0.0	50.0	65.0	25201	1
31	60.0	0.0	50.0	0.0	29100	5
35	60.0	0.0	50.0	60.0	29222	6
32	60.0	0.0	50.0	60.0	29400	2
19	60.0	0.0	65.0	0.0	30602	1
20	45.0	0.0	85.0	0.0	30780	1
36	80.0	0.0	15.0	85.0	40082	2
21	75.0	0.0	65.0	0.0	1260	5
23	75.0	0.0	65.0	0.0	1980	5
35	75.0	0.0	65.0	0.0	633	5
22	75.0	0.0	65.0	0.0	30602	5
21	80.0	0.0	15.0	80.0	1980	6
24	80.0	0.0	15.0	80.0	1260	6
31	80.0	0.0	15.0	80.0	453	6
28	80.0	0.0	15.0	80.0	30780	6
23	80.0	0.0	15.0	80.0	30602	6
23	60.0	0.0	50.0	0.0	1080	3
26	60.0	0.0	50.0	0.0	2160	3
28	60.0	0.0	50.0	0.0	2228	3
34	60.0	0.0	50.0	0.0	1408	3
36	60.0	0.0	50.0	0.0	633	3
21	60.0	0.0	50.0	0.0	29222	3
24	60.0	0.0	50.0	0.0	30602	3
25	80.0	0.0	15.0	85.0	1260	4
29	80.0	0.0	15.0	85.0	1800	4
19	80.0	0.0	15.0	85.0	30780	4
20	30.0	0.0	80.0	0.0	453	3
32	30.0	0.0	80.0	0.0	1408	3
30	80.0	0.0	15.0	85.0	633	2
33	80.0	0.0	15.0	85.0	1350	2
33	80.0	0.0	15.0	85.0	30302	2

28	25.0	0.0	85.0	0.0	30302	1
22	85.0	0.0	10.0	90.0	30602	2
36	80.0	0.0	50.0	0.0	30602	3
36	85.0	0.0	16.0	92.0	29040	6
19	85.0	0.0	16.0	92.0	29400	6
23	85.0	0.0	16.0	92.0	30302	6
12	85.0	0.0	10.0	0.0	210	1
12	85.0	0.0	10.0	0.0	3651	1
18	80.0	0.0	10.0	85.0	7201	4
34	80.0	0.0	10.0	85.0	30602	4
18	85.0	0.0	20.0	0.0	7211	3
18	85.0	0.0	20.0	0.0	349209	3
12	85.0	0.0	20.0	0.0	561	3
25	80.0	0.0	20.0	85.0	29702	5
22	85.0	0.0	20.0	0.0	29282	6
32	75.0	0.0	65.0	0.0	30602	3
33	60.0	0.0	50.0	0.0	29222	1
19	30.0	0.0	80.0	0.0	30602	5
31	80.0	0.0	10.0	10.0	8553	3
32	80.0	0.0	10.0	10.0	12573	3
20	80.0	0.0	10.0	10.0	9753	3
21	80.0	0.0	10.0	10.0	12453	3
20	80.0	0.0	10.0	10.0	8553	3
17	80.0	0.0	10.0	10.0	186	3
22	75.0	0.0	0.0	10.0	12573	3
23	75.0	0.0	0.0	10.0	9753	3
24	75.0	0.0	0.0	10.0	12453	3
25	75.0	0.0	0.0	10.0	8553	3
18	75.0	0.0	0.0	10.0	3780	3
23	75.0	0.0	0.0	10.0	480	3
26	50.0	0.0	90.0	50.0	17553	3
28	50.0	0.0	90.0	50.0	17673	3
29	50.0	0.0	90.0	50.0	12453	3
15	50.0	0.0	90.0	50.0	12453	3
19	50.0	0.0	90.0	50.0	7200	3
19	50.0	0.0	90.0	50.0	1200	3
17	90.0	0.0	60.0	0.0	8553	5
19	90.0	0.0	60.0	0.0	8553	5
18	90.0	0.0	60.0	0.0	3600	5
16	90.0	0.0	60.0	0.0	3780	5
15	90.0	0.0	60.0	0.0	7200	5
16	90.0	0.0	60.0	0.0	4080	5

16	90.0	0.0	60.0	0.0	1800	5
16	90.0	0.0	60.0	0.0	900	5
25	15.0	0.0	90.0	50.0	4200	1
26	15.0	0.0	90.0	50.0	7800	1
28	15.0	0.0	90.0	50.0	1986	1
34	15.0	0.0	90.0	50.0	660	1
24	50.0	50.0	0.0	0.0	900	4
23	50.0	50.0	0.0	0.0	4500	4
25	50.0	50.0	0.0	0.0	5700	4
22	50.0	50.0	0.0	0.0	5100	4
26	50.0	50.0	0.0	0.0	1380	4
30	50.0	50.0	0.0	0.0	29702	4
32	50.0	50.0	0.0	0.0	29282	4
31	50.0	50.0	0.0	0.0	29702	4
23	50.0	0.0	30.0	20.0	5700	6
36	50.0	0.0	30.0	20.0	4500	6
25	50.0	0.0	30.0	20.0	4740	6
26	50.0	0.0	30.0	20.0	4500	6
26	50.0	0.0	30.0	20.0	1986	6
33	50.0	0.0	30.0	20.0	1080	6
19	50.0	0.0	30.0	20.0	29282	6
24	50.0	0.0	30.0	20.0	29702	6
17	30.0	0.0	20.0	50.0	4500	3
34	30.0	0.0	20.0	50.0	4740	3
31	30.0	0.0	20.0	50.0	4500	3
22	30.0	0.0	20.0	50.0	5100	3
33	30.0	0.0	20.0	50.0	4560	3
24	30.0	0.0	20.0	50.0	4740	3
25	30.0	0.0	20.0	50.0	5400	3
26	30.0	0.0	20.0	50.0	4080	3
25	30.0	0.0	20.0	50.0	1206	3
32	30.0	0.0	20.0	50.0	780	3
31	30.0	0.0	20.0	50.0	29222	3
23	30.0	0.0	20.0	50.0	30302	3
16	0.0	99.9	0.0	0.0	10806	5
16	0.0	99.9	0.0	0.0	7200	5
32	0.0	99.9	0.0	0.0	8700	5
16	0.0	99.9	0.0	0.0	5100	5
21	0.0	99.9	0.0	0.0	420	5
23	0.0	99.9	0.0	0.0	2106	5
30	0.0	99.9	0.0	0.0	36360	5
26	0.0	99.9	0.0	0.0	36482	5

29	0.0	99.9	0.0	0.0	36482	5
20	0.0	99.9	0.0	0.0	29282	5
21	0.0	99.9	0.0	0.0	35222	5
27	0.0	99.9	0.0	0.0	31572482	5
27	0.0	99.9	0.0	0.0	31572482	5
15	75.0	0.0	0.0	25.0	2106	6
22	75.0	0.0	0.0	25.0	8406	6
24	75.0	0.0	0.0	25.0	5706	6
22	75.0	0.0	0.0	25.0	1986	6
34	75.0	0.0	0.0	25.0	31860	6
35	75.0	0.0	0.0	25.0	29282	6
23	35.0	0.0	30.0	35.0	1206	5
25	35.0	0.0	30.0	35.0	1746	5
26	35.0	0.0	30.0	35.0	1206	5
31	35.0	0.0	30.0	35.0	1206	5
33	35.0	0.0	30.0	35.0	2106	5
32	35.0	0.0	30.0	35.0	1206	5
34	35.0	0.0	30.0	35.0	2106	5
24	35.0	0.0	30.0	35.0	726	5
21	35.0	0.0	30.0	35.0	29400	5
21	35.0	0.0	30.0	35.0	29102	5
35	0.0	90.0	0.0	10.0	1986	4
36	0.0	90.0	0.0	10.0	1986	4
32	0.0	90.0	0.0	10.0	1986	4
31	0.0	90.0	0.0	10.0	2106	4
25	0.0	90.0	0.0	10.0	726	4
24	0.0	90.0	0.0	10.0	1206	4
29	0.0	90.0	0.0	10.0	1206	4
21	0.0	90.0	0.0	10.0	1206	4
30	0.0	90.0	0.0	10.0	37860	4
24	0.0	90.0	0.0	10.0	29702	4
33	0.0	90.0	0.0	10.0	29282	4
32	0.0	90.0	0.0	10.0	29282	4
32	0.0	90.0	0.0	10.0	36482	4
13	80.0	0.0	10.0	10.0	7205	1
12	80.0	0.0	10.0	10.0	14401	1
15	80.0	0.0	10.0	10.0	7200	3
15	80.0	0.0	10.0	10.0	3605	3
16	80.0	0.0	10.0	10.0	130	3
17	80.0	0.0	10.0	10.0	1080	3
12	75.0	0.0	25.0	0.0	188	4
18	75.0	0.0	25.0	0.0	1808	4

19	75.0	0.0	25.0	0.0	4988	4
15	75.0	0.0	25.0	0.0	1688	4
19	75.0	0.0	25.0	0.0	1920	4
15	70.0	0.0	50.0	15.0	1993	5
14	60.0	0.0	40.0	0.0	3600	6
16	60.0	0.0	40.0	0.0	720	6
20	50.0	0.0	50.0	0.0	1080	5
18	80.0	0.0	10.0	90.0	2503	1
19	80.0	0.0	10.0	90.0	360	1
21	60.0	0.0	40.0	0.0	29400	5
20	80.0	0.0	60.0	20.0	30602	4
12	60.0	0.0	70.0	10.0	14400	2
12	60.0	0.0	70.0	30.0	272	5
30	55.0	0.0	60.0	0.0	30780	4
36	90.0	0.0	10.0	70.0	30780	5
35	75.0	0.0	35.0	0.0	1680	2
21	90.0	0.0	10.0	70.0	30780	3
35	80.0	0.0	30.0	75.0	30602	2
34	90.0	0.0	10.0	0.0	29040	2
32	90.0	0.0	10.0	0.0	30780	2
26	90.0	0.0	10.0	0.0	29702	2
13	60.0	0.0	50.0	65.0	25201	1
31	60.0	0.0	50.0	0.0	29100	5
35	60.0	0.0	50.0	60.0	29222	6
32	60.0	0.0	50.0	60.0	29400	2
19	60.0	0.0	65.0	0.0	30602	1
20	45.0	0.0	85.0	0.0	30780	1
36	80.0	0.0	15.0	85.0	40082	2
21	75.0	0.0	65.0	0.0	1260	5
23	75.0	0.0	65.0	0.0	1980	5
35	75.0	0.0	65.0	0.0	633	5
22	75.0	0.0	65.0	0.0	30602	5
21	80.0	0.0	15.0	80.0	1980	6
24	80.0	0.0	15.0	80.0	1260	6
31	80.0	0.0	15.0	80.0	453	6
28	80.0	0.0	15.0	80.0	30780	6
23	80.0	0.0	15.0	80.0	30602	6
23	60.0	0.0	50.0	0.0	1080	3
26	60.0	0.0	50.0	0.0	2160	3
28	60.0	0.0	50.0	0.0	2228	3
34	60.0	0.0	50.0	0.0	1408	3
36	60.0	0.0	50.0	0.0	633	3

21	60.0	0.0	50.0	0.0	29222	3
24	60.0	0.0	50.0	0.0	30602	3
25	80.0	0.0	15.0	85.0	1260	4
29	80.0	0.0	15.0	85.0	1800	4
19	80.0	0.0	15.0	85.0	30780	4
20	30.0	0.0	80.0	0.0	453	3
32	30.0	0.0	80.0	0.0	1408	3
30	80.0	0.0	15.0	85.0	633	2
33	80.0	0.0	15.0	85.0	1350	2
33	80.0	0.0	15.0	85.0	30302	2
28	25.0	0.0	85.0	0.0	30302	1
22	85.0	0.0	10.0	90.0	30602	2
36	80.0	0.0	50.0	0.0	30602	3
36	85.0	0.0	16.0	92.0	29040	6
19	85.0	0.0	16.0	92.0	29400	6
23	85.0	0.0	16.0	92.0	30302	6
12	85.0	0.0	10.0	0.0	210	1
12	85.0	0.0	10.0	0.0	3651	1
18	80.0	0.0	10.0	85.0	7201	4
34	80.0	0.0	10.0	85.0	30602	4
18	85.0	0.0	20.0	0.0	7211	3
18	85.0	0.0	20.0	0.0	349209	3
12	85.0	0.0	20.0	0.0	561	3
25	80.0	0.0	20.0	85.0	29702	5
22	85.0	0.0	20.0	0.0	29282	6
32	75.0	0.0	65.0	0.0	30602	3
33	60.0	0.0	50.0	0.0	29222	1
19	30.0	0.0	80.0	0.0	30602	5
31	80.0	0.0	10.0	10.0	8553	3
32	80.0	0.0	10.0	10.0	12573	3
20	80.0	0.0	10.0	10.0	9753	3
21	80.0	0.0	10.0	10.0	12453	3
20	80.0	0.0	10.0	10.0	8553	3
17	80.0	0.0	10.0	10.0	186	3
22	75.0	0.0	0.0	10.0	12573	3
23	75.0	0.0	0.0	10.0	9753	3
24	75.0	0.0	0.0	10.0	12453	3
25	75.0	0.0	0.0	10.0	8553	3
18	75.0	0.0	0.0	10.0	3780	3
23	75.0	0.0	0.0	10.0	480	3
26	50.0	0.0	90.0	50.0	17553	3
28	50.0	0.0	90.0	50.0	17673	3

29	50.0	0.0	90.0	50.0	12453	3
15	50.0	0.0	90.0	50.0	12453	3
19	50.0	0.0	90.0	50.0	7200	3
19	50.0	0.0	90.0	50.0	1200	3
17	90.0	0.0	60.0	0.0	8553	5
19	90.0	0.0	60.0	0.0	8553	5
18	90.0	0.0	60.0	0.0	3600	5
16	90.0	0.0	60.0	0.0	3780	5
15	90.0	0.0	60.0	0.0	7200	5
16	90.0	0.0	60.0	0.0	4080	5
16	90.0	0.0	60.0	0.0	1800	5
16	90.0	0.0	60.0	0.0	900	5
25	15.0	0.0	90.0	50.0	4200	1
26	15.0	0.0	90.0	50.0	7800	1
28	15.0	0.0	90.0	50.0	1986	1
34	15.0	0.0	90.0	50.0	660	1
24	50.0	50.0	0.0	0.0	900	4
23	50.0	50.0	0.0	0.0	4500	4
25	50.0	50.0	0.0	0.0	5700	4
22	50.0	50.0	0.0	0.0	5100	4
26	50.0	50.0	0.0	0.0	1380	4
30	50.0	50.0	0.0	0.0	29702	4
32	50.0	50.0	0.0	0.0	29282	4
31	50.0	50.0	0.0	0.0	29702	4
23	50.0	0.0	30.0	20.0	5700	6
36	50.0	0.0	30.0	20.0	4500	6
25	50.0	0.0	30.0	20.0	4740	6
26	50.0	0.0	30.0	20.0	4500	6
26	50.0	0.0	30.0	20.0	1986	6
33	50.0	0.0	30.0	20.0	1080	6
19	50.0	0.0	30.0	20.0	29282	6
24	50.0	0.0	30.0	20.0	29702	6
17	30.0	0.0	20.0	50.0	4500	3
34	30.0	0.0	20.0	50.0	4740	3
31	30.0	0.0	20.0	50.0	4500	3
22	30.0	0.0	20.0	50.0	5100	3
33	30.0	0.0	20.0	50.0	4560	3
24	30.0	0.0	20.0	50.0	4740	3
25	30.0	0.0	20.0	50.0	5400	3
26	30.0	0.0	20.0	50.0	4080	3
25	30.0	0.0	20.0	50.0	1206	3
32	30.0	0.0	20.0	50.0	780	3

31	30.0	0.0	20.0	50.0	29222	3
23	30.0	0.0	20.0	50.0	30302	3
16	0.0	99.9	0.0	0.0	10806	5
16	0.0	99.9	0.0	0.0	7200	5
32	0.0	99.9	0.0	0.0	8700	5
16	0.0	99.9	0.0	0.0	5100	5
21	0.0	99.9	0.0	0.0	420	5
23	0.0	99.9	0.0	0.0	2106	5
30	0.0	99.9	0.0	0.0	36360	5
26	0.0	99.9	0.0	0.0	36482	5
29	0.0	99.9	0.0	0.0	36482	5
20	0.0	99.9	0.0	0.0	29282	5
21	0.0	99.9	0.0	0.0	35222	5
27	0.0	99.9	0.0	0.0	31572482	5
27	0.0	99.9	0.0	0.0	31572482	5
15	75.0	0.0	0.0	25.0	2106	6
22	75.0	0.0	0.0	25.0	8406	6
24	75.0	0.0	0.0	25.0	5706	6
22	75.0	0.0	0.0	25.0	1986	6
34	75.0	0.0	0.0	25.0	31860	6
35	75.0	0.0	0.0	25.0	29282	6
23	35.0	0.0	30.0	35.0	1206	5
25	35.0	0.0	30.0	35.0	1746	5
26	35.0	0.0	30.0	35.0	1206	5
31	35.0	0.0	30.0	35.0	1206	5
33	35.0	0.0	30.0	35.0	2106	5
32	35.0	0.0	30.0	35.0	1206	5
34	35.0	0.0	30.0	35.0	2106	5
24	35.0	0.0	30.0	35.0	726	5
21	35.0	0.0	30.0	35.0	29400	5
21	35.0	0.0	30.0	35.0	29102	5
35	0.0	90.0	0.0	10.0	1986	4
36	0.0	90.0	0.0	10.0	1986	4
32	0.0	90.0	0.0	10.0	1986	4
31	0.0	90.0	0.0	10.0	2106	4
25	0.0	90.0	0.0	10.0	726	4
24	0.0	90.0	0.0	10.0	1206	4
29	0.0	90.0	0.0	10.0	1206	4
21	0.0	90.0	0.0	10.0	1206	4
30	0.0	90.0	0.0	10.0	37860	4
24	0.0	90.0	0.0	10.0	29702	4
33	0.0	90.0	0.0	10.0	29282	4

32	0.0	90.0	0.0	10.0	29282	4
32	0.0	90.0	0.0	10.0	36482	4
13	80.0	0.0	10.0	10.0	7205	1
12	80.0	0.0	10.0	10.0	14401	1
15	80.0	0.0	10.0	10.0	7200	3
15	80.0	0.0	10.0	10.0	3605	3
16	80.0	0.0	10.0	10.0	130	3
17	80.0	0.0	10.0	10.0	1080	3
12	75.0	0.0	25.0	0.0	188	4
18	75.0	0.0	25.0	0.0	1808	4
19	75.0	0.0	25.0	0.0	4988	4
15	75.0	0.0	25.0	0.0	1688	4
19	75.0	0.0	25.0	0.0	1920	4
15	70.0	0.0	50.0	15.0	1993	5
14	60.0	0.0	40.0	0.0	3600	6
16	60.0	0.0	40.0	0.0	720	6
20	50.0	0.0	50.0	0.0	1080	5
18	80.0	0.0	10.0	90.0	2503	1
19	80.0	0.0	10.0	90.0	360	1
21	60.0	0.0	40.0	0.0	29400	5
20	80.0	0.0	60.0	20.0	30602	4
12	60.0	0.0	70.0	10.0	14400	2
12	60.0	0.0	70.0	30.0	272	5
30	55.0	0.0	60.0	0.0	30780	4
36	90.0	0.0	10.0	70.0	30780	5
35	75.0	0.0	35.0	0.0	1680	2
21	90.0	0.0	10.0	70.0	30780	3
35	80.0	0.0	30.0	75.0	30602	2
34	90.0	0.0	10.0	0.0	29040	2
32	90.0	0.0	10.0	0.0	30780	2
26	90.0	0.0	10.0	0.0	29702	2
13	60.0	0.0	50.0	65.0	25201	1
31	60.0	0.0	50.0	0.0	29100	5
35	60.0	0.0	50.0	60.0	29222	6
32	60.0	0.0	50.0	60.0	29400	2
19	60.0	0.0	65.0	0.0	30602	1
20	45.0	0.0	85.0	0.0	30780	1
36	80.0	0.0	15.0	85.0	40082	2
21	75.0	0.0	65.0	0.0	1260	5
23	75.0	0.0	65.0	0.0	1980	5
35	75.0	0.0	65.0	0.0	633	5
22	75.0	0.0	65.0	0.0	30602	5

21	80.0	0.0	15.0	80.0	1980	6
24	80.0	0.0	15.0	80.0	1260	6
31	80.0	0.0	15.0	80.0	453	6
28	80.0	0.0	15.0	80.0	30780	6
23	80.0	0.0	15.0	80.0	30602	6
23	60.0	0.0	50.0	0.0	1080	3
26	60.0	0.0	50.0	0.0	2160	3
28	60.0	0.0	50.0	0.0	2228	3
34	60.0	0.0	50.0	0.0	1408	3
36	60.0	0.0	50.0	0.0	633	3
21	60.0	0.0	50.0	0.0	29222	3
24	60.0	0.0	50.0	0.0	30602	3
25	80.0	0.0	15.0	85.0	1260	4
29	80.0	0.0	15.0	85.0	1800	4
19	80.0	0.0	15.0	85.0	30780	4
20	30.0	0.0	80.0	0.0	453	3
32	30.0	0.0	80.0	0.0	1408	3
30	80.0	0.0	15.0	85.0	633	2
33	80.0	0.0	15.0	85.0	1350	2
33	80.0	0.0	15.0	85.0	30302	2
28	25.0	0.0	85.0	0.0	30302	1
22	85.0	0.0	10.0	90.0	30602	2
36	80.0	0.0	50.0	0.0	30602	3
36	85.0	0.0	16.0	92.0	29040	6
19	85.0	0.0	16.0	92.0	29400	6
23	85.0	0.0	16.0	92.0	30302	6
12	85.0	0.0	10.0	0.0	210	1
12	85.0	0.0	10.0	0.0	3651	1
18	80.0	0.0	10.0	85.0	7201	4
34	80.0	0.0	10.0	85.0	30602	4
18	85.0	0.0	20.0	0.0	7211	3
18	85.0	0.0	20.0	0.0	349209	3
12	85.0	0.0	20.0	0.0	561	3
25	80.0	0.0	20.0	85.0	29702	5
22	85.0	0.0	20.0	0.0	29282	6
32	75.0	0.0	65.0	0.0	30602	3
33	60.0	0.0	50.0	0.0	29222	1
19	30.0	0.0	80.0	0.0	30602	5
31	80.0	0.0	10.0	10.0	8553	3
32	80.0	0.0	10.0	10.0	12573	3
20	80.0	0.0	10.0	10.0	9753	3
21	80.0	0.0	10.0	10.0	12453	3

20	80.0	0.0	10.0	10.0	8553	3
17	80.0	0.0	10.0	10.0	186	3
22	75.0	0.0	0.0	10.0	12573	3
23	75.0	0.0	0.0	10.0	9753	3
24	75.0	0.0	0.0	10.0	12453	3
25	75.0	0.0	0.0	10.0	8553	3
18	75.0	0.0	0.0	10.0	3780	3
23	75.0	0.0	0.0	10.0	480	3
26	50.0	0.0	90.0	50.0	17553	3
28	50.0	0.0	90.0	50.0	17673	3
29	50.0	0.0	90.0	50.0	12453	3
15	50.0	0.0	90.0	50.0	12453	3
19	50.0	0.0	90.0	50.0	7200	3
19	50.0	0.0	90.0	50.0	1200	3
17	90.0	0.0	60.0	0.0	8553	5
19	90.0	0.0	60.0	0.0	8553	5
18	90.0	0.0	60.0	0.0	3600	5
16	90.0	0.0	60.0	0.0	3780	5
15	90.0	0.0	60.0	0.0	7200	5
16	90.0	0.0	60.0	0.0	4080	5
16	90.0	0.0	60.0	0.0	1800	5
16	90.0	0.0	60.0	0.0	900	5
25	15.0	0.0	90.0	50.0	4200	1
26	15.0	0.0	90.0	50.0	7800	1
28	15.0	0.0	90.0	50.0	1986	1
34	15.0	0.0	90.0	50.0	660	1
24	50.0	50.0	0.0	0.0	900	4
23	50.0	50.0	0.0	0.0	4500	4
25	50.0	50.0	0.0	0.0	5700	4
22	50.0	50.0	0.0	0.0	5100	4
26	50.0	50.0	0.0	0.0	1380	4
30	50.0	50.0	0.0	0.0	29702	4
32	50.0	50.0	0.0	0.0	29282	4
31	50.0	50.0	0.0	0.0	29702	4
23	50.0	0.0	30.0	20.0	5700	6
36	50.0	0.0	30.0	20.0	4500	6
25	50.0	0.0	30.0	20.0	4740	6
26	50.0	0.0	30.0	20.0	4500	6
26	50.0	0.0	30.0	20.0	1986	6
33	50.0	0.0	30.0	20.0	1080	6
19	50.0	0.0	30.0	20.0	29282	6
24	50.0	0.0	30.0	20.0	29702	6

17	30.0	0.0	20.0	50.0	4500	3
34	30.0	0.0	20.0	50.0	4740	3
31	30.0	0.0	20.0	50.0	4500	3
22	30.0	0.0	20.0	50.0	5100	3
33	30.0	0.0	20.0	50.0	4560	3
24	30.0	0.0	20.0	50.0	4740	3
25	30.0	0.0	20.0	50.0	5400	3
26	30.0	0.0	20.0	50.0	4080	3
25	30.0	0.0	20.0	50.0	1206	3
32	30.0	0.0	20.0	50.0	780	3
31	30.0	0.0	20.0	50.0	29222	3
23	30.0	0.0	20.0	50.0	30302	3
16	0.0	99.9	0.0	0.0	10806	5
16	0.0	99.9	0.0	0.0	7200	5
32	0.0	99.9	0.0	0.0	8700	5
16	0.0	99.9	0.0	0.0	5100	5
21	0.0	99.9	0.0	0.0	420	5
23	0.0	99.9	0.0	0.0	2106	5
30	0.0	99.9	0.0	0.0	36360	5
26	0.0	99.9	0.0	0.0	36482	5
29	0.0	99.9	0.0	0.0	36482	5
20	0.0	99.9	0.0	0.0	29282	5
21	0.0	99.9	0.0	0.0	35222	5
27	0.0	99.9	0.0	0.0	31572482	5
27	0.0	99.9	0.0	0.0	31572482	5
15	75.0	0.0	0.0	25.0	2106	6
22	75.0	0.0	0.0	25.0	8406	6
24	75.0	0.0	0.0	25.0	5706	6
22	75.0	0.0	0.0	25.0	1986	6
34	75.0	0.0	0.0	25.0	31860	6
35	75.0	0.0	0.0	25.0	29282	6
23	35.0	0.0	30.0	35.0	1206	5
25	35.0	0.0	30.0	35.0	1746	5
26	35.0	0.0	30.0	35.0	1206	5
31	35.0	0.0	30.0	35.0	1206	5
33	35.0	0.0	30.0	35.0	2106	5
32	35.0	0.0	30.0	35.0	1206	5
34	35.0	0.0	30.0	35.0	2106	5
24	35.0	0.0	30.0	35.0	726	5
21	35.0	0.0	30.0	35.0	29400	5
21	35.0	0.0	30.0	35.0	29102	5
35	0.0	90.0	0.0	10.0	1986	4

36	0.0	90.0	0.0	10.0	1986	4
32	0.0	90.0	0.0	10.0	1986	4
31	0.0	90.0	0.0	10.0	2106	4
25	0.0	90.0	0.0	10.0	726	4
24	0.0	90.0	0.0	10.0	1206	4
29	0.0	90.0	0.0	10.0	1206	4
21	0.0	90.0	0.0	10.0	1206	4
30	0.0	90.0	0.0	10.0	37860	4
24	0.0	90.0	0.0	10.0	29702	4
33	0.0	90.0	0.0	10.0	29282	4
32	0.0	90.0	0.0	10.0	29282	4
32	0.0	90.0	0.0	10.0	36482	4
13	80.0	0.0	10.0	10.0	7205	1
12	80.0	0.0	10.0	10.0	14401	1
15	80.0	0.0	10.0	10.0	7200	3
15	80.0	0.0	10.0	10.0	3605	3
16	80.0	0.0	10.0	10.0	130	3
17	80.0	0.0	10.0	10.0	1080	3
12	75.0	0.0	25.0	0.0	188	4
18	75.0	0.0	25.0	0.0	1808	4
19	75.0	0.0	25.0	0.0	4988	4
15	75.0	0.0	25.0	0.0	1688	4
19	75.0	0.0	25.0	0.0	1920	4
15	70.0	0.0	50.0	15.0	1993	5
14	60.0	0.0	40.0	0.0	3600	6
16	60.0	0.0	40.0	0.0	720	6
20	50.0	0.0	50.0	0.0	1080	5
18	80.0	0.0	10.0	90.0	2503	1
19	80.0	0.0	10.0	90.0	360	1
21	60.0	0.0	40.0	0.0	29400	5
20	80.0	0.0	60.0	20.0	30602	4
12	60.0	0.0	70.0	10.0	14400	2
12	60.0	0.0	70.0	30.0	272	5
30	55.0	0.0	60.0	0.0	30780	4
36	90.0	0.0	10.0	70.0	30780	5
35	75.0	0.0	35.0	0.0	1680	2
21	90.0	0.0	10.0	70.0	30780	3
35	80.0	0.0	30.0	75.0	30602	2
34	90.0	0.0	10.0	0.0	29040	2
32	90.0	0.0	10.0	0.0	30780	2
26	90.0	0.0	10.0	0.0	29702	2
13	60.0	0.0	50.0	65.0	25201	1

31	60.0	0.0	50.0	0.0	29100	5
35	60.0	0.0	50.0	60.0	29222	6
32	60.0	0.0	50.0	60.0	29400	2
19	60.0	0.0	65.0	0.0	30602	1
20	45.0	0.0	85.0	0.0	30780	1
36	80.0	0.0	15.0	85.0	40082	2
21	75.0	0.0	65.0	0.0	1260	5
23	75.0	0.0	65.0	0.0	1980	5
35	75.0	0.0	65.0	0.0	633	5
22	75.0	0.0	65.0	0.0	30602	5
21	80.0	0.0	15.0	80.0	1980	6
24	80.0	0.0	15.0	80.0	1260	6
31	80.0	0.0	15.0	80.0	453	6
28	80.0	0.0	15.0	80.0	30780	6
23	80.0	0.0	15.0	80.0	30602	6
23	60.0	0.0	50.0	0.0	1080	3
26	60.0	0.0	50.0	0.0	2160	3
28	60.0	0.0	50.0	0.0	2228	3
34	60.0	0.0	50.0	0.0	1408	3
36	60.0	0.0	50.0	0.0	633	3
21	60.0	0.0	50.0	0.0	29222	3
24	60.0	0.0	50.0	0.0	30602	3
25	80.0	0.0	15.0	85.0	1260	4
29	80.0	0.0	15.0	85.0	1800	4
19	80.0	0.0	15.0	85.0	30780	4
20	30.0	0.0	80.0	0.0	453	3
32	30.0	0.0	80.0	0.0	1408	3
30	80.0	0.0	15.0	85.0	633	2
33	80.0	0.0	15.0	85.0	1350	2
33	80.0	0.0	15.0	85.0	30302	2
28	25.0	0.0	85.0	0.0	30302	1
22	85.0	0.0	10.0	90.0	30602	2
36	80.0	0.0	50.0	0.0	30602	3
36	85.0	0.0	16.0	92.0	29040	6
19	85.0	0.0	16.0	92.0	29400	6
23	85.0	0.0	16.0	92.0	30302	6
12	85.0	0.0	10.0	0.0	210	1
12	85.0	0.0	10.0	0.0	3651	1
18	80.0	0.0	10.0	85.0	7201	4
34	80.0	0.0	10.0	85.0	30602	4
18	85.0	0.0	20.0	0.0	7211	3
18	85.0	0.0	20.0	0.0	349209	3

12	85.0	0.0	20.0	0.0	561	3
25	80.0	0.0	20.0	85.0	29702	5
22	85.0	0.0	20.0	0.0	29282	6
32	75.0	0.0	65.0	0.0	30602	3
33	60.0	0.0	50.0	0.0	29222	1
19	30.0	0.0	80.0	0.0	30602	5
31	80.0	0.0	10.0	10.0	8553	3
32	80.0	0.0	10.0	10.0	12573	3
20	80.0	0.0	10.0	10.0	9753	3
21	80.0	0.0	10.0	10.0	12453	3
20	80.0	0.0	10.0	10.0	8553	3
17	80.0	0.0	10.0	10.0	186	3
22	75.0	0.0	0.0	10.0	12573	3
23	75.0	0.0	0.0	10.0	9753	3
24	75.0	0.0	0.0	10.0	12453	3
25	75.0	0.0	0.0	10.0	8553	3
18	75.0	0.0	0.0	10.0	3780	3
23	75.0	0.0	0.0	10.0	480	3
26	50.0	0.0	90.0	50.0	17553	3
28	50.0	0.0	90.0	50.0	17673	3
29	50.0	0.0	90.0	50.0	12453	3
15	50.0	0.0	90.0	50.0	12453	3
19	50.0	0.0	90.0	50.0	7200	3
19	50.0	0.0	90.0	50.0	1200	3
17	90.0	0.0	60.0	0.0	8553	5
19	90.0	0.0	60.0	0.0	8553	5
18	90.0	0.0	60.0	0.0	3600	5
16	90.0	0.0	60.0	0.0	3780	5
15	90.0	0.0	60.0	0.0	7200	5
16	90.0	0.0	60.0	0.0	4080	5
16	90.0	0.0	60.0	0.0	1800	5
16	90.0	0.0	60.0	0.0	900	5
25	15.0	0.0	90.0	50.0	4200	1
26	15.0	0.0	90.0	50.0	7800	1
28	15.0	0.0	90.0	50.0	1986	1
34	15.0	0.0	90.0	50.0	660	1
24	50.0	50.0	0.0	0.0	900	4
23	50.0	50.0	0.0	0.0	4500	4

Anexo XII: *Dataset* utilizado para efetuar o clustering de utilizadores (funcionalidade 1 da plataforma de análise de dados).

ID	Geénero	Idade	EA Predominante	SeusConteudosVisuais	SeusConteudosAurais	SeusConteudosReadwrite	SeusConteudosKinesthetic
13	masculino	36	<i>kinesthetic</i>	1,53	0,00	0,00	98,47
14	masculino	23	<i>kinesthetic</i>	100,00	0,00	0,00	0,00
17	masculino	22	<i>kinesthetic</i>	85,12	0,00	0,00	14,88
18	feminino	20	<i>kinesthetic</i>	87,55	0,00	0,00	12,45
19	masculino	21	<i>kinesthetic</i>	4,54	0,00	92,05	3,41
20	masculino	26	<i>kinesthetic</i>	49,13	12,28	38,58	0,00
21	masculino	25	<i>kinesthetic</i>	78,83	21,17	0,00	0,00
22	feminino	26	<i>readwrite</i>	84,62	0,00	0,00	15,38
23	feminino	42	<i>readwrite</i>	88,17	3,94	0,00	7,89
24	masculino	20	visual	92,75	4,83	0,00	2,42
25	masculino	22	<i>readwrite</i>	80,98	3,17	3,17	12,68
26	masculino	24	visual	73,34	6,67	13,33	6,67
27	masculino	22	<i>kinesthetic</i>	0,00	100,00	0,00	0,00
28	masculino	21	visual	96,37	0,00	3,63	0,00
29	masculino	26	<i>readwrite</i>	0,00	71,58	14,21	14,21
30	masculino	27	<i>kinesthetic</i>	15,16	30,33	15,16	39,34
31	masculino	25	<i>kinesthetic</i>	37,58	47,39	0,00	15,03
32	masculino	23	visual	42,97	28,64	7,16	21,23
33	masculino	24	<i>readwrite</i>	3,71	92,57	0,00	3,71
34	masculino	21	visual	21,12	0,00	5,28	73,60
35	masculino	19	<i>readwrite</i>	45,44	54,56	0,00	0,00
36	masculino	18	<i>readwrite</i>	33,33	50,01	0,00	16,66

Anexo XIII: *Dataset* utilizado para gerar regras de associação sobre os dados dos utilizadores (feature 3 da plataforma de análise de dados).

Idade	EA	Media de Notas
36-40	<i>kinesthetic</i>	bom
36-40	<i>kinesthetic</i>	insatisfaz
17-21	<i>kinesthetic</i>	satisfaz
17-21	visual	insatisfaz
22-26	<i>kinesthetic</i>	satisfaz
22-26	<i>kinesthetic</i>	satisfaz
22-26	visual	satisfaz
>41	<i>readwrite</i>	muito bom
17-21	visual	insatisfaz
22-26	<i>readwrite</i>	fraco
22-26	visual	satisfaz
22-26	<i>kinesthetic</i>	satisfaz
17-21	visual	insatisfaz
22-26	<i>readwrite</i>	satisfaz
27-30	visual	satisfaz
22-26	visual	satisfaz
22-26	visual	satisfaz
22-26	<i>readwrite</i>	satisfaz
17-21	visual	insatisfaz
17-21	visual	satisfaz
17-21	<i>readwrite</i>	insatisfaz

Anexo XIV: *Dataset* utilizado para treinar a rede neuronal que permite prever as notas dos alunos (feature 9 da plataforma de análise de dados).

Id User	Id Disciplina	EA predominante	OA 100	OA N	Numero OAs Consultados	class (passou ou não à disciplina)
13	1	<i>kinesthetic</i>	TRUE	FALSE	<5	TRUE
13	2	<i>kinesthetic</i>	TRUE	FALSE	<5	TRUE
18	1	<i>kinesthetic</i>	FALSE	FALSE	6-10	FALSE
19	2	<i>visual</i>	FALSE	FALSE	<5	FALSE
20	1	<i>kinesthetic</i>	FALSE	FALSE	11-15	FALSE
20	2	<i>kinesthetic</i>	FALSE	FALSE	11-15	FALSE
21	1	<i>kinesthetic</i>	FALSE	FALSE	11-15	TRUE
21	2	<i>kinesthetic</i>	FALSE	FALSE	11-15	TRUE
22	2	<i>visual</i>	FALSE	FALSE	<5	FALSE
23	2	<i>readwrite</i>	FALSE	FALSE	6-10	TRUE
24	1	<i>aural</i>	FALSE	FALSE	11-15	FALSE
24	2	<i>aural</i>	FALSE	FALSE	11-15	FALSE
25	2	<i>readwrite</i>	FALSE	FALSE	6-10	FALSE
26	1	<i>aural</i>	FALSE	FALSE	11-15	TRUE
26	2	<i>aural</i>	FALSE	FALSE	11-15	TRUE
27	2	<i>kinesthetic</i>	FALSE	FALSE	<5	TRUE
28	1	<i>visual</i>	FALSE	FALSE	11-15	FALSE
28	2	<i>visual</i>	FALSE	FALSE	11-15	FALSE
29	1	<i>readwrite</i>	FALSE	FALSE	6-10	TRUE
29	2	<i>readwrite</i>	FALSE	FALSE	6-10	TRUE
30	1	<i>visual</i>	FALSE	FALSE	6-10	TRUE
30	2	<i>visual</i>	FALSE	FALSE	6-10	TRUE
31	2	<i>visual</i>	FALSE	FALSE	11-15	TRUE
32	2	<i>aural</i>	FALSE	FALSE	6-10	TRUE
33	1	<i>readwrite</i>	FALSE	FALSE	6-10	FALSE
33	2	<i>readwrite</i>	FALSE	FALSE	6-10	FALSE
34	1	<i>aural</i>	FALSE	FALSE	11-15	FALSE
34	2	<i>aural</i>	FALSE	FALSE	11-15	FALSE
35	1	<i>visual</i>	FALSE	FALSE	6-10	TRUE
35	2	<i>visual</i>	FALSE	FALSE	6-10	TRUE
36	2	<i>readwrite</i>	FALSE	FALSE	6-10	FALSE

Anexo XV: *Dataset* utilizado para efetuar o clustering sobre os dados dos utilizadores e suas consultas (funcionalidade 10 da plataforma de análise de dados).

USER	USERVISUAL	USERAURAL	USERREADWRITE	USERKINESTHETIC	OASMEDIAVISUAL	OASDESIVOVISUAL	OASMEDIAAURAL	OASDESIVOAURAL	OASMEDIAREADWRITE	OASDESVIOREADWRITE	OASMEDIAKINESTHETIC	OASDESVIOKINESTHETIC	USERIDSCONSULTADOS	USERPONTUACAO
18	91,60	8,30	8,30	91,60	82,50	7,50	0,00	0,00	42,50	17,50	0,00	0,00	2,00	6,95
19	60,00	2,00	29,00	30,00	67,50	30,65	15,00	33,54	19,33	20,26	32,83	39,62	6,00	4,50
20	4,00	18,00	57,00	82,00	55,00	17,11	0,00	0,00	52,86	31,38	10,71	16,57	7,00	7,00
21	50,00	51,00	21,00	90,00	57,00	31,80	18,99	38,04	17,00	20,64	25,50	27,34	10,00	7,00
22	85,00	18,00	84,00	79,00	59,17	17,18	8,33	18,63	10,83	16,94	19,17	16,69	6,00	6,67
23	51,00	26,00	79,00	14,00	54,38	26,27	18,74	34,77	29,38	25,55	8,13	12,23	8,00	8,95
24	56,00	89,00	75,00	36,00	54,38	28,99	17,50	31,92	15,63	19,91	26,25	26,07	8,00	4,00
25	7,00	30,00	36,00	20,00	45,50	27,70	14,00	29,39	26,50	27,39	31,00	26,34	10,00	2,33
26	73,00	93,00	55,00	19,00	42,50	13,69	0,00	0,00	46,25	26,43	30,63	17,40	8,00	5,50
28	80,00	67,00	21,00	2,00	50,63	25,91	0,00	0,00	45,63	36,86	39,38	24,29	8,00	4,00
29	25,00	39,50	50,00	0,00	51,25	31,70	22,50	38,97	26,25	37,31	42,50	28,39	4,00	6,33
30	81,00	51,00	40,00	53,00	35,83	29,21	31,65	44,85	23,33	22,30	25,00	31,89	6,00	6,00
31	79,00	30,00	42,00	47,00	48,13	24,99	17,50	31,92	26,88	28,39	29,38	27,21	8,00	7,00
32	7,00	67,00	42,00	3,00	36,88	34,18	34,99	45,26	22,50	27,27	15,63	19,91	8,00	5,33
33	49,00	15,00	86,00	58,00	48,33	22,48	0,00	0,00	21,67	6,24	56,67	20,95	3,00	5,33
34	30,00	65,00	57,00	22,00	58,13	19,52	0,00	0,00	36,25	34,62	29,38	19,60	8,00	4,33
35	51,00	9,30	39,50	0,00	37,50	37,50	45,00	45,00	32,50	32,50	5,00	5,00	2,00	5,50
36	40,00	34,00	69,00	26,00	51,25	32,67	22,50	38,97	24,50	29,69	34,63	30,61	8,00	4,33

Anexo XVI: *Dataset* utilizado no sistema de recomendações

OA consultado 0	OA consultado 1	OA consultado 2	OA consultado 3	OA consultado 4	OA consultado 5
LO 171	LO 147	LO 180	LO 190	LO 194	LO 199
LO 199	LO 178	LO 146	LO 172	LO 205	LO 204
LO 195	LO 178	LO 119	LO 122	LO 130	LO 176
LO 100	LO 103	LO 158	LO 108	LO 112	LO 156 ¹⁵

¹⁵ Este *dataset* contém apenas os dados dos utilizadores próximos ao utilizador de id=22 isto é, utilizadores predominantemente visuais e inscritos na disciplina de Sistemas Digitais.
