THE GRAPH HEURISTIC WITH GREAT DELUGE ALGORITHM TO SOLVE THE ITC 2007 EXAMI

PERPUSTAKAAN UMP
0000103201

EM

CHUAN WEI KIAT

THESIS SUBMITTED IN FULFILMENT OF BACHELOR OF COMPUTER SCIENCE (GRAPHIC AND MULTIMEDIA TECHNOLOGY) WITH HONORS

FACULTY OF COMPUTER SYSTEM AND SOFTWARE ENGINEERING
UNIVERSITI MALAYSIA PAHANG

2014

ABSTRACT

Examination timetabling problem involve allocating the exams into a specific or limited number of timeslot and rooms. The produce examination timetable should meet all the hard constraints and achieve the soft constrains as much as possible. In this research, dataset of examination track from second International Timetabling Competition (ITC 2007) will be use. This dataset contains of total of twelve dataset which have different features (i.e. number of room, exam and timeslot). The graph heuristic with great deluge algorithm will be used to solve this problem. We hope that the proposed technique will be able to produce better result than other technique in the literature.

ABSTRAK

Masalah jadual waktu peperiksaan melibatkan mengaturkan peperiksaan ke dalam jumlah slot masa dan bilik-bilik tertentu atau terhad. Hasil jadual waktu peperiksaan harus memenuhi semua kekangan keras dan mencapai kekangan lembut sebanyak mungkin. Dalam kajian ini, dataset trek pemeriksaan dari kedua Pertandingan jadual waktu Antarabangsa (ITC 2007) akan digunakan. Dataset ini mengandungi daripada jumlah dua belas dataset yang mempunyai ciri-ciri yang berbeza(iaitu jumlah bilik, peperiksaan dan slot masa). Graf Heuristic dengan Great Deluge algoritma akan digunakan untuk menyelesaikan masalah ini. Kami berharap bahawa teknik yang dicadangkan akan dapat menghasilkan keputusan yang lebih baik berbanding teknik yang lain dalam kesusasteraan.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Background of Study

Nowadays, almost all education institutes have problem about exam timetabling, especially university. Generally, the university examination timetabling problems played an important role which involves allocating examination to timeslots subject to fulfill several constrains problem which are hard constraints and soft constrain. To ensure the feasibility of timetable the hard constraint must all be met by the resulting timetable. As an example, a student cannot sit for two examinations at the same time. On the others hand, the requirement for soft constraint is not necessary but the quality of the timetable is needed to improve by soft constraint as far as possible. For example, a student should not take more than two exams in a day.

There are two versions of examination problem exist which known as capacitated and the un-capacitated version. Capacitated problem is more close to the real world problem for example the room capacities can be consider as one of the constraints but there are less of the researchers put their awareness on the capacitated version. This situation happen might be the lack of capacitated benchmark dataset. In addition, there are many constraints in capacitated problem makes it more difficult to compare to un-capacitated problem. The difficulties when solving the capacitated problem involves the consideration of constraints such as the amount and the size of the room are those constraints that increase the difficulties when solving capacitated problem. There is the result of survey from Burke et al.'s (1996) shown that 73% of the universities agreed that scheduling examination timetable is a hard task. The un-capacitated version of the problem does not take room capacities into consideration like what the capacitated version does. Researchers are focusing on the algorithm and algorithmic performance to produce a quick and effective solution. Thus, there are lots of researchers able to solve the un-capacitated exams problem.

We will use the capacitated problem dataset, International Timetabling Competition 2007(ITC2007) in this project. ITC 2007 dataset has been added several new constraints instead of those that are found in scientific literature.

1.2 Problem Statement

Examination timetabling problem has been concerned and attracted many researchers in this few years. There are some un-capacitated dataset is used by many reported in the literature. One of the examples for this un-capacitated dataset is Toronto dataset. However, this dataset does not mimic the real world examination timetabling problem. Besides that, there is some researchers study on capacitated dataset such as Nottingham and Melbourne dataset but it has a constraint that only a maximum number of seats are included in a day. This does not really similar the real world problem as we need to consider the individual room capacity. Therefore, there will be a gap between research and real practical.

The ITC 2007 examination dataset will be concerned because this dataset is a real world capacitated examination dataset. ITC 2007 contain different constraints from others dataset constraint like Toronto, Nottingham and Melbourne that seen in the literature. In ITC 2007, there are two categories of hard constraints which are "Period Related Hard Constraints" and "Room Related Hard Constraints". The example of these hard constraints is student cannot sit more than one exam at the same time, the exam capacity should not exceed the room capacity, exam length should not violate the timeslot length, a sequence or ordering of an exams must be respected and schedule exam into specified room. On the others hand, the soft constraints are two exams in a row, two exams in a day, spreading of exam and mixed duration.

1.3 Objective

- To study on the examination track of the Second International Timetabling Competition (ITC 2007).
- To develop timetabling problem by using graph heuristic method with Great Deluge Algorithm.
- To evaluate the timetable produce whether it satisfied all the hard constraints and soft constraints as much as possible

1.4 Scope

In this research, we will study the examination dataset from ITC 2007. ITC 2007 is a capacitated dataset that will consider room size and also the number of rooms as the hard constraints. We will implement graph heuristic with great deluge algorithm. method to develop a schedule for the dataset.

1.5 Thesis Organization

There are total six chapters in this thesis. Basically, Chapter 1 will discuss about the introduction of the study. In Chapter 2 will involve the literature review and we will discuss more on methodology in Chapter 3. In addition, Chapter 4 will further explain on design and implementation. Chapter 5 will show the result and discussion of the research. Finally, there will be a conclusion of the research in Chapter 6.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview of timetabling

Timetable or we call schedule is a basic time management tool that consists of time for all events, tasks and when it takes place. It can be categorized into different types. Such as personal timetabling, educational timetabling (Qu el al 2009), transportation timetabling and so on. These timetabling may have different constraint or requirement that need to be satisfied. (Burke, Kingston and deWerra2004).

Basically, timetabling problem contain two types of constraints. One of the constraints is hard constrains and the others one is soft constrains. Hard constrains are constrains that must be completely included without any break. A successful timetable must be fulfilling all the hard constrains. Example of hard constrains in exam timetabling is a student won't be able to attend two exams at the same time. On the others hand, soft constrains are not necessary to fulfill all but should be satisfied as much as possible. It will increase the quality of the timetable by reduce the violations to the soft constrains. Example of soft constrains is a student more prefer two days one exam which the exam period is not too close until they not enough time to study and not too far until it take a long time to finish the exam.

2.2University timetabling problems

University timetabling can divide into two types, one is course timetabling and the other one is exam timetabling. Both of these have similar core issues and also feature (e.g., Carter and Laporte, 1996). The main concern of these timetabling problems is to avoid the same students sitting two examinations at the same time. But there are some major differences between course timetabling and exam timetabling so that we still can differentiate them. The first different is constrains that we need to follow. There are some example of hard and soft constraints of the course timetabling problems (Abdullah, 2006) and the examination timetabling problems (Qu et al., 2009) that show in table 2.1 and 2.2.

Besides than the constrains, there are some different in method when building course and examination tabling. In the building itself can be divided into process environment, scheduling instances and modeling. For the process environment, the producer of the exam tabling and course tabling is different person. Exam tabling is produced by academic department while course tabling is produce by school independently according to McCollum (2007) and Burke et al (1996). For scheduling instances, all courses that had been offered will formed out an examination timetabling, while we just need to schedule the lecturer session, tutorial session, laboratory session and also courses that offered in course time tabling. (McCollum, 2007). In modeling, examination timetabling is based on total number of courses and the registered course by the students while course time tabling formed based on how many students is taking the course. (McCollum, 2007).

As we can see, there is a different between examination tabling and course tabling, but complexity of problem is directly proportional to the degree of freedom. Open registration (Laporte and Desroches, 1984) that allows students to register their course base on their own timetable make the problem become more complex. There will be more difficult to form a feasible timetable if the student is given more freedom to register their course.

Table 2.1 Examples of constraints of the course timetabling problems (Abdullah, 2006)

Hard Constraints
• A teacher and student should not be assigned in more than one place at the same timeslot.
• Each timeslot can only have one course with only one schoolroom.
• Capacity of each classroom must be able to accommodate the total number of students that attend the course at a certain timeslot by having equal capacity or more than that.
• The classroom should have the suitable equipment and features to fulfill the course that being assigned in.
Soft Constraints
• Each student should have more than one course per day.
• A student should be avoiding attending two or more consecutive courses on a day.
• Each student should be avoiding to be scheduled to attend a course which is being allocated to the final timeslot of the day.

2.3Examination timetabling

In education institution, an examination timetable plays an important rule. Everything regarding exam timetable will directly affect administrators, lecturer and student.

Based on Schaerf(1990), "The examination timetabling problem requires the scheduling of a given number of exams (one for each course) within a given amount of time" , the statement above status that the goal of exam timetabling is to arrange all the exams into a series of rooms as well as timeslots which satisfied the hard constrains and also soft constrains as much as possible. (e.g., Carter and Laporte, 1996a; Qu et al., 2009). Table 2.2 is shows some of the constraints in examination timetabling problems.

Table 2.2 Examples of constraints of the examination timetabling problems (Qu et al, 2009)

<u>Hard Constraints</u>

- There are no collaborative resources (e.g. Students) in exams being assigned simultaneously.
- There are sufficient resources to be used for examination timetable (e.g. the number of students that take the exam must be less or equal to the room capacity of that exam).

<u>Soft Constraints</u>

- The exams should not in any consecutive period slots or days and should spread as much as possible.
- The exams in same group must be held at the same period, day or at same place.
- Consecutive all of the exams.
- Every exam should be scheduled first or the largest exams should be scheduled at early time compared to others small exams.
- Satisfied all prior exams condition.
- Every timeslot should limit the numbers of exams and students.
- Some specific exams must be place in certain timeslots as request by the school.
- Located conflicting exams on the same day as near as possible.
- Might be able to split the exams over nearby or similar places.
- Combined the exams with the same length into same room as long as got sufficient room capacity for students.
- Resource requirements should be fulfill as many as possible.

2.3.1 Constrains and objective investigated in examination timetabling problem

We can found a lot of constrains of exam timetabling in the literature. This is because of every institution may have their own constrains that need to be fulfill. In addition, the parties which are affected by the timetable might have different preference to produce a high quality timetable. For an example, the administrator takes the responsibility to construct an examination schedule and should ensure that no student should sit for two exams in the same time. For student, they hope to have a bigger gap between the exam days so that they have enough time to do their revision. Now, we need to discuss about the common constraints in timetabling problems which usually been used by institution.

There are some most common used datasets in the examination timetabling research community which are Toronto dataset (Carter, Laporte and Lee, 1996), Nottingham (Burke, Newall and weare, 1996) and Melbourne (Merlot et al. 2003). Amount these, the researchers mostly focus on Toronto dataset compare to other two datasets. In year 2008, McCollum et al (2008) had introduced the Second International Timetabling Competition (ITC2007) dataset which is more realistic problems than the benchmark problems. In addition, there are other examination datasets, such as UKM (Ayoub et al, 2007) and UiTM (Kendall and Hussin, 2004; Hu Xin, 2005).

2.4 Capacitated and un- capacitated examination timetabling problem

Un-capacitated problem had been researched in a lot of the literature. In terms of producing a solution that effectively and quickly, it is more focus on the algorithm and algorithmic performance. There is why most of the researchers' not dealing with all the side views of problems but only focus on a simplified version of examination problem (McCollum (2007), Carter and Laporte (1996)).In addition, those survey forms only include some of the common hard constraints. For example, the quantities of student sitting in the exams must not over the size of the room, none of student can take more than one subject at the same time and other else. While in soft constraints, they will normally use separate the examination gap as far as possible, so that student will not having two exam in a day.

On the others hand, capacitated problem is having the different characteristic of problem compare with un-capacitated problem. It has included capacity as the constraints and this will make the system more close to the real world. Yet, those researchers will less to do the research on capacitated problem maybe is because of the lack of benchmark dataset. Furthermore, solving capacitated problem will be more hard compare to un-capacitated problem. In Burke ea al (1996a) survey paper, it shows that accommodating exams is really a hard problem which is agreed by around 73% of universities. Since capacitated problems need to consider about the capacity of room and also the less complex problems such as student and exam list, so it needed more overall data. According to McCollum (2007), it might be hard to collect that additional information. Burke et al (1996) had mentioned that due to the lack of halls availability and divide the student of taking exam into more than one room which will cause other constraints likes splitting an exam onto different sites or taking into account between rooms. This will be the hardness of examination scheduling.

To make sure that the dataset will be more suitable to the realistic problems and also because of the interest in capacitated problem, Burke, Newall, and Weare, (1996) had made some improvement on the benchmark dataset (i.e. Toronto dataset), which consists of the overall capacity if in one big room/hall all exams will be held in. However, this also shows that we need to consider about the individual room capacity when we simplify the timetabling problems (Merlot et al (2003)).

2.4.1 Toronto dataset

Toronto dataset is a dataset that consist of thirteen real-world examination timetabling problems which five from the Canadian institution, three from the Canadian highs schools, one from the London School of Economics, one from King Fahd University, Dhahran and one more from Purdue University, Indiaana (Carter, Laporte and Lee, 1996). The information of the Toronto datasets is showed in table 2.3.Qu et al (2009) to make sure it will have the real comparison between the scientific communities it was classified the problem instances into I and II. In year 1996, Carte, Laporte and Lee, was introduced the Toronto dataset. To reduce the timeslots and to spread the conflicting exams within the timeslots, they had investigated two variants of the objectives with the purpose. Di Gaspero and Schaerf using tabu search to solve this dataset in the year 2001 by only consider the constraints conflict. When is the student need to take two exams at the same timeslots as a hard constraint and when a student needs to take two exams in continuous periods as a soft constraint. Some objective had included to the original dataset by Burke, Newall and Weare, (1996).The maximum quantity of the room capacity per timeslot and second-order conflict of same day constraints is started considered by them. Merlot et al. (2003) applying the hybridize of constraint programming, hill climbing (HC) and simulated annealing (SA) to solve the dataset with the objective that to spreading the conflict examinations within limited number of timeslots, reduce the needed of timeslots, and to minimize second-order conflict at the same day.

Table 2.3 Toronto Dataset (Qu et al., 2009)

Problem Instance	Exams	Students	Enrollments	Conflic Density	TimeSlots
car91 I	682	16925	56877	0.13	35
car91 II	682	16925	56242/56877	0.13	35
car92 I	543	18419	55522	0.14	32
car92 II	543	18149	55189/55522	0.14	32
ear83 I	190	1125	8109	0.27	24
ear83 II	189	1108	8014	0.27	24
hec92 I	81	2823	10632	0.42	18
hec92 II	80	2823	10625	0.42	18
kfu93	461	5349	25113	0.03	42
lse91	381	2726	10918	0.06	18
pur93 I	2419	30029	120681	0.03	42
pur93 II	2419	30029	120686/120681	0.03	42
rye92	486	11483	45051	0.07	23
sta83 I	139	611	5751	0.14	13
sta 83 II	138	549	5689	0.14	13
tre92	261	4360	14901	0.18	23
uta92 I	622	21266	58979	0.13	35
uta 92 II	638	21329	59144	0.13	35
ute92	184	2749	11793	0.08	10
yor83 I	181	941	6034	0.29	21
yor83 II	180	919	6012	0.29	21

2.4.2 Dataset of University of Melbourne

Melbourne dataset is a dataset that introduced by Merlot et al., (2003). Basically, it has two different dataset that consist of two timeslots on weekday and also the capacities of each timeslots are different. Period exclusive constraints which in specific session in the exam will be assigned or it is limited set of session of the exam to held in are included in this dataset. To reduce the second-order conflict on the same day or overnight is the aim of this dataset. Table 2.4 shows some information of the University of Melbourne examination dataset. The dataset can be downloaded from http://www.or.ms.unimelb.edu.au/timetabling. Furthermore, Merlot et al., (2003), Cote, Wong and Saboun, (2005) have researched the dataset by using bi-objective evolutionary algorithm which tabu search (TS) and variable neighborhood decent (VND) were used.

Table 2.4 University of Melbourne datasets

Problem Instances	Exams	Students	Enrolments	Timeslots
I	521	20656	62248	23
II	526	19816	60637	31

2.4.3 Dataset of University of Nottingham

Burke, Newall and Weare, (1996) had introduced a dataset called Nottingham dataset. In this dataset contain total of 23 timeslots and three timeslots a day (weekdays).The total capacity and no clashing is the constraints that being included in this dataset. Besides that, reduce the number of second order conflicts on the same day is the purpose for this dataset. In table 2.5 will show the information of the University of Nottingham examination dataset and the dataset can be downloaded from http://www.asp.cs .nott.ac.uk/resources/data.shtml. Burke and Newall applying graph heuristic (i.e CD,LD and SD) in year 1999 to reduce the second order of conflicts on the same day. This method also used by Merlot in year 2003 to describe previously to the Nottingham dataset. Burke et al. (2004) try to use great deluge algorithm (GDA)also to reduce the second order conflicts on same day try solve the dataset .

Table 2.5 University of Nottingham datasets

Exams	Students	Enrolments	Conflict Density	Timeslots	Capacity
800	7896	34265	0.03 (3%)	23	1550

2.4.4 Dataset of University Kebangsaan Malaysia (UKM)

Ayob et al., (2007) had introduced a capacitated dataset called UKM dataset. This dataset required to schedule all the exams. No student is able to take more than one exam simultaneously and sitting three continuous exams in a same day are the constraints of this dataset. The room exclusive constraints which are the specific room for exams must be fulfilled and the students must be assigned in the same room for those who sit for consecutive exams. The goal of this dataset is to evenly distributed the exams and reduce the number of students who having consecutive exams in the same day. Table 2.6 shows the information of the UKM dataset and the room capacity of the dataset is shows in table 2.7.

Table 2.6 University Kebangsaan Malaysia datasets (UKM06-01)

Exams	Students	Enrolments	Timeslots	Capacity
818	14047	75857	42	1550

Table 2.7 Room capacity of datasets (UKM06-01)

Room	Room Capacity
DPBestari	850
DGemilang	610
Dewan (DECTAR)	610
LobiUtama (DECTAR)	270
PSeni (DECTAR)	152
LobiA (DECTAR)	70
LobiB (DECTAR)	70

2.4.5 Dataset of University Teknologi MARA (UiTM)

Kendall and Hussin (2004) had introduced a capacitated dataset, UiTM dataset. These dataset are almost same to dataset of UKM which requires all the exams to be scheduled. Besides that, the constraints involved are the first order conflict and coincidence (ie. Exams that required scheduling together should be assigned in the same timeslot.) The goal of this dataset is to spread the exam as even as possible. The spreading of dataset is based on the calculation of the proximity value as in Carter, Laporte and Lee, (1996) and the exams that are scheduled others than weekdays will get a penalty. The examination dataset of UiTM is showed at table 2.8.

Table 2.8 University Teknologi Malaysia (UiTM) dataset

Exams	Students	Enrolments	Timeslots
2063	84675	357761	40

2.4.6 Second International Timetabling Competition (ITC2007) dataset

Second international timetabling competition (ITC2007) is a dataset which can divide into examination timetabling and course timetabling but we will focus on examination dataset in this project. To establish a platform for those researchers to apply their algorithms on real world timetabling problems is the main purpose of ITC2007. There are some constraints that consist in ITC2007 examination set. First, is at the same time, no student can sit more than one exam. Second, the exam capacity cannot exceed the room capacity. Third, the exam time should not violet the timeslot length. Fourth, some specific requirement need to be done in this for example exam A can be assigned after exam B or Exam A must be in room 10 etc. To reduce second-order conflicts in same day, reduce mixed duration of exams within a timeslots, reduce the specific timeslots or room to be used and large examination arrange as early as possible is the aim of this dataset. In McCollum et al., (2008), we can get the information of the examination competition track. There are some researchers who study and investigated this data. McCollum et al., (2009) is one of the researchers who uses iterated forward search, hill climbling and great deluge algorithm. A multistage approach which consist GRASP, simulated annealing and mathematical programming has been used by Gogos, AleFragis and Housos, (2008). A two-phase approach with adaptive heuristic ordering as the constructive phase is applied by McCollum et al., (2009) and the solution is improved by applying an extended great deluge algorithm. Table 2.9 shows the data of the ITC2007 datasets (examination track) and the hard and also the soft constraints of the dataset is showed in table 2.10 and table 2.11 respectively .While, Table 2.12 shows the summary of some of the dataset.

Table 2.9 International Timetabling competition dataset

Instance	Conflict Density (%)	Exams	Students	Periods	Rooms	Period HC	Room HC
Exam-1	5.05	607	7891	54	7	12	0
Exam-2	1.17	870	12743	40	49	12	2
Exam-3	2.62	934	16439	36	48	170	15
Exam-4	15	273	5045	21	1	40	0
Exam-5	0.87	1018	9253	42	3	27	0
Exam-6	6.16	242	7909	16	8	23	0
Exam-7	1.93	1096	14676	80	15	28	0
Exam-8	4.55	598	7718	80	8	20	1
Exam-9	7.48	169	655	25	3	10	0
Exam-10	4.97	214	1577	32	48	58	0
Exam-11	2.62	934	16439	26	40	170	15
Exam-12	18.45	78	1653	12	50	9	7

Table 2.10 Hard constraints of ITC 2007

Hard Constraints	
H1	Student cannot sit more than one exam at the same time.
H2	The exams capacity should not exceed the room capacity.
H3	The exam length should not violate the timeslot length.
H4	A sequence or ordering of an exams must be respected, e.g. schedule Exam A after Exam B;
H5	Schedule exam into specified room (room related hard constraints) e.g. Exam A should schedule to Room 11.

Table 2.11 Soft constraints of ITC2007

Soft Constraints	
S1	*Two exams in a row*: minimize student sitting consecutive exams in the same day.
S2	*Two exams in a day*: minimize student sitting more than two exams in a day (only applied if more than two timeslot per day).
S3	*Spreading of exams*: Each set of student examinations should be spread as evenly as possible over the exam period.
S4	*Mixed duration*: minimize number of exams with different duration that are scheduled into the same room.
S5	*Larger examination schedule late in the timetable:* minimize the number of large exams appear 'late' of the timetable.
S6	*Period penalty*: minimize the number of exams scheduled in period with penalty.
S7	*Room penalty*: minimize the number of exams scheduled in room with penalty.

Table 2.12 Summary of datasets

	Constraints	Toronto	Nottingham	Melbourne	UKM	ITC2007	UMP
Examinations	Clash free	Hard	Hard	Hard	Hard	Hard	Hard
	Scheduled all exams	-	Soft	Soft	Hard		Hard
	Exam preference - Specified arrangement: *sa* - Specified room: *sr* - Large exam schedule first: *lf* - Restriction on exam in particular timeslot: *rt* - Scheduled combined exam in the same timeslots: *ct*	-	-	Hard (*rt*)	-	Hard(*sa*) Soft (*lf*)	-
	Consecutive exam - Two exam in a row:*2r* - Two exam in a day:*2d* - Two exam in a row overnight:*2n* - Three exam in a day:*3d*	-	Soft *(2d & 2n)*	Soft *(2d & 2n)*	Hard *(3d)* *Soft* *(2r)*	Soft *(2r and 2d)*	-
Timeslot Related	Timeslot preference - Minimize/avoid usage :*tu*	-	-	-	-	Soft(*tu*)	-
	Timeslot length - Mixed duration of exams in one timeslot: *mt*	-	-	-	-	Hard Soft(*mt*)	-
	Spreading - Specified spread: *ss*	Soft	Hard *(ss)*	Soft	Soft	Soft(*ss*)	Soft