

## Problem statement

The goal of this project is to write tools to assist in finding schedule conflicts in proposed teaching schedules for UNBC.

### 0.1 A cautionary note

This is a real-life problem. Expect the specification details to be incomplete and to change over time.

### 0.2 An overview of the problem

Scheduling for Fall and Winter courses at UNBC starts approximately a year in advance. For instance, in Fall 2015, the Chair of Computer Science worked with Computer Science faculty members to determine what courses they would teach in Spring 2016, Fall 2016, and Winter 2017.

At the same time as the course workload was being finalized by the Chair and the Dean in conjunction with the faculty, the Registrar's Office was collecting information related to the scheduling of courses.

After collecting this information, they then produce draft schedules which get sent to faculty members, chairs, and academic advisors who review them looking for problematic conflicts, necessary room changes and the like.

The tools to be built in this project are to help review draft schedules. One common way to look for conflicts is to build weekly timetables for one or more resource, and then overlay the timetables. In the past (and possibly currently) some academic advisors have done this by creating multiple coloured transparencies, each with a particular resource, say all of the second-year major Computer Science courses; and then physically overlaying transparencies and seeing what conflicts exist. A goal of this project is to be able to do this "virtually" on a computer. To be more concrete, we want to be able to

1. make weekly timetables, showing say, all of the second-year Computer Science classes, or all of the classes in room 8-164; and
2. overlay timetables;

and we want to be able to do this graphically, preferably with an easy to understand user interface.

### 0.3 Timetables

It isn't completely clear what a timetable should look like. Usually they have days across the top giving column headers, typically Monday to Friday, and they have times down the left side, typically from 8:30 to 21:00. Each class time block is marked by a rectangle, with most time blocks being one of 1h, 1.5h and 3h.

However there are also courses that run early in the morning, and courses that run on weekends; courses that have 2h blocks, and other oddities. It is probably a good design strategy to start with a tool that handles Mon–Fri × 08:30–21:00 timetables only, but to be aware when designing that you may wish to expand to larger timetables.

### 0.4 The overall input file format

The Registrar's Office captures draft schedules in Microsoft Excel files stored in either `.xls` or `.xlsx` format. This is shown in Figure 1.

These can be easily converted to comma-separated-value files (`.csv` files). The latter contain a textual representation of an Excel file with each row being a line in the file, and cells in each row being separated by commas.

There are six kinds of lines in a draft schedule file. They are

- A date header line. There is exactly one of these.
- A line that says “Draft Course Schedule”. There is exactly one of these.
- A line that helps explain whether a particular proposed time is forced (typical first-year courses, but also for some professional programs), or not forced. There is exactly one of these.
- A header line explaining the columns below.
- There is exactly one of these.
- \* The next two kinds of lines have multiple occurrences and are inter-mixed.
  - A forced-time course line.
  - A non-forced time course line.

### 0.5 Individual line formats

**The date header line** This consists of one cell in the 7<sup>th</sup> column. It consists of a month name spelled out, a space, and a four-digit year.

**The Draft Course Schedule line** This consists of one cell in the 7<sup>th</sup> column. It consists of the text phrase “Draft Course Schedule”.

**The forced/unforced line** This consists of two cells, one in the 6<sup>th</sup> column—“Forced Times”, and one in the 9<sup>th</sup> column—“Pattern Times”.

**The header line** This consists of thirteen text cells, from left to right:

- Course ID
- Component ID

January 2017												
Draft Course Schedule												
Forced Times					Pattern Times							
Course ID	Component ID	Start Date	End Date	Day	Start Time	Duration	Pattern Day	Pattern Start Time	Pattern Duration	Building ID	Room Number	Surname
ANTH102	A1	2017-01-04	2017-04-26				Wednesday	11:30	01:30	AGO	7-212	Bouchard
ANTH102	A1	2017-01-04	2017-04-26				Monday	11:30	01:30	AGO	7-212	Bouchard
ANTH102	A1	2017-04-11	2017-04-26									Bouchard
ANTH200	A1	2017-01-04	2017-04-07				Wednesday	16:00	01:30	AGO	7-150	ANTH
ANTH200	A1	2017-01-04	2017-04-07				Monday	16:00	01:30	AGO	7-150	ANTH
ANTH200	A1	2017-04-11	2017-04-26									
ANTH201	A1	2017-01-04	2017-04-26				Wednesday	14:30	01:30	TAL	10-4034	ANTH
ANTH201	A1	2017-01-04	2017-04-26				Monday	14:30	01:30	TAL	10-4034	ANTH
ANTH201	A1	12/30/1899	12/30/1899									ANTH
ANTH211	A1	2017-04-04	2017-04-26									Gibson
ANTH211	A1	2017-01-04	2017-04-26				Monday	18:00	03:00	AGO	7-158	Gibson
BIOL104	A1	2017-01-04	2017-04-07	Friday	11:30	01:00				CFC	6-213	Rea
BIOL104	A1	2017-01-04	2017-04-07	Wednesday	11:30	01:00				CFC	6-213	Rea
BIOL104	A1	2017-01-04	2017-04-07	Monday	11:30	01:00				CFC	6-213	Rea
BIOL104	A1	2017-04-11	2017-04-26									
BIOL104	A2	2017-01-04	2017-04-07	Friday	10:30	01:00				CFC	6-213	Rea
BIOL104	A2	2017-01-04	2017-04-07	Wednesday	10:30	01:00				CFC	6-213	Rea
BIOL104	A2	2017-01-04	2017-04-07	Monday	10:30	01:00				CFC	6-213	Rea
BIOL104	A2	2017-04-11	2017-04-26									
BIOL110	A1	2017-01-04	2017-04-07				Friday	08:30	01:00	LAB8	8-164	Coxson
BIOL110	A1	2017-01-04	2017-04-07				Wednesday	08:30	01:00	LAB8	8-164	Coxson
BIOL110	A1	2017-01-04	2017-04-07				Monday	08:30	01:00	LAB8	8-164	Coxson

Figure 1: Life as we know it

- Start Date
- End Date
- Day
- Start Time
- Duration
- Pattern Day
- Pattern Start Time
- Pattern Duration
- Building ID
- Room Number
- Surname

**Data lines** Both the forced time lines and the unforced time lines have ten non-blank entries, the difference being that the forced time lines have blank entries in columns 8–10, whereas the unforced time lines have blank entries in columns 5–7.

The ten non-blank entries go in the following order:

- A course id (such as “CPSC101”).
- A component id (such as “A1”).
- A start date in yyyy-mm-dd format.
- An end date in yyyy-mm-dd format.
- A day of the week (such as “Friday”).
- A start time in hh:mm 24-hour format.
- A *duration* in hh:mm format.
- A Building ID (such as “LAB8”).
- A room number (such as “10-4074”).
- An instructor surname.

## 0.6 Details of data lines

Here are some points that apply.

- Each line corresponds to one time block of one kind of meeting of one course. For instance, there are three lines corresponding to the lecture part of the Computer Science 101 course, one for each of the three days it is on.
- Some courses are taught online or by directed reading and lack rooms and times. These can be ignored for this project.
- Course ids consist of a four letter code (like “ECON”), followed by a three digit code. For digit codes less than 500, the course is an undergraduate course, and first digit of the code is the year level of the course. 600–799 are Master’s level graduate courses, and 800-999 are PhD level courses.

The course ids are important because a request to contrast second-year Mathematics with third-year Computer Science requires finding the MATH2xx ids and the CPSC3xx ids.

- Following the course id is the component id, which typically consists of a single letter followed by a digit. Common letters are A for lecture, L for lab, and T for tutorial. Multi-session courses like MATH100 have more than one A-code. For a given combination of course id and component id, there are as many lines as there times that the class meets per week.
- The start date and the end date are probably irrelevant to this project. Some professional courses teach in non-standard blocks.
- The start time and duration determine when the class is. Durations are rounded up to the nearest half hour, so a class from 3:30pm to 4:20pm is indicated by 15:30 and 01:00.
- Building ids and room numbers indicate *where* the class is. Current scheduling software does not allow a multi-day course component to be offered in different places. The building id can be deduced from the first part of the room number, and vice versa. The second part of the room number is 3 digits for some buildings, 4 digits for others.
- The instructor surname gives some indication of who has been assigned to teach a course. All capital names like CPSCLAB are place holders. Others are probably real persons' names. It is likely a mistake if the same surname is teaching in two different places at the same time, but there are cases where different people have the same surname.