

Proyecto Horarios

Cota Martínez Guillermo Oswaldo
Macías Gómez Jorge

6 de febrero de 2021

Resumen

Analizamos los datos de las materias del ciclo 2021-1 de la Facultad de Ciencias de la UNAM, para crear un programa que acomode el horario de un alumno, dado su avance escolar y sus intereses para materias optativas.

Vectorizamos los temarios por materias para crear un recomendador de materias según intereses.

1. Documentación Ejecutiva

1.1. Objetivo

Queremos desarrollar una aplicación que pueda ayudar a los estudiantes de la carrera de Matemáticas a elegir sus optativas y sus materias, organizando su horario y considerando los gustos del alumno.

Analizaremos los datos de las materias del ciclo 2021-1 de la Facultad de Ciencias de la UNAM, para crear un programa que acomode el horario de un alumno, dado su avance escolar y sus intereses para materias optativas.

1.2. Introducción.

En la facultad de ciencias existen mas de 100 materias diferentes cada una con entre 1 y 15 horarios diferentes, por lo que puede resultar difícil armar un horario para un semestre.

Creamos un programa que recomienda horarios según las materias ya cursadas, los gustos del alumno y los horarios.

1.3. Nuestros datos.

Tomamos los horarios de la carrera de Matemáticas de la Facultad de ciencias para el semestre 2021-1, un único documento con los horarios del semestre en formato pdf, que pasamos a un txt. Tomamos 154 temarios de materias de la carrera de Matemáticas de la Facultad de Ciencias, los temarios inicialmente estaban en formato pdf. Los convertimos a formato txt con ayuda de Python, esto para facilitar el trabajo que realizaremos con el texto, cada uno de estos temarios contiene información cómo la siguiente:

- Nombre de la materia
- Clave de la materia

- Semestre
- Optativo o Obligatorio
- Asignaturas subsecuentes
- Asignaturas Antecedentes
- Objetivos
- Lista de Unidades temáticas
- Bibliografía

De estos los primeros 4 nos ayudaran a clasificar a las materias, mientras que los últimos 5 nos ayudaran a identificar los temas de las materias para lograr dar una mejor recomendación de materias según los temas de interés de los alumnos.

Elegimos estos datos por ser nuestra facultad de origen, teníamos facilidad de acceso a los horarios de la facultad por esta misma razón, consideramos que era la mejor opción.

1.4. Recomendador de asignaturas

Buscamos extraer algún tipo de características de los datos de texto anteriores para poder calcular la similitud y / o disimilitud entre ellos y usaremos un vector para representar los textos, al que llamaremos vector de palabras.

Como sugiere el nombre, los vectores de palabras son representaciones vectorizadas de palabras en un documento. Los vectores llevan consigo un significado semántico. Por ejemplo, el hombre y el rey tendrán representaciones vectoriales cercanas entre sí, mientras que el hombre y la mujer tendrán representaciones alejadas entre sí.

Calcularemos los vectores de frecuencia de cada documento. Esto nos dará una matriz donde cada

columna representa una palabra en el vocabulario general (todas las palabras que aparecen en al menos un documento), y cada columna representa un temario.

Usamos la matriz para crear una matriz de similitudes, creamos un método que recibe una serie de palabras o intereses y aplicamos un análisis sobre nuestros datos para regresar una lista de materias que van acorde a los intereses recibidos.

1.5. Uso de recursos

Estamos creando estructuras de datos que almacenaran los datos de los horarios, grupos y materias del plan, esto implica un uso de memoria proporcional a la cantidad de grupos y horarios. En el caso del uso de tiempo, estamos probando combinaciones hasta encontrar una combinación valida. Dados los tamaños de nuestros datos, es de un orden de minutos.

2. Conclusiones

Logramos crear un recomendador de materias y un estructurador de horarios que te sugiere materias según tus intereses, permite elegir que materias de la recomendación quieres acomodar en un horario, es posible expandir el proyecto a mas carreras y/o facultades. Logramos desarrollar con éxito las ideas de nuestro proyecto sin dificultades.

3. Documentación Técnica

3.1. Creador de horarios

Una vez que el documento de los horarios del semestre que usaremos fue parseado a un texto plano, lo usamos en conjunto del documento "materias" para dividir los horarios en semestres y optativas. Creamos una función que crea un diccionario {clave_materia: nombre_materia} Para poder acceder en $O(1)$ a las materias segun su clave.

Usamos expresiones regulares y las funciones de análisis de patrones para parsear la información contenida en los temarios en un diccionario que se ve:

```
1 {'clave': '0007',
2  'materia': 'algebra superior i',
3  'grupo': '4000',
4  'profesor': 'eugenia marmolejo rivas',
5  'horario': {'lu': [[9.0, 10.0]],
6              'ma': [[9.0, 10.0]],
7              'mi': [[9.0, 10.0]],
8              'ju': [[9.0, 10.0]],
9              'vi': [[9.0, 10.0]],
10             'sa': []}
11 }
```

Una vez hecho esto, se proceden a crear una clase para cada materia, y asignarlo a un diccionario donde la llave es la clave y el valor, el objeto. Esto es, que se creó una clase `materia` cuyos objetos son las claves de cada asignatura, así como se hizo un loop sobre cada grupo

para agruparlo con su correspondiente materia y a su vez, agregarlo a un diccionario con cada materia por día y hora: Ejemplo, el objeto `dict_materias['0007']` representa el objeto *álgebra superior i* y en el objeto, accediendo a `dict_materias['0007'].grupos` se accede a la información de cada grupo de la materia, a su vez, `dict_materias['0007'].horarios` regresa los horarios de dicha materia por hora y día, así como qué grupo está en cada horario.

De igual manera, para continuar con el programa, se establecieron algunas funciones que realizan lo siguiente:

- Verificar si una hora de algún día en específico se cruza con un conjunto de horas de ese mismo día.
- Dados dos conjuntos de horas de un día, verificar si ambos conjuntos se intersectan en algún intervalo de tiempo.
- Dados varios conjuntos de horas a lo largo de varios días, verificar si son disjuntos dos a dos.

Finalmente, se traduce un conjunto de grupos a conjuntos de horas y con ayuda de la última función se predice si el horario es aceptable o no.

3.2. La matemática del algoritmo

Ahora discutiremos un poco del problema a resolver algorítmicamente:

Dado un conjunto de materias a cursar, decidir si hay alguna combinación de grupos (uno para cada materia) en el cual no se intersecte ningún horario y de ser posible, dar alguna combinación.

Es un problema que se abordó haciendo uso de los siguientes elementos:

- **Estructuras de datos que permiten acceder a la información en tiempo constante.** Para esto, hicimos amplio uso de los diccionarios, pues como cada materia se identifica unívocamente con su clave, y a su vez cada grupo de cada materia se identifica unívocamente con su código de grupo, entonces optamos por hacer amplio uso de esta estructura de datos, pues cada vez que queríamos hacer una consulta de la información de una materia, bastó con introducir su clave y se obtuvo todo. Enseguida, para obtener la materia de cada grupo, es similar que lo descrito anteriormente, pues se propuso una función que convierte de grupo a materia en tiempo constante (igual usando diccionarios) y se prosiguió a obtener la materia de ese modo. Después para conseguir el horario de cada grupo para cada materia, se logra en tiempo lineal, donde la linealidad está en el número de grupos en total, pero conseguir el horario de un grupo en específico se hace en tiempo constante, nuevamente con ayuda de los diccionarios.

- **Exploración Exhaustiva** El algoritmo de crear horarios, que es la función 'mágica' que hace que

nuestro algoritmo funcione, recibe las claves de las materias que queremos y enseguida hace una exploración exhaustiva sobre las combinaciones de elegir un grupo de cada materia. Esto no es lo más óptimo y se consideró hacer un backtrack sobre los horarios pero se hizo la siguiente consideración coherente sobre el problema: EL horario está pensado para alumnos de la licenciatura de matemáticas donde cada alumno tiene en promedio 4 materias por semestre, además de que en el peor de los casos, se tendrían que acomodar 12 materias, pues es el máximo de materias permitidas por la facultad. Entonces el recorte de estados que se hizo fue considerar únicamente un grupo por materia y elegir exactamente tantos grupos como materias, de manera que a grande número de grupos a considerar, se reduce a combinaciones de grupos con materias distintas. Este algoritmo es el correcto, pues prueba cada combinación de materias, en caso de siempre tener conjuntos no disjuntos dos a dos, se le informa al usuario que no hay horarios que se puedan crear.

Estos métodos nos permiten tener una complejidad de espacio del orden de $O(m + g)$ donde g son los grupos que existen, m las materias (hay más grupos que materias) pues guardamos diccionarios que tienen materias y diccionarios que contienen grupos. Adicionalmente si se considera el notebook en donde se hace la extracción (Extraccion.ipynb) se tiene una complejidad del orden de los grupos, pues el parser hace una variable para cada grupo. Finalmente este es omitido, pues ese notebook contiene las instrucciones para guardar el diccionario de los grupos en su propio archivo distinto que finalmente es leído por el programa principal. En tiempo se tiene una complejidad en esperanza de $O(n^4)$ donde n es el promedio de grupos por materia y está elevada a la cuarta potencia por la consideración de que la gran mayoría de los alumnos meten 4 materias. Pero con el análisis del peor de los casos, tiene complejidad $O(n^{12})$ con las 12 materias que se pueden meter como máximo, sin embargo este es un escenario prácticamente irreal para el usuario final.

4. Resultados

Un ejemplo de la ejecución es la siguiente:

Hola este es el súper creador
de horarios 9000!

Elige materias de algún semestre:

```
1 -> Primer Semestre
2 -> Segundo Semestre
3 -> Tercer Semestre
4 -> Cuarto Semestre
5 -> Quinto Semestre
6 -> Sexto Semestre
7 -> Optativas
```

8 -> Recomendador de optativas

Eliga algún número: 1

Las materias del 1° semestre son:

```
0 -> algebra superior i
1 -> calculo diferencial e integral i
2 -> geometria analitica i
3 -> geometria moderna i
```

Por favor usa los indices para elegir que desees cursar:

Elige las materias que desees tomar,
separados por comas: 0,1

.-.-.-.-.-.-.-.-.-.-.

¿Quiere seguir eligiendo materia? [S/N]s

Hasta ahora has seleccionado

algebra superior i

calculo diferencial e integral i

Claves: ['0007', '0091']

Elige materias de algún semestre:

```
1 -> Primer Semestre
2 -> Segundo Semestre
3 -> Tercer Semestre
4 -> Cuarto Semestre
5 -> Quinto Semestre
6 -> Sexto Semestre
7 -> Optativas
8 -> Recomendador de optativas
```

Eliga algún número: 8

Escibe algunos intereses que tengas:

teorema limite central

Buscaremos horarios para estos intereses:

teorema limite central

.-.-.-.-.-.-.-.-.-.-.

El algoritmo sugiere las siguientes materias para ti:

```
0 -> probabilidad i
1 -> calculo diferencial e integral i
2 -> ecuaciones diferenciales iii
3 -> geometria moderna i
4 -> procesos estocasticos ii
5 -> sistemas dinamicos discretos i
6 -> variable compleja i
7 -> geometria diferencial i
8 -> calculo diferencial e integral ii
```

Por favor usa los indices para elegir
que desees cursar:

Elige las materias que desees tomar,
separados por comas: 0

.-.-.-.-.-.-.-.-.-.-

Elegiste las siguientes opciones: 0
¿Quiere seguir eligiendo materia? [S/N]n

.-.-.-.-.-.-.-.-.-.-

Crearemos un horario en el que entres
a las: 8 y salgas a las: 20

Creamos el siguiente horario con los
siguientes grupos:

```

~~.~~.~~.~~.~~.~~.~~.~~.~~.~~.~~.
Calculo diferencial e integral i
Grupo 4040 Profesor(a):Kenya Veronica
Espinosa Hurtado
~~.~~.~~.~~.~~.~~.~~.~~.~~.~~.~~.
Algebra superior i
Grupo 4005 Profesor(a):Francisco
Larrion Riveroll
~~.~~.~~.~~.~~.~~.~~.~~.~~.~~.~~.
Probabilidad i
Grupo 9024 Profesor(a):Maria
Asuncion BegonA Fernandez Fernandez
.-.-.-.-.-.-.-.-.-.-

```

Lunes -> (17.0, 19.0) -> calculo
diferencial e integral i
Lunes -> (9.0, 10.0) -> algebra superior i
Lunes -> (14.0, 15.0) -> probabilidad i

Martes -> (17.0, 18.0) -> calculo
diferencial e integral i
Martes -> (9.0, 10.0) -> algebra superior i
Martes -> (14.0, 15.0) -> probabilidad i

Miercoles -> (17.0, 19.0) -> calculo
diferencial e integral i
Miercoles -> (9.0, 10.0) -> algebra superior i
Miercoles -> (14.0, 15.0) -> probabilidad i

Jueves -> (17.0, 18.0) -> calculo
diferencial e integral i
Jueves -> (9.0, 10.0) -> algebra superior i
Jueves -> (14.0, 15.0) -> probabilidad i

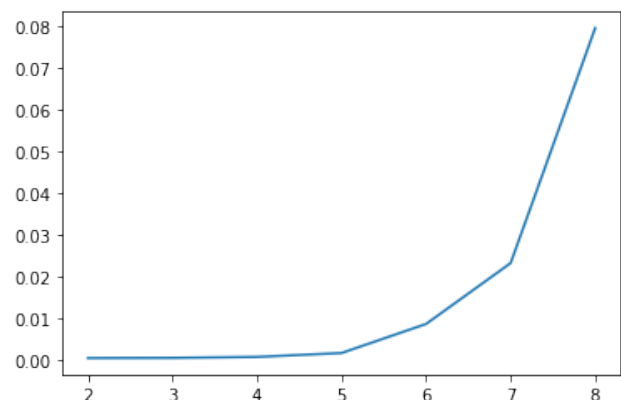
Viernes -> (17.0, 19.0) -> calculo
diferencial e integral i
Viernes -> (9.0, 10.0) -> algebra superior i
Viernes -> (14.0, 15.0) -> probabilidad i

Sabado -> (7.0, 8.0) -> calculo
diferencial e integral i

Como vemos, nuestro recomendador de horarios tiene una buena recomendación y además en tiempos de ejecución se hicieron algunas pruebas que demuestran su desempeño:

- Número de materias: 2,
Tiempo promedio: 0.000406739234924
- Número de materias: 3,
Tiempo promedio: 0.0004784350395202
- Número de materias: 4,
Tiempo promedio: 0.001254510879516
- Número de materias: 5,
Tiempo promedio: 0.001463096141815
- Número de materias: 6,
Tiempo promedio: 0.00411743736267
- Número de materias: 7,
Tiempo promedio: 0.0582787427902
- Número de materias: 8,
Tiempo promedio: 0.0976520895957

Figura 1. Ejecución según número de materias



El tiempo en segundos, demuestra el tiempo promedio en encontrar horario para 500 combinaciones distintas para cada cálculo. De esta manera vemos que tiene un buen desempeño para el usuario promedio

5. Conclusiones

Para el caso de estudio, este algoritmo resultó efectivo y capaz de realizar la tarea, algunas mejoras para futuros usos sería el de estandarizar los datos de entrada para acoplarlo no solamente a la facultad de ciencias, si no en distintos casos, pero por la complejidad del scrapping del texto, resulta una tarea ardua. Algo

que notar de este algoritmo es que, a juzgar por los distintos horarios, vienen en el mismo formato, por lo que no solo sería efectivo para el ciclo 2021-1, sino para futuros usos.

Referencias

- [1] NOTICIAS(Cuentame INEGI).<http://cuentame.inegi.org.mx/monografias/informacion/bc/territorio/clima.aspx?tema=mee=02>
- [2] NOTAS DE CLASE