

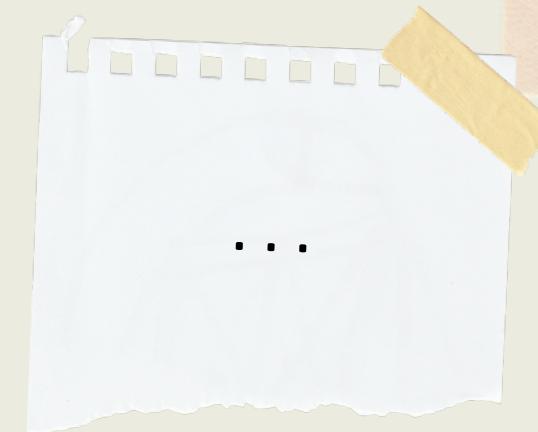
Patrones

secuenciales

Equipo 8

# ¿En qué consiste?

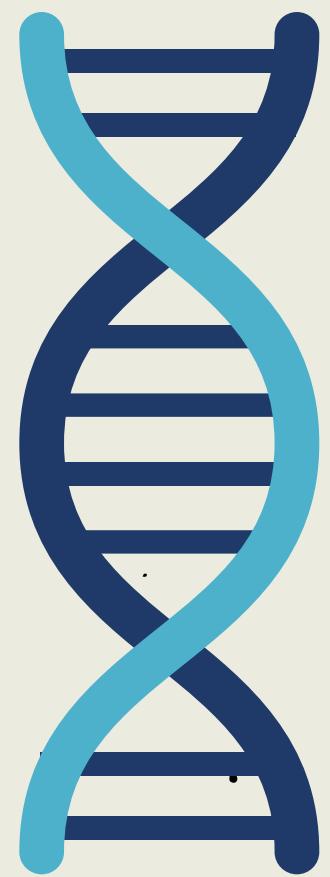
Consiste en encontrar patrones estadísticamente relevantes en colecciones de datos que están representados de forma secuencial. Debido la frecuencia con que aparecen este tipo de datos en escenarios de aplicaciones reales, esta técnica constituye uno de los métodos más populares de descubrimiento de patrones.



# ¿Qué es un patrón secuencial?

Es la búsqueda de asociaciones de la forma: "*si sucede el evento X en el instante de tiempo t entonces sucederá el evento Y en el instante t+n*".





## Ejemplos

- Secuencias genéticas como las de ADN y ARN.
- Historial de compras de un cliente determinado
- Navegación de un visitante del sitio web



# Características

- Objetivo: encontrar patrones en secuencia
- Una secuencia es una lista ordenada de itemsets, donde cada itemset es un elemento de la secuencia
- El tamaño de una secuencia es su cantidad de elementos (itemsets)
- La longitud de una secuencia es su cantidad de items
- El soporte de una secuencia es el porcentaje de secuencias que la contienen en un conjunto de secuencia S
- Las secuencias frecuentes (o patrones secuenciales) son las subsecuencias de una secuencia que tienen un soporte mínimo.



# Extracción de patrones secuenciales

Dado un una base de datos de secuencias y un umbral de soporte mínimo , encontrar el conjunto completo de subsecuencias frecuentes

## Base de datos de secuencias

ID	Secuencia
10	<a( <b>abc</b> )(ac)d(cf)>
20	<(ad)c(bc)(ae)>
30	<(ef)( <b>ab</b> )(df) <b>cb</b> >
40	<eg(af)cbc>

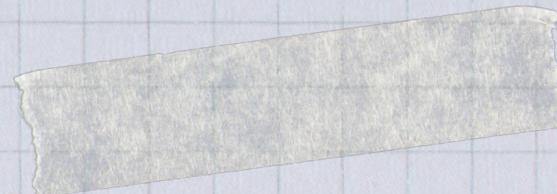
- Una **secuencia** es: <(ef) (ab) (df) c b>
- Un **elemento** puede contener un conjunto de items( también llamados eventos)
  - Los items dentro de un elemento pueden estar desordenados, por conveniencia los ordenaremos alfabeticamente.
- <**a(bc)dc**> es una **subsecuencia** de <**a(abc)(ac)d(cf)**>

Dado un umbral de **soporte mínimo** min\_sup=2, <(ab)c> es un **patron secuencial**

# Algoritmos representativos

- GSP (Generalized Sequential Patterns)
- APrioriAll
- APrioriSome
- DynamicSome
- PrefixSpan
- SPAM (Sequential PAttern Mining).
- SPADE (Sequential PAttern Discovery using Equivalence classes)

# Ejemplo



# Preguntas

¿Qué es una secuencia?

Pregunta 1

¿Cuál es el objetivo de esta técnica?

Pregunta 4

¿Por qué es importante el tiempo?

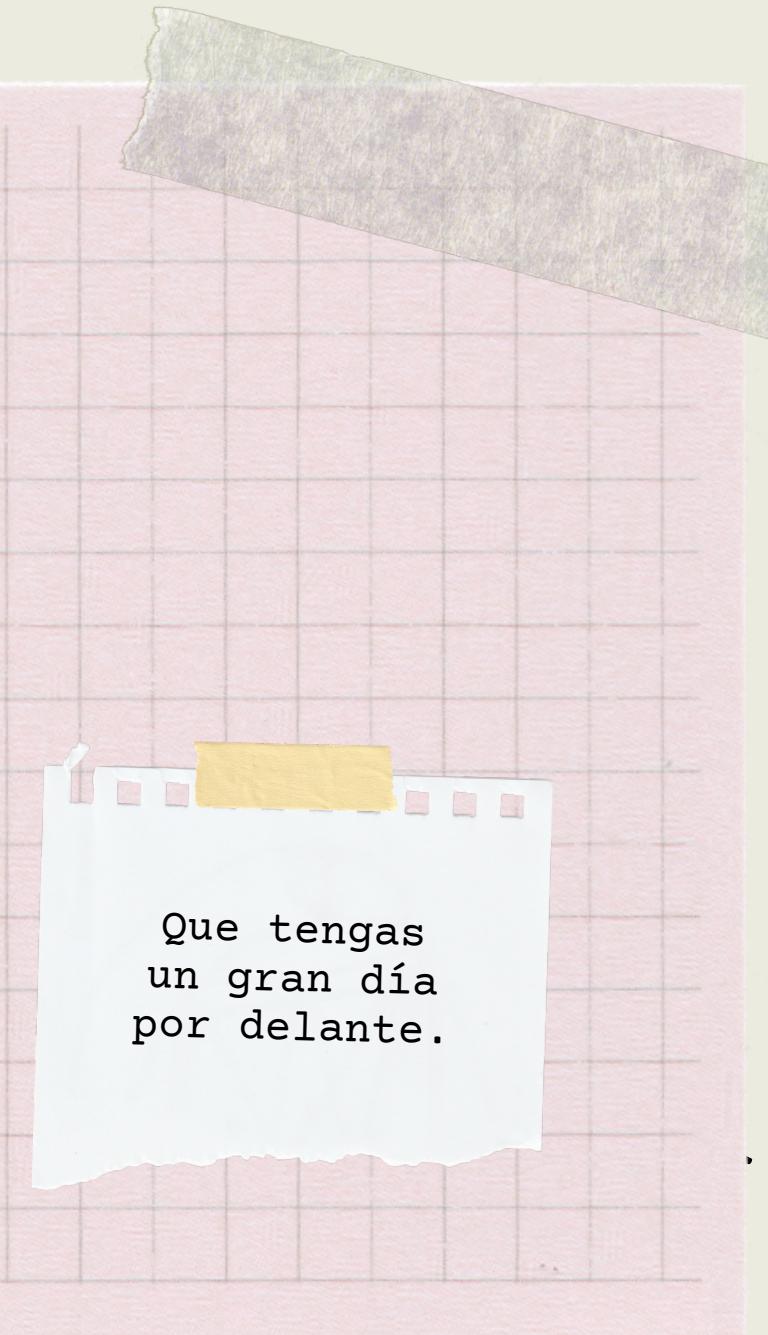
Pregunta 2

Menciona dos ejemplos de patrones secuenciales

Pregunta 3

Pregunta 5

!Muchas  
gracias!



# Bibliografía

1

Mineria De Datos Secuenciales  
(slideshare.net)

2

<https://amp.ww.es.freejournal.org/5543174/1/mineria-de-secuencias.html>

3

<https://elvex.ugr.es/idbis/dm/slides/22%20Pattern%20Mining%20-%20Sequences.pdf>

4

<https://repositorio.uniandes.edu.co/bitstream/handle/1992/20214/u672297.pdf?sequence=1>

