

# MINADO Y ANÁLISIS DE ASOCIACIONES DE PALABRAS

Salvador López Mendoza

Mayo de 2018

# RELACIONES BÁSICAS

- Paradigmática.

Las palabras A y B tienen una *relación paradigmática* si pueden ser sustituidas la una por la otra.

*gato* y *perro*; *lunes* y *sábado*.

- Sintagmática.

Las palabras A y B tienen una *relación sintagmática* si pueden ser combinadas la una con la otra.

*perro* y *comer*; *auto* y *conducir*.

Son relaciones complementarias.

# MINADO DE ASOCIACIONES

- Mejora la precisión de muchas tareas de procesamiento de lenguaje natural.
- Se usa directamente en muchas aplicaciones de minería y recuperación de textos.
  - Recuperación de textos. Se usa la asociación de palabras para sugerir variaciones en las consultas.
  - Construcción automática de mapas de tópicos para navegación. Las palabras son los nodos y las asociaciones las aristas.
  - Comparar y resumir opiniones. ¿Qué palabras tienen relaciones más fuertes con la palabra *batería* en las reseñas del Galaxy 7?

# INTUICIÓN

Relaciones paradigmáticas: contextos similares.

Considerar las siguientes frases:

*Mi **gato** come pez los sábados.*

*Su **gato** come pavo los martes.*

*Mi **perro** come carne los domingos.*

*Su **perro** come pavo los martes.*

¿Qué tan similar es el contexto **gato** y el contexto **perro**?

¿Qué tan similar es el contexto **gato** y el contexto **auto**?

## INTUICIONES (II)

Relaciones sintagmáticas: ocurrencias correlacionadas.

Considerar las siguientes frases:

*Mi gato come pez los sábados.*

*Su gato come pavo los martes.*

*Mi perro come carne los domingos.*

*Su perro come pavo los martes.*

¿Qué palabras pueden aparecer a la izquierda de *come*?

¿Qué palabras pueden aparecer a la derecha de *come*?

Cada vez que aparece la palabra *come*, ¿qué otras palabras pueden aparecer?

¿Qué tanto ayuda la aparición de la palabra *come* para predecir la aparición de la palabra *carne*? ¿y para predecir la aparición de la palabra *texto*?

# IDEAS GENERALES

- Paradigmática.
  - Se representa cada palabra por su contexto.
  - Se calcula la similaridad de los contextos.
  - Las palabras cuya similaridad de contextos sea grande pueden tener relación paradigmática.
- Sintagmática.
  - Contar cuántas veces aparecen juntas dos palabras en un contexto.
  - Comparar sus co-ocurrencias con sus apariciones individuales.
  - Aquellas palabras que tengan co-ocurrencias grandes pero apariciones individuales relativamente bajas pueden tener relaciones sintagmáticas.
- Aquellas palabras que están relacionadas paradigmáticamente tienden a tener relaciones sintagmáticas con la misma palabra.  
Se necesita buscar a la vez las dos relaciones.