

Programa 1:

Modelo Booleano de Recuperación de Información

Objetivo: Implementar el modelo booleano de recuperación de información en el lenguaje de programación Python.

- 1.- Leer m documentos (corpus) desde algún repositorio local o remoto.
- 2.- Generar el diccionario de términos de todo el corpus
(Generar un documento de texto donde se guarde el diccionario)
- 3.- Aplicar eliminación de palabras vacías (stopword)
(Generar un documento de texto)
- 4.- Aplicar una técnica de stemming para reducir el “tamaño” de las palabras.
- 5.- Obtener una matriz binaria donde cada fila sea la representación de un documento formado por n “términos” (del total de la colección). Recordar que si el término se encuentra dentro del documento se representa con 1 y si no está, entonces con 0.
(Generar un documento de texto donde se pueda ver esta matriz de stems)
- 6.- Diseñar una tabla hash que permita obtener mediante su función lo menos posibles colisiones. Posteriormente, almacenar la información dentro de la tabla para obtener acceso a las consultas de “términos (stems)” lo más rápido posible.
- 7.- Leer la consulta booleana Q
- 8.- Aplicar stopwords y stemming a la consulta Q
- 9.- Aplicar la notación postfijo para el procesamiento de recuperación para la consulta dada.(mostrar en el mismo archivo los resultados de los pasos 7,8,9)

10.-Presentar los nombres de los documentos obtenidos por Q.
(mostrar en un archivo los nombre de los documentos recuperados aparte de la consola o interfaz de usuario).

REALIZAR UN REPORTE EN LATEX, DONDE SE ESPECIFIQUE:

- 1.- PORTADA
- 2.- EL OBJETIVO DEL PROGRAMA,
- 3.- METODOLOGÍA
 - 3.1.- Diagrama de módulos
 - 3.2.- Descripción de cada uno de los módulos del programa(no agregar código).
- 4.- RESULTADOS
- 5.- CONCLUSIONES DE ACUERDO A SUS OBSERVACIONES.
- 6.- APÉNDICES:
 - A. CONCEPTOS BÁSICOS
 - B. MANUAL DE USUARIO
 - C. MANUAL TÉCNICO
 - D. CÓDIGO (BIEN COMENTARIADO)

FECHA DE ENTREGA:

VIERNES 29 DE SEPTIEMBRE