

ACCESO INTELIGENTE A LA INFORMACIÓN

EJERCICIO WHOOSH I

Tenemos un directorio con una lista de correos electrónicos. Los nombres de los ficheros son números (p.e. 34.txt). Cada fichero contiene un correo y tiene la siguiente estructura:

- 1º línea: e-mail del remitente
- 2ª línea: e-mails de los destinatarios, separados por espacios en blanco.
- 3ª línea: fecha del correo (en formato YYYYMMDD)
- 4ª línea: Asunto del correo.
- 5ª línea y posteriores: Cuerpo del correo.

Tenemos otro directorio que contiene un fichero llamado agenda.txt que tiene los nombres y apellidos y los e-mails de la agenda. El formato del fichero es el siguiente:

- 1º línea: e-mail(1)
- 2ª línea: Nombre y Apellidos(1)
- 1º línea: e-mail(2)
- 2ª línea: Nombre y Apellidos(2)
- ...

Usando Whoosh, crear un esquema para este tipo de documentos y crear un índice para poder llevar a cabo consultas.

Se desea construir un programa en Python, usando Whoosh, que permita al usuario hacer las siguientes consultas:

- a) Mostrar los remitentes y asuntos de todos los correos que incluyan en el Asunto o el Cuerpo la consulta de usuario. Los remitentes se presentan con Nombre y Apellidos.
- b) Mostrar los e-mails de remitentes, e-mails de destinatarios y asuntos de los correos que sean posteriores a una fecha. La fecha la introduce el usuario con el formato YYYYMMDD.
- c) Se desea mostrar los nombres de los ficheros que contienen correos spam. Los correos spam son aquellos que contienen en su Asunto alguna de las palabras que el usuario introduce. Por ejemplo: 'Contrato Gracias compraventa'