

# Informe práctica 2

## Programa de Indexación

### Consultas espaciales

Para crear los índices que permitan realizar las consultas espaciales se añaden al documento clases de tipo *"DoublePoint"* con las etiquetas de las cuatro coordenadas obtenidas de los campos *"LowerCorner"* y *"UpperCorner"*. Para eso se adaptarán las funciones realizadas en la práctica anterior. De esta forma tenemos campos *"North"*, *"East"*, *"West"* y *"South"* de tipo *"DoublePoint"* con sus respectivos valores.

### Consultas temporales

Para el segundo apartado se buscan añadir consultas temporales, empleando los campos *"created"* e *"issued"*, con las fechas correspondientes. Antes de añadir las clases *"StringField"* correspondientes a estos valores, se convierten los valores a cadenas de texto sin guiones, de esta forma, evitar problemas con el Analyzer.

### Necesidades de información

Para el tercer apartado emplearemos los índices creados hasta ahora en los apartados previos y en la práctica anterior.

# Programa de Búsqueda

## Consultas espaciales

Tras declarar el nombre *Spatial*: para realizar la consulta y obtener así las coordenadas indicadas, mediante splits se analiza el texto y se asigna cada uno de los “trozos” obtenidos a oeste, este, sur y norte respectivamente, dado que todas las coordenadas están en el mismo orden, podemos saber cual es cual. Luego se añaden los rangos a la *Query* y se realiza la operación necesaria para calcular si el área de la consulta coincide en algún punto con el área que indica el documento.

## Consultas temporales

En este caso, el programa de indexación ya se encarga de realizar todo lo necesario.

## Necesidades de información

Para obtener el contenido del fichero “*necesidadesInformacion.xml*”, se han empleado las clases *documento* y *documentBuilder* para obtener el texto e identificador del fichero, aprovechando que este tenía extensión “*xml*”.

Para crear las consultas a partir de las necesidades se han utilizado distintas expresiones regulares y, se ha empleado la herramienta POS Tagger para obtener los tipos de cada elemento de las necesidades. Con esto preservar nombres adjetivos y números relevantes y además eliminar toda la información entre paréntesis ya que no se considera relevante.

## Resultados Zagan para las Necesidades de Información

Tras realizar varias pruebas e ir cambiando las expresiones regulares en torno a si consideramos la información suficiente, excesiva o que no era comprensible; estos son los resultados que hemos obtenido.

1) ¿Qué trabajos o tesis existen sobre el caciquismo, la dictadura, o la represión política relacionados con Huesca, o en su defecto España, y centrados si es posible en el siglo XX?

**Consulta: trabajos tesis caciquismo, dictadura, represión política relacionados Huesca, defecto España, centrados posible siglo XX?**

2274 total matching documents

2) ¿Qué trabajos relacionados con la Ingeniería Biomédica se han desarrollado preferentemente en un Departamento de Ingeniería sobre el diagnóstico de enfermedades neurodegenerativas? Ejemplos de estas enfermedades son el Alzheimer o el Parkinson.

**Consulta: trabajos relacionados Ingeniería Biomédica Departamento Ingeniería diagnóstico enfermedades neurodegenerativas? Ejemplos enfermedades Alzheimer Parkinson.**

2181 total matching documents

3) Análisis de la evolución económica de España desde la crisis de 2008 hasta 2019 en relación con los diferentes partidos políticos que han gobernado el país durante este periodo o con cualquier otro aspecto de relevancia social.

**Consulta: Análisis evolución económica España crisis 2008 2019 relación partidos políticos país periodo aspecto relevancia social.**

2270 total matching documents

4) Estoy interesado en trabajos académicos sobre Bioinformática (también conocida como Biología Computacional, Bioinformatics o Computational Biology) o Filogenética (Phylogenetics), publicados entre 2010 y 2018.

**Consulta: interesado trabajos académicos Bioinformática publicados 2010 2018.**

2475 total matching documents

5) Busco Trabajos de Fin de Grado de los últimos 10 años en español. Deben ser del campo de la informática (Computer Science) realizados o dirigidos por alguien llamado Javier.

**Consulta: Busco Trabajos Fin Grado 10 años español. campo informática realizados dirigidos llamado Javier.**

3188 total matching documents

Podemos observar como para la consulta se han conservado nombres, adjetivos y números, se ha eliminado el resto de elementos, y los comentarios entre paréntesis en la necesidad.

Podemos observar que para estas consultas se devuelven todos los documentos que contengan algunos de estos términos. Lo que puede llevar a entregar documentos no relevantes. Para ordenar los resultados, y entregar primero los resultados más relevantes se ha empleado el algoritmo por defecto Okapi BM25.