

USO DE MACHINE LEARNING EN ACTUACIONES SUPERVISORAS:

Herramienta de Tratamiento y Revisión de EXpedientes (TyREX)

Seminar on the Applications and Development of
Big Data and Data Science in Central Banking

03 de Junio de 2021



1. Actuaciones supervisoras de Conducta

**2. Herramienta de Tratamiento y Revisión de
EXpedientes (TyREX)**

3. Conclusiones y lecciones aprendidas

CICLO DE SUPERVISIÓN



- Inspecciones y Actividades Supervisoras

- Hipotecas / Consumo / Tarjetas
- Formalización del préstamo
- Términos contractuales
- Condiciones financieras
-



Obtención de Evidencia

Basada en:

- Análisis de procedimientos y controles internos
- Análisis de inventario de operaciones
- Análisis muestral de expedientes

Desafío:

- Gran cantidad de documentación a revisar
- Incremento de tamaño de muestras de expedientes con diversidad de casuísticas y puntos de revisión.
- Recursos limitados y plazos ajustados

- **Oportunidad de automatización: Revisión cláusulas suelo**
 - Dificultad de manejo de escrituras de préstamos hipotecarios.
 - Clausulado poco estandarizado
 - Necesidad de análisis de mayores tamaños de muestra
- **Requerimientos:**
 - **Entrenamiento:**
 - *Marcaje de la información relevante para entrenamiento*
 - *Codificación de la información asimilable por los algoritmos*
 - **Presentación y validación resultados:**
 - *Validación manual de las detecciones*
 - *Trazabilidad de los marcajes y detecciones*
- Prueba de concepto de la herramienta de

Tratamiento y Revisión de Expedientes

Árbol de navegación: Located on the left side of the interface, showing a hierarchical list of document elements and their associated file names.

Formulario de validación: A modal window titled 'Validación del Dato' that appears over the main text, containing a text area and a 'Delete' button.

Expresiones de Búsqueda: A table listing search codes and their corresponding results.

Marcajes de información: A table listing information tags and their corresponding results.

Capa de texto: A callout pointing to a specific line of text in the main document view.

Opciones de filtrado: A panel on the right side of the interface containing various filters and checkboxes for document processing.

Opciones de búsqueda avanzada: A section at the bottom right of the interface providing advanced search options like 'Búsqueda en el documento actual' and 'Búsqueda Difusa'.

Información del documento: A panel at the bottom of the interface displaying document statistics such as page, line, and word counts.

Código	Res.
BOW.DIGIT	1620760
SUB_FUZZY.VAL_INTERES	798743
SUB.VAL_INTERES	678599
FUZZY.BOW.NUMERO	825258
IMP.LIJJ	55427
BOW.NUMERO	149489
COMP	644
FUZZY.BOW.BAIADA	15246
IMP.LETRAS	25559
FUZZY.BOW.%	26293
SUB.TIPO_DE_INTERES	25562

Tipo	Resultado
Adv. NO Suelo	1
Clausula Suelo	175
Novación Suelo	159
Tipo Novación	77
Tipo Nuevo	117
Tipo Subrogación Novativa	1
Tipo Subrogación	38

Doc: 4 | Pág: 16 | Line: 14 | Word: 50,215 | Pos: 130,825

2. TRATAMIENTO Y REVISIÓN DE EXPEDIENTES -TYREX

NAVEGACIÓN POR LOS DOCUMENTOS

The screenshot displays the TyREX 2.0 interface. On the left, the 'Árbol de navegación' (Navigation Tree) lists document elements (001-015) with their respective file names. The central 'Área de visualización PDF' (PDF Visualization Area) shows a document page with text in a monospaced font. On the right, there are search and filter options, including a 'Regex' table and a 'Marcjes' table. At the bottom, the 'Información del documento' (Document Information) bar shows document statistics and file details.

Árbol de navegación

Área de visualización PDF

Información del documento

Código	Res.
<input type="checkbox"/> BOW.DIGIT	1620760
<input type="checkbox"/> SUB_FUZZY.VAL_INTERES	798743
<input type="checkbox"/> SUB.VAL_INTERES	678599
<input type="checkbox"/> FUZZY.BOW.NUMERO	825258
<input type="checkbox"/> IMP.LIJN	55427
<input type="checkbox"/> BOW.NUMERO	149489
<input type="checkbox"/> COMP	644
<input type="checkbox"/> FUZZY.BOW.BAJADA	15246
<input type="checkbox"/> IMP.LETRAS	25559
<input type="checkbox"/> FUZZY.BOW.%	26293
<input type="checkbox"/> SUB.LIMITE	7319
<input type="checkbox"/> SUR_FUZZY.LIMITF	7586
<input type="checkbox"/> FUZZY.BOW.SUELO	15007
<input type="checkbox"/> SUB_FUZZY.TIPO_DE_INTERES	31071
<input type="checkbox"/> SUB.TIPO_DE_INTERES	25562

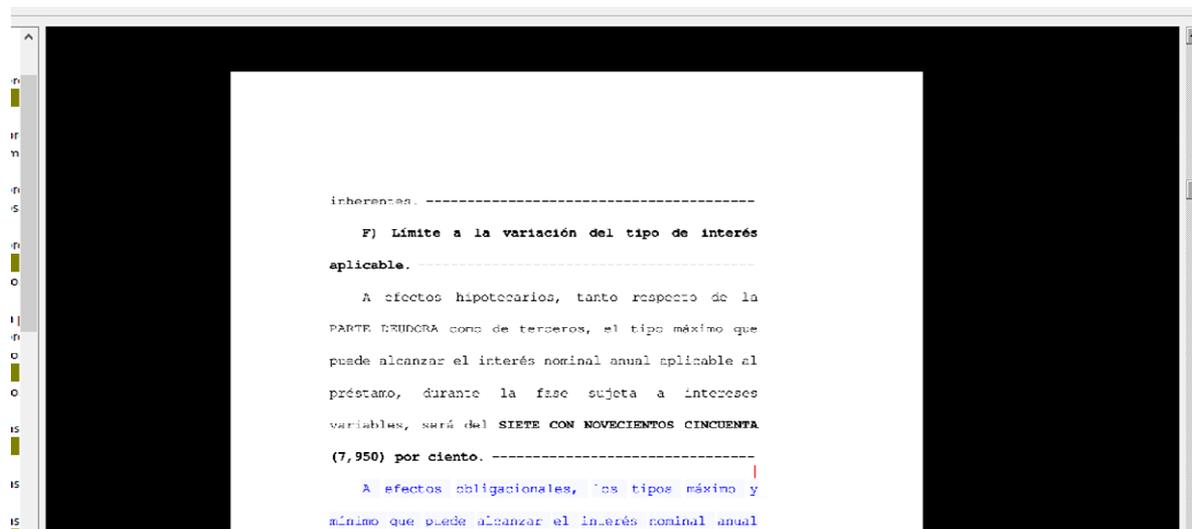
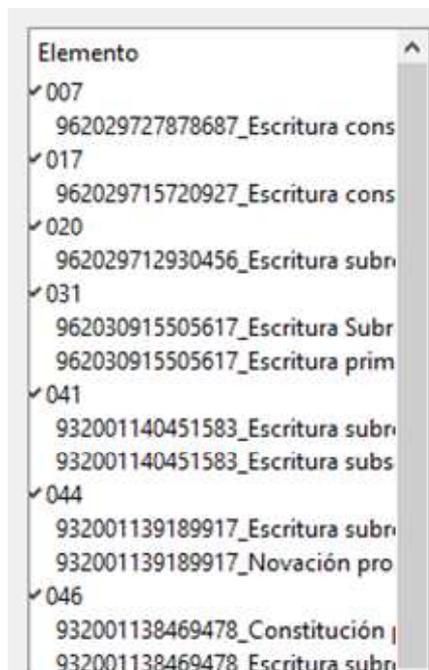
Tipo	Resultados
<input type="checkbox"/> Adv. NO Suelo	1
<input type="checkbox"/> Clausula Suelo	176
<input type="checkbox"/> Novación Suelo	159
<input type="checkbox"/> Tipo Novación	77
<input type="checkbox"/> Tipo Nuevo	117
<input type="checkbox"/> Tipo Subrogacion Novativa	1
<input type="checkbox"/> Tipo Subrogación	38

Doc: 25 Pag: 14 Line: 8 Word: 347,755 Pos: 816,953 Tipo de fichero Escritura Mostrar solo: Todos PDF

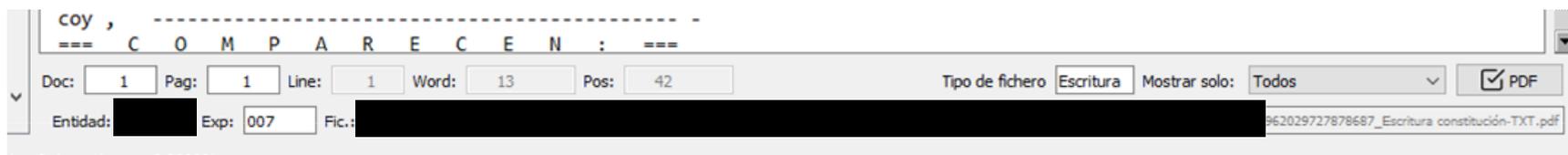
Entidad: Exp: 009 Fic:

- Navegación entre diferentes documentos y expedientes

- *Árbol de navegación, área de visualización PDF*



- *Cuadro inferior de información del documento*



- Marcaje de información

Elemento Valor

- 001 Esc. Novación 20031201-TXT.pdf
Esc. Préstamo 19970401-TXT.pdf
- 002 Esc. Novación 20070507-TXT.pdf
Esc. Préstamo 19970607-TXT.pdf
- 003 ASM 20150715-TXT.pdf
Esc. Préstamo 19980515-TXT.pdf
NNJ 20031015-TXT.pdf
NNJ 20050515-TXT.pdf
- 004 ASM 20150711-TXT.pdf
Esc. Novación 20041111-TXT.pdf
Esc. Préstamo 19980611-TXT.pdf
- 005 Esc. Novación 20141112-TXT.pdf
Esc. Préstamo 20060213-TXT.pdf
- 006 ASM 201507xx-TXT.pdf
Esc. Novación 20030904-TXT.pdf
Esc. Préstamo 19980923-TXT.pdf
NNJ 20051026-TXT.pdf
- 007 Esc. Matriz 19950629-TXT.pdf
Esc. Novación 20050411-TXT.pdf
Esc. Préstamo 19981210-TXT.pdf
- 008 Esc. Novación 20060113-TXT.pdf
Esc. Préstamo 19990216-TXT.pdf
- 009 ACS 20150728-TXT.pdf
Esc. Novación 20050713-TXT.pdf
Esc. Préstamo 19990819-TXT.pdf
- 010 ASM 201507xx-TXT.pdf
Esc. Préstamo 20010113-TXT.pdf
- 011 ACS 20170901-TXT.pdf
Esc. Novación 20051207-TXT.pdf
Esc. Préstamo 20010627-TXT.pdf
- 012 Esc. Novación 20060823-TXT.pdf
Esc. Préstamo 20011025-TXT.pdf
- 013 Esc. Préstamo 20011120-TXT.pdf
- 014 ACS 20150807-TXT.pdf
Esc. Novación 20050912-TXT.pdf
Esc. Préstamo 19960312-TXT.pdf
- 015 ACS 20150910-TXT.pdf

205200625

grafe, se pacta expresamente que el interés resultante no podrá superar el TRECE POR CIENTO nominal anual ni ser inferior al CINCO POR CIENTO nominal anual. -----
d) Comunicación de los tipos a la parte prestataria. -----
La parte prestataria conocerá los índices de referencia mediante la publicación de los mismos por el Banco de España, mensualmente, en el Boletín Oficial del Estado. -----
Transcurrido el plazo de un mes desde la publicación en el Boletín Oficial del Estado sin haber recibido la Caja oposición de la parte prestataria al nuevo tipo, se entenderá que se ha aceptado por ésta

Doc: 16 Pag: 14 Line: 2 Word: 210.291 Pos: /99.141 Tipo de fichero Escritura Mostrar solo: Todos PDF

Entidad: Exp: 006 Fic: (006)Esc. Préstamo 19980923-TXT.pdf

Tipo Texto	Original	Filtro	Ninguno
<input type="checkbox"/> BOW.DIGIT	1620760		
<input type="checkbox"/> SUB_FUZZY.VAL_INTERES	798743		
<input type="checkbox"/> SUB_VAL_INTERES	678599		
<input type="checkbox"/> FUZZY.BOW.NUMERO	825258		
<input type="checkbox"/> IMP.IJN	55427		
<input type="checkbox"/> BOW.NUMERO	149489		
<input type="checkbox"/> COMP	644		
<input type="checkbox"/> BOW.BAJADA	15246		
<input type="checkbox"/> RAS	25559		
<input type="checkbox"/> BOW.%	26293		
<input type="checkbox"/> ITE	7319		
<input type="checkbox"/> ZZY.LIMITE	7586		
<input type="checkbox"/> BOW.SUELO	15007		
<input type="checkbox"/> ZZY.TIPO_DE_INTERES	31071		
<input type="checkbox"/> D_DE_INTERES	25562		

Tipo	Resultados
<input type="checkbox"/> Adv. NO Suelo	1
<input type="checkbox"/> Clausula Suelo	176
<input type="checkbox"/> Novación Suelo	159
<input type="checkbox"/> Tipo Novación	77
<input type="checkbox"/> Tipo Nuevo	117
<input type="checkbox"/> Tipo Subrogacion Novativa	1
<input type="checkbox"/> Tipo Subrogación	38

Nuevo Marcaje

Extracto:
el interés resultante no podrá superar el TRECE POR CIENTO nominal anual ni ser inferior al CINCO POR CIENTO nominal anual.

Valor(es):
5%

Tipo de DATO
Clausula Suelo

Words: 0 Filtro

En expediente: Añadir resultados a: regexp: Marcajes:

Búsqueda en el documento actual Diferenciar Mayúsculas/Minúsculas Búsqueda Difusa: Núm. Dif. (máx. 3) 0

Búsqueda con Anclajes Visuales Cuadros de búsqueda

- Marcaje de información

grafe, se pacta expresamente que el interés resultante no podrá superar el TRECE POR CIENTO nominal anual ni ser inferior al CINCO POR CIENTO nominal anual. -----

d) Comunicación de los tipos a la parte prestataria. -----

La parte prestataria conocerá los índices de referencia mediante la publicación de los mismos por el Banco de España, mensualmente, en el Boletín Oficial del Estado. -----

Transcurrido el plazo de un mes desde la publicación en el Boletín Oficial del Estado sin haber recibido la Caja oposición de la parte prestataria al

Nuevo Marcaje

Extracto:

el interés resultante no podrá superar el TRECE POR CIENTO nominal anual ni ser inferior al CINCO POR CIENTO nominal anual. -----

Valor(es):

5%

Tipo de DATO

Clausula Suelo

Icons: Document with checkmark, NA, Document with pencil

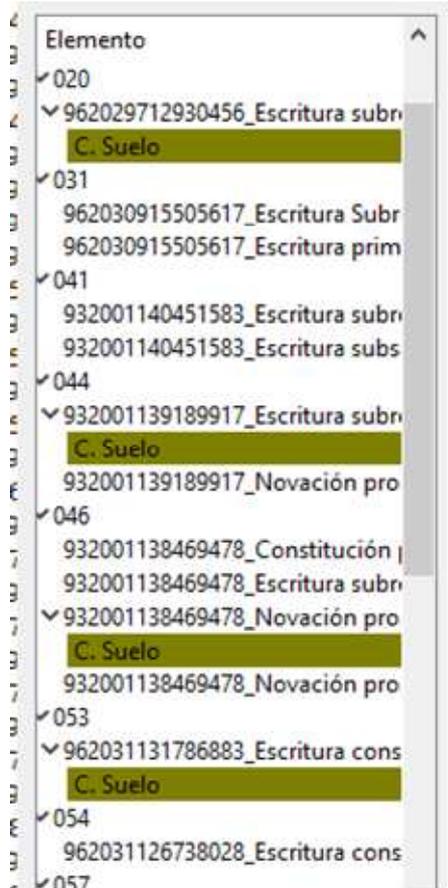
NAVEGACIÓN POR LOS MARCAJES

Árbol de navegación con marcajes integrados

The screenshot displays the TYREX 2.0 interface for document processing. On the left, a tree view shows a list of document elements, with 'Clausula Suelo' selected. The main window shows a document page with a highlighted text block. A red box labeled 'Marcaje realizado' points to a specific sentence in the text. On the right, a table titled 'Marcajes de información' lists various codes and their corresponding results.

Código	Res.	Tipo	Resultados
<input type="checkbox"/> BOW.DIGIT	1620760	<input type="checkbox"/> Adv. NO Suelo	1
<input type="checkbox"/> SUB_FUZZY.VAL_INTERES	798743	<input checked="" type="checkbox"/> Clausula Suelo	176
<input type="checkbox"/> SUB.VAL_INTERES	678599	<input type="checkbox"/> Novación Suelo	159
<input type="checkbox"/> FUZZY.BOW.NUMERO	825258	<input type="checkbox"/> Tipo Novación	77
<input type="checkbox"/> IMP_LI_N	55427	<input type="checkbox"/> Tipo Nuevo	117
<input type="checkbox"/> BOW.NUMERO	149489	<input type="checkbox"/> Tipo Subrogación Novativa	1
<input type="checkbox"/> COMP	644	<input type="checkbox"/> Tipo Subrogación	38
<input type="checkbox"/> FUZZY.BOW.BAJADA	15246		
<input type="checkbox"/> IMP_I.FTRAS	25559		
<input type="checkbox"/> FUZZY.BOW.%	26293		
<input type="checkbox"/> SUB.LIMITE	7319		
<input type="checkbox"/> SUB_FUZZY.LIMITE	7586		
<input type="checkbox"/> FUZZY.BOW.SUELO	15007		
<input type="checkbox"/> SUB_FUZZY.TIPO_DE_INTERES	31071		
<input type="checkbox"/> SUB.TIPO_DE_INTERES	25562		

- Navegación entre diferentes marcas de información.



<input type="checkbox"/>	ML_Suelo NO Obl.	2
<input type="checkbox"/>	ML_Adv. NO Suelo	2
<input type="checkbox"/>	ML_Adv. Suelo	6
<input checked="" type="checkbox"/>	C. Suelo	43
<input type="checkbox"/>	tipo sustitutivo	1

« « 10 de 43 » »

Words: 0 Filtro

Árbol de navegación con marcajes y código de color

Formulario de validación

Opciones de filtrado

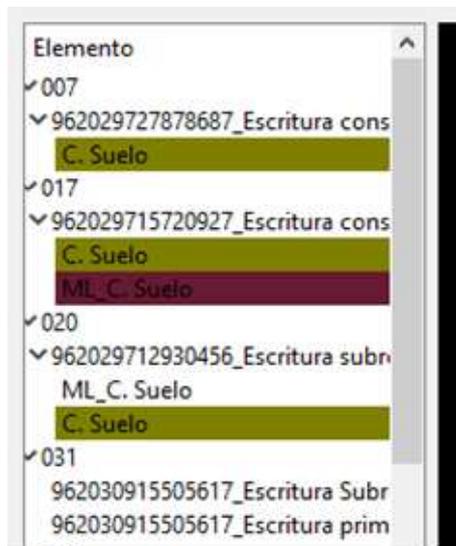
Marcaje propuesto

Marcajes de información (ML)

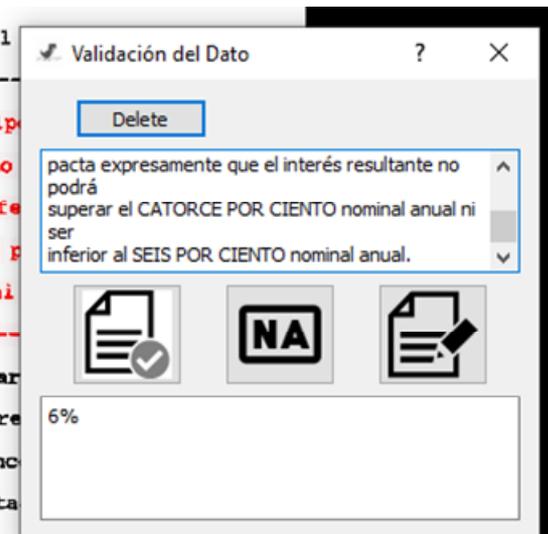
The screenshot displays the TyREX 2.0 interface for document processing. On the left, a tree view lists document elements with status indicators like 'Tanto', 'Si dej', and 'Sin pa'. The main window shows a document page with a 'Validación del Dato' dialog box overlaid, containing a 'Delete' button and a text area with a dropdown menu. On the right, a 'Filtro' panel shows a table of 'Regex' with columns for 'Código', 'Res.', and 'Tipo'. Below this, a 'Marcajes' section shows a table of 'Tipo' with columns for 'Tipo' and 'Resultados:'. The bottom status bar shows document statistics: 'Doc: 4', 'Pag: 16', 'Line: 14', 'Word: 50.215', 'Pos: 130.825'. The Windows taskbar at the bottom includes icons for various applications and the system clock showing '16:48' on '26/05/2021'.

- Validación de Marcajes

- Validar si una propuesta de marca de información es correcta.



el tipo de referencia por cualquier otra causa, el de interés a aplicar será el último vigente.-----
Tanto en el supuesto de que se aplique el tipo de referencia, MIBOR a un año, definido en el apartado los índices sustitutivos previstos en ese epígrafe pacta expresamente que el interés resultante no podrá superar el CATORCE POR CIENTO nominal anual ni inferior al SEIS POR CIENTO nominal anual.-----
d).- Comunicación de los tipos a la parte prestataria. La parte prestataria conocerá los índices de referencia mediante la publicación de los mismos por el Banco de España, mensualmente, en el Boletín Oficial de España.



- Código de colores:

- Fondo verde: Marcaje validado OK
- Fondo rojo: Marcaje validado NOK
- Sin color de fondo: Sin validar

2. TRATAMIENTO Y REVISIÓN DE EXPEDIENTES -TYREX

OPCIONES DE BÚSQUDA AVANZADA

The screenshot displays the TyREX 2.0 interface for document processing. The main window shows a document with text in Spanish, including phrases like "por razones distintas a las especificadas en los dos primeros párrafos de este apartado, en el futuro en el B.O.E., el tipo base a aplicar será el que corresponda al expresado para la revisión y resulte publicado oficialmente por el Banco de España o Ministerio de Economía y Hacienda, por este orden, y en el caso de que no exista tal publicación o resulte imposible determinar el tipo de referencia por cualquier otra causa, el tipo de interés a aplicar será el último vigente." and "Tanto en el supuesto de que se aplique el tipo de referencia, MIBOR a un año, definido en el apartado a) o los índices sustitutos previstos en ese epígrafe, se pacta expresamente que el interés resultante no podrá superar el CATORCE POR CIENTO nominal anual ni ser inferior al SEIS POR CIENTO nominal anual.".

On the right side, there is a search configuration panel with the following sections:

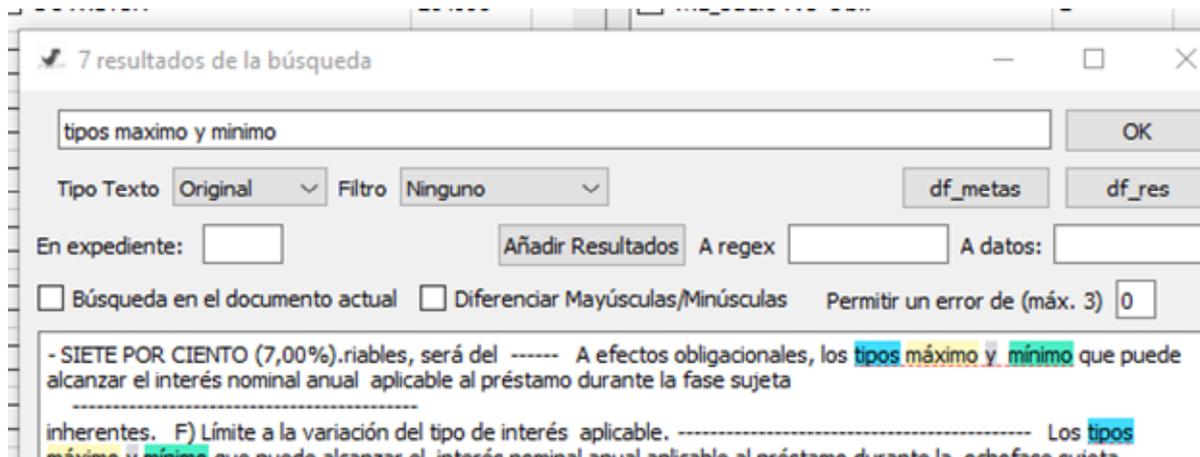
- Exportar metas to Excel** and **Export to Excel (by doc)** buttons.
- Regex** section with a list of codes and results:

Código	Res.
<input type="checkbox"/> BOW.DIGIT	1620760
<input type="checkbox"/> SUB_FUZZY.VAL_INTERES	798743
<input type="checkbox"/> SUB.VAL_INTERES	678599
<input type="checkbox"/> FUZZY.BOW.NUMERO	825258
<input type="checkbox"/> IMP.LIÑ	55427
<input type="checkbox"/> BOW.NUMERO	149489
<input type="checkbox"/> COMP	644
<input type="checkbox"/> FUZZY.BOW.BAJADA	15246
<input type="checkbox"/> IMP.LETRAS	25559
<input type="checkbox"/> FUZZY.BOW.%	26293
<input type="checkbox"/> SUB.LIMITE	7319
<input type="checkbox"/> SUB_FUZZY.LIMITE	7586
<input type="checkbox"/> FUZZY.BOW.SUELO	15007
<input type="checkbox"/> SUB_FUZZY.TIPO_DE_INTERES	31071
<input type="checkbox"/> SUB.TIPO_DE_INTERES	25562
- Marcajes** section with a list of types and results:

Tipo	Resultados
<input type="checkbox"/> Adv. NO Suelo	1
<input checked="" type="checkbox"/> Clausula Suelo	175
<input type="checkbox"/> Novación Suelo	159
<input type="checkbox"/> Tipo Novación	77
<input type="checkbox"/> Tipo Nuevo	117
<input type="checkbox"/> Tipo Subrogacion Novativa	1
<input type="checkbox"/> Tipo Subrogación	38

At the bottom right, a callout box with an orange border contains the text: "Opciones de búsqueda avanzada".

- Búsquedas:



- Opciones de búsqueda:

- Búsqueda en toda la muestra/documento actual
- Búsqueda en un expediente concreto
- Case Sensitive
- **Búsqueda Difusa:** Permitir un número de diferencias

- Opción de guardar los resultados como marcajes

- Tratamiento de documentos semi-estructurados:

- Definición de zonas de búsqueda

The screenshot displays two main components of the TYREX system interface:

- Left Panel (Document Processing):** Shows a document with various fields. A green box highlights the text "Ordain gutuna Carta de pago". Below it, a table lists interest types and rates:

berandutea	gke
demora	e.e.r.
9,00%	1,43%

 Another table shows interest rates:

diolts / divisa	aldia / periodo	Euros
kapitale / capital	02-11-17 / 02-12-17	243,30
interesak / intereses		130,35
kuota guztira / total cuota		373,65
ordainzeleku guztira / total a pagar		373,65
kontuan zordandutako zenbatekoa		373,65
importe adeudado en cuenta		373,65
- Right Panel (Search Results):** Shows a search window with 177 results. The search criteria include "Cadena de búsqueda" and "Búsqueda Difusa: Num Dif. (máx. 3) 0". The results list interest rates such as 1,889549%, 1,887584%, 1,857111%, 1,935733%, and 3,900000%. A dropdown menu for "Total cuota" is visible, showing options like "Interes Demora", "Interes Nominal", and "Principal".

- Exportación a Excel

- Diferentes opciones de exportación

	D	E	F	H	I
	nombre	tipo	entidad	TAEs	InteresNominal
o\INS	TIP-8522729556-TXT.pdf	Rev. Tipos		0,52	0,520000
o\INS	TIP - 8522339782-TXT.pdf	Rev. Tipos		0,48	0,482000
o\INS	TIP-8521950200-TXT.pdf	Rev. Tipos		0,34	0,346000
o\INS	TIP-8521465113-TXT.pdf	Rev. Tipos		0,28	0,282000

- Resultados

- Más de 7000 escrituras analizadas
- Número relativamente alto de falsos positivos, pero útil para agilizar el análisis
 - *Exactitud: 95% , Precisión 65%*
- Permite agilizar la revisión de grandes cantidades de documentación
- Muy buena aceptación por parte de los analistas

- Exactitud: predicciones correctas / número real de casos.
- Precisión: predicciones correctas / total predicciones.

- Alta potencialidad:

- Aumento de trazabilidad e incremento de productividad.
- No reemplaza al analista
- Base para la obtención de evidencia.
- Ha trascendido su finalidad inicial

Caso de éxito

LECCIONES APRENDIDAS

- **Retos:**

- Alta dependencia del proceso de OCR y la calidad de documentos de origen
- Elevado coste de entrenamiento (Algoritmos de machine learning es lo más fácil)

- **Soluciones:**

- La extracción de características (features) por “sliding windows”
- Expresiones regulares como parte de las features
- Retro-alimentación de las validaciones
- Adaptar “entrenamiento” y entregables.
 - *Tan importante como la calidad de los resultados, es la forma de presentarlos*

- **Proceso:**

- Reflexión previa.
- Desarrollo incremental.
- Proximidad con equipo supervisor.

GRACIAS POR SU ATENCIÓN

BRUNO COUTINHO

PABLO YOLDI



- Procesado y entrenamiento:

- “*Bag of words*” por “*Sliding Windows*”
- Clasificación de expresiones parecidas, pero conceptualmente diferentes:
 - *Cláusula suelo, cláusula suelo no obligacional, advertencia notario, novación suelo. Todos lo demás -> “Noise”.*
- Diccionarios de corrección errores, “*lematización*”, *stopwords*.
- Integración de expresiones regulares en el vector de *features*.
- Algoritmos de obtención de “*Noise*”.
- Comparación automática de diferentes modelos y parámetros de entrenamiento
 - *Red neuronal, Random forest, modelo de regresión logística.*