



Facultad de Ciencias Empresariales

EL PROCESAMIENTO DE TEXTOS JURÍDICOS: NATURAL LANGUAGE PROCESSING & LEGAL TECH

ESTADO DE LA CUESTIÓN, OPORTUNIDADES DE FUTURO Y REALIDAD PRÁCTICA

Autor: Alejandro Fuentes González

5º E3 Analytics

Tutor: José Portela González

RESUMEN

La aparición de técnicas relacionados con el procesamiento de lenguaje natural supone una verdadera revolución para numerosos sectores, entre los que podemos destacar la industria legal.

Para analizar este impacto es necesario conocer los fundamentos básicos del procesamiento de lenguaje natural. Esto permite hacerse ya una idea de las posibilidades que el empleo de estas técnicas ofrece para transformar la práctica jurídica.

El número de empresas (en su gran mayoría *startups*) especializadas en este tipo de algoritmos destinados a la automatización del sector legal es cada vez mayor. El amplio abanico de posibilidades que permite el uso de estos algoritmos hace que haya muchas empresas que ofrecen servicios diferenciados: desde automatizar la redacción de un contrato, a hacerlo con la búsqueda de la jurisprudencia, pasando por el diseño de los primeros *chatbots* legales.

La realidad práctica nos indica que se han realizado ya algunos estudios empíricos sobre la precisión de este tipo de técnicas y que los despachos de referencia están apostando por unas tecnologías que, si bien no pueden sustituir al abogado, sí pueden ayudarle a ofrecer un servicio de mayor valor.

Palabras clave: procesamiento de lenguaje natural, *Machine Learning*, *LegalTech*, automatización del sector jurídico, Inteligencia Artificial

ABSTRACT

The emergence of techniques related to natural language processing is a real revolution for many sectors, among which we can highlight the legal industry.

In order to analyse this impact, it is necessary to understand the basic fundamentals of natural language processing. This already gives an idea of the possibilities that the use of these techniques offers for transforming legal practice.

The number of companies (mostly startups) specialising in this type of algorithms for automating the legal sector is growing. The wide range of possibilities offered by the use of these algorithms means that there are many companies offering different services: from automating the drafting of a contract, to doing so with the search for case law, to designing the first legal chatbots.

Practical reality indicates that some empirical studies have already been carried out on the accuracy of these types of techniques and that leading law firms are betting on technologies that, although they cannot replace the lawyer, they can help him or her to offer a more valuable service.

Key words: natural language processing, Machine Learning, LegalTech, legal industry automation, Artificial Intelligence

ÍNDICE

CAPÍTULO I. INTRODUCCIÓN	6
I.1. MOTIVACIÓN	6
I.2. METODOLOGÍA	7
I.3. OBJETIVOS	7
CAPÍTULO II. FUNDAMENTOS DEL PROCESAMIENTO DE LENGUAJE NATURAL (NLP).....	9
II.1. ¿QUÉ ES Y PARA QUÉ SIRVE EL PROCESAMIENTO DE LENGUAJE NATURAL?	9
II.2. CONCEPTOS CLAVE EN NLP	10
CAPÍTULO III. NLP EN LEGAL TECH	15
III.1. ¿QUÉ ES EL LEGAL TECH?	15
III.2. APLICACIONES: ¿CÓMO PODEMOS EMPLEAR ESTAS TÉCNICAS PARA GENERAR VALOR?.....	18
III.3. ESTADO ACUAL: UNA VISIÓN INTERNACIONAL.....	28
CAPÍTULO IV. CASOS DE USO	34
IV.1. TAXY.IO, DERECHO TRIBUTARIO ALEMÁN	34
IV.2. ANÁLISIS PREDICTIVO: TRIBUNAL EUROPEO DE DERECHOS HUMANOS.....	39
IV.3. CUATRE CASAS: REFERENTE EN ESPAÑA	41
CAPÍTULO V. CONCLUSIONES	46
BIBLIOGRAFÍA	48
ANEXOS	51
ANEXO I. ENTREVISTA SOBRE <i>LEGALTECH</i> e INTELIGENCIA ARTIFICIAL EN LOS DESPACHOS DE ABOGADOS	51

ÍNDICE DE FIGURAS

Figura 1.....	13
Figura 2.....	13
Figura 3.....	14
Figura 4.....	23
Figura 5.....	28
Figura 6.....	35
Figura 7.....	36
Figura 8.....	37
Figura 9.....	38

CAPÍTULO I. INTRODUCCIÓN

I.1. MOTIVACIÓN

El procesamiento de lenguaje natural (más conocido por su terminología inglesa, *natural language processing* o **NLP**) es una técnica de *machine learning* en constante desarrollo en los últimos años. Sus aplicaciones son muy diversas: desde la minería de datos al análisis de sentimientos, pasando por las crecientes aplicaciones en el mundo jurídico. Será este el campo en el que centraremos nuestro análisis.

Aparentemente, el Derecho y el *Machine Learning* nos parecen ámbitos del conocimiento no demasiado interconectados. Al fin y al cabo, pensar en el nuevo mundo científico que abren las novedosas técnicas de inteligencia artificial, algoritmos, ... nos lleva a pensar automáticamente en datos numéricos. Sin embargo, el ordenamiento jurídico está compuesto eminentemente por textos sin apenas referencias de este tipo.

Esta aparente lejanía entre ambos campos no es tal. A pesar de que el desarrollo de las técnicas de *Machine Learning* aplicadas al Derecho no es tan avanzado como en otras ramas del conocimiento, no debemos obviar los avances que se han producido en los últimos años, se están produciendo en el presente y, sobre todo, aquellos que se producirán en el futuro.

Las posibilidades que el procesamiento de lenguaje natural abre son infinitas, también para los operadores del mundo jurídico. La enorme cantidad de horas que los abogados emplean en los procesos de *due diligence* es solo un ejemplo de los ámbitos en los que estas técnicas permitirían mejorar en la práctica jurídica.

Esta unión entre Derecho y Tecnología ha dado lugar a lo que se conoce como *Legal Tech*, una industria con un enorme potencial de crecimiento. El abogado del futuro deberá ser capaz de manejar las herramientas de procesamiento de lenguaje natural para mantener una posición competitiva en el mercado.

¿Acaso se imaginan a un abogado que en el año 2021 no sepa manejar un procesador de textos informático como Office Word? ¿Qué nos lleva a pensar que en el futuro no ocurrirá lo mismo con otros *softwares* que hagan mucho más fácil y eficiente la tarea de los abogados?

I.2. METODOLOGÍA

Para poder analizar estos cambios, de vertiginoso ritmo en algunos casos, que se vienen produciendo en los últimos años y, sin duda, seguirán ocurriendo en el futuro (cada vez con mayor repercusión) es necesario acudir al origen de los mismos; es decir, comenzaremos la investigación acudiendo al concepto de **procesamiento de lenguaje natural**. Esto nos permitirá asentar las bases de la tecnología que estudiamos.

Posteriormente, expuestos ya los orígenes y fundamentos del concepto de **NLP**, ahondaremos en la unión, y potencial simbiosis, que está surgiendo entre esta técnica de procesamiento de textos y el Derecho. Analizaremos qué es el *Legal Tech*, cuáles son las aplicaciones que se están empleando actualmente para generar valor en la práctica jurídica y cuáles son las oportunidades que las tecnologías ofrecen a los agentes jurídicos de cara al futuro.

Por último, consideramos necesario dar unas breves pinceladas de la realidad, de la práctica. Para ello, además de analizar varios casos reales, contamos con la colaboración de un despacho de abogados español de reconocido prestigio, Cuatrecasas, Gonçalves Pereira, S.L.P. (en adelante, **Cuatrecasas**), uno de los más avanzados en el empleo de estas técnicas. En el **Anexo I** se encontrará la entrevista realizada a dos de las máximas responsables de este Despacho en cuanto a la implantación de la tecnología en el sector legal.

I.3. OBJETIVOS

El objetivo último de este Trabajo es el análisis de la repercusión que el empleo de técnicas de procesamiento de lenguaje natural puede tener en el ejercicio del Derecho. Para poder abordar este análisis, para alcanzar nuestro objetivo último, consideramos necesario desgranar este objetivo final en metas de menor alcance.

Es por ello que, como decíamos en el apartado anterior, comenzamos el Trabajo exponiendo los fundamentos científicos y teóricos de las técnicas de NLP. Posteriormente, trataremos ya de ir abordando conceptos muy relacionados con nuestro objetivo final, como puede ser el *Legal Tech*.

Este estudio del estado de la cuestión desde un plano teórico será completado en todo momento por una visión eminentemente práctica de la misma. Queremos dejar claro que los usos de las técnicas de procesamiento de lenguaje natural aplicadas al Derecho son de

enorme relevancia. Para ello, recogeremos estudios de directorios jurídicos y asociaciones, así como un testimonio directo de un despacho español.

CAPÍTULO II. FUNDAMENTOS DEL PROCESAMIENTO DE LENGUAJE NATURAL (NLP)

II.1. ¿QUÉ ES Y PARA QUÉ SIRVE EL PROCESAMIENTO DE LENGUAJE NATURAL?

Los seres humanos nos comunicamos unos con otros a través del uso del lenguaje natural. El avance de las técnicas de almacenamiento de información ha provocado que nos encontremos con repositorios con más y más información disponible. Pese a esta cantidad creciente de información disponible, la ausencia de técnicas que facilitaran el tratamiento de la misma no permitía aprovechar su enorme potencial.

Todo esto cambió con la aparición del procesamiento de lenguaje natural. Esta técnica “involucra una transformación [del texto en lenguaje natural] a una representación formal, manipula esta representación y por último, si es necesario, lleva los resultados nuevamente a lenguaje natural” (Hernández & Gómez, 2013, pág. 1). En definitiva, el procesamiento de lenguaje natural convierte texto no estructurado en una representación formal que los ordenadores pueden comprender y analizar (Nay, 2018, pág. 1).

Uno de los principales problemas del análisis de textos de lenguaje natural es que los datos que se generan en estos datos son **datos no estructurados**. De acuerdo con López Yse (2019), los datos no estructurados no se ajustan a la estructura tradicional de filas y columnas de las bases de datos relacionales. Sin embargo, este mismo autor expone que las revoluciones producidas en disciplinas como el *Machine Learning* han permitido comenzar a manipular estos datos no estructurados. Además, han introducido un nuevo cambio de gran importancia: anteriormente, el análisis de textos se realizaba a partir de sus palabras claves (*keywords*) pero la aparición de nuevas técnicas de procesamiento de lenguaje natural permite ahondar en el significado real de todas las palabras (lo que López Yse llama *the cognitive way*).

Vemos, por tanto, que el procesamiento de lenguaje natural es posible gracias al avance del *Machine Learning*, que puede ser definido como el proceso de entrenar un modelo computacional para cumplir una tarea con datos (Nay, 2018, pág. 3). Dentro de las técnicas existentes de *Machine Learning*, distinguimos entre las de aprendizaje supervisado y las de aprendizaje no supervisado.

En el **aprendizaje supervisado** nos encontramos con variables predictivas y variables de resultado, cuya combinación permite entrenar al modelo y que este mismo modelo pueda

tomar las nuevas observaciones y, según sus variables predictivas, predecir su variable resultado. Dependiendo del tipo de resultado que nazca de este modelo podemos dividir los modelos de aprendizaje supervisado en modelos de regresión (cuando el resultado es un valor numérico) o modelos de clasificación (el resultado es una categoría).

De acuerdo con Nay (2018), el objetivo primario del aprendizaje supervisado es enseñar a un modelo, que podrá generalizar, a partir de la muestra de datos llamada *training data* (conjunto de entrenamiento), en el *testing data* (conjunto de testeo). Una vez hecho esto, el modelo puede ser utilizado en situaciones reales donde se conocen las variables predictivas, pero no las de resultado, para predecir estas últimas.

Por su parte, el **aprendizaje no supervisado** no utiliza variables de resultado, solo encontramos variables predictivas que caracterizan a las diferentes observaciones del conjunto de datos (Nay, 2018, pág. 4). Esto aumenta enormemente los conjuntos de datos que pueden tratarse: lo habitual es que nos encontremos con datos no estructurados, sin una variable de resultado. Esto es especialmente común, como decíamos anteriormente, cuando nos referimos al procesamiento de textos.

Es especialmente difícil, de acuerdo con Nay (2018), medir el rendimiento de los modelos de aprendizaje no supervisado, especialmente comparándolos con los supervisados.

II.2. CONCEPTOS CLAVE EN NLP

2.1. Método one-hot-encoding y bag of words

El método *one-hot-encoding* permite representar una colección de palabras (como una oración) como una lista de 0s y 1s de la misma longitud que todo el vocabulario que tenemos en cuenta. El 0 aparece cuando la palabra no aparece en el texto que se analiza (por ejemplo, la oración) y el 1 cuando sí aparece. Es por ello que una oración siempre aparece representada con una larga lista de muchos 0s y algunos 1s, lo que se conoce como *sparse representation*. (Nay, 2018, pág. 5). Debemos tener en cuenta el concepto de *term frequency representation*, que simplemente hace referencia al número de veces que una palabra aparece en el texto que se analiza.

El empleo del método *one-hot-encoding* nos lleva al concepto de *bag of words*: la representación de un documento como una lista de palabras y la frecuencia con la que se repiten en el mismo, sin tener en cuenta su orden ni la gramática (Nay, 2018, pág. 5) (López Yse, 2019).

Si bien utilizamos la *term frequency* (TF) para indicar el número de veces que una palabra ocurre, es importante entender el concepto de *term frequency-inverse document frequency* (TF-IDF). Este concepto nos permite restar relevancia a palabras que se repiten mucho en todos los textos que analizamos, lo que nos permite destacar, dentro de un documento concreto, aquellos términos raros o únicos (López Yse, 2019).

El cálculo de la **tf-idf** se hace de la siguiente forma (Nay, 2018, pág. 22):

1. TF (*term frequency*): número de veces que una palabra aparece en un documento
2. IDF (*inverse document frequency*): logaritmo del total de documentos dividido entre el número de documentos en los que aparece el término. Por ello, si aparece en más documentos el resultado tiende a 0 (logaritmo de un cociente tendente a 1), mientras que, si aparece en pocos documentos, se aleja del 0
3. TF-IDF (*term frequency-inverse document frequency*): multiplicación de la TF por la IDF. Los términos con mayor TF-IDF son aquellos con una alta TF (aparecen muchas veces) y alta IDF (aparecen en pocos documentos, son términos únicos o raros).

2.2. Tokenización y eliminación de Stop Words

El proceso de **tokenización** es muy sencillo: se separan todos los términos separados por espacios en blanco y, al mismo tiempo, se eliminan algunos caracteres que no tienen interés para quien analiza el texto (por ejemplo, los signos de puntuación). De acuerdo con López Yse, esto puede provocar algunos problemas, pues hay algunos términos que deben entenderse conjuntamente (sirva de ejemplo el nombre de dos ciudades estadounidenses como San Francisco o Nueva York).

El proceso de **eliminación de stop words** es, asimismo, sencillo a la par que muy útil. Consiste en la eliminación de palabras muy comunes en una lengua: artículos, pronombres, preposiciones o conjunciones (López Yse, 2019).

Es habitual encontrar listas predeterminadas de *stop words*, preestablecidas según el idioma con el que vayamos a trabajar. Pese a ello, podemos modificar esta lista predeterminada según el ámbito de conocimiento en el que centremos nuestro análisis.

2.3. Stemming y lematización

El proceso de *stemming* tiene, de nuevo, el objetivo de depurar el texto y hacer que su análisis posterior sea más sencillo. Consiste en la eliminación de los afijos (tanto prefijos como sufijos) que formen los diferentes *tokens* del documento (López Yse, 2019).

Es importante señalar algunos de los problemas que López Yse remarca sobre el proceso de *stemming*. Podemos diferenciar entre afijos flexivos o derivativos: mientras que los primeros apenas añaden significado, los segundos crean palabras nuevas, con matices que pueden ser de interés para el posterior análisis.

Lo habitual es que los programas utilizados para el procesamiento de lenguaje natural (como R o Python) cuenten con listas de afijos; sin embargo, al igual que con el proceso de eliminación de *stop words* podemos editar estas listas según nuestras necesidades.

A pesar de los diferentes procesos que habremos utilizado, la cantidad de palabras a analizar aún será enorme. Por ello, surgió la **lematización**. Este proceso consiste en agrupar todas las palabras que tengan la misma raíz/lexema en una sola. Esto nos permite, por ejemplo, que los verbos en pasado se conviertan al presente (López Yse, 2019).

Es evidente que este proceso requiere de un mayor esfuerzo computacional, es el proceso más complicado de los que hemos analizado hasta ahora.

2.4. Word embeddings

La técnica *Word embeddings* consistente en representar palabras, individualmente consideradas, como vectores de números reales en un espacio determinado (Brownlee, 2019).

Uno de los métodos más eficientes de esta técnica es la que se conoce como **Word2Vec**, un método estadístico que permite vectorizar las palabras de un texto (Brownlee, 2019). Se basa en la creación de una red neuronal para determinar el vector que le corresponde a cada palabra, de modo que palabras con significados relacionados tengan vectores similares.¹

Ilustraremos esta técnica con un ejemplo (Enzyme Advising Group, 2019):

¹ Vid. Similitud coseno

Se tiene el vector correspondiente a la palabra “Rey”. Este está asociado al vector de la palabra “Hombre”. Así, si se le resta el vector “Hombre” y se le suma el vector “Mujer”, quedaría entonces el vector “Reina”.

Cuanto mayor sea el conjunto de datos, mayor será la precisión del algoritmo a la hora de vectorizar las palabras y entender su verdadero significado. Así, destacan algoritmos como **ELMo**, que genera el vector de cada palabra según su contexto.

Un ejemplo de lo anterior (Enzyme Advising Group, 2019):

Así, el vector para “banco” se genera de forma distinta según si se refiere a la institución bancaria o a un banco para sentarse.

2.5. Similitud coseno

Un concepto muy interesante relacionado con esta vectorización de las palabras es el que nos permite calcular la similitud (más que una similitud ortográfica, buscamos que se trate de palabras relacionadas entre sí). Una vez representadas todas las palabras en vectores (siguiendo el modelo *bag of words*), podemos calcular el grado de similitud entre ellas.

Siguiendo el enfoque de Ghatak (2019), se utiliza la siguiente fórmula para calcular la similitud:

Figura 1: fórmula de similitud (Ghatak, 2019, pág. 232)

$$sim_{cosine}(u, v) = \frac{u^T v}{\|u\|_2 \|v\|_2} = cosine(\theta)$$

Se calcula, por tanto, el coseno del ángulo que forman los dos vectores (Ghatak, 2019, pág. 232).

Así, podemos apreciar ejemplos de muy fácil comprensión (**Figura 2**):

Figura 2: resultado de similitud (Ghatak, 2019, pág. 233)

```
cosine similarity(man, woman) 0.8323494
```

```
cosine similarity(france-paris, india-delhi) 0.6974226
```

Respecto de la **Figura 2**: parece razonable entender que la similitud entre “hombre” y “mujer” sea alta, al igual que ocurre entre “Francia-París” e “India-Delhi”. Recordamos

que no hacemos referencia a que sean palabras similares gramáticamente, sino que sus significados tengan relación.

El empleo de esta técnica nos permite, de hecho, buscar cuál es la palabra más cercana a otra si le damos un ejemplo de palabras similares, lo que se conoce como analogías de palabras (Ghatak, 2019, pág. 233).

Figura 3: resultado de analogías de palabras (Ghatak, 2019, pág. 233)

```
[1] "italy -> italian : spain -> spanish"
```

```
[1] "small -> smaller : large -> larger"
```

```
[1] "india -> delhi : china -> beijing"
```

En la **Figura 3** se aprecia muy bien cómo funciona esto. Se ofrece un primer ejemplo y el algoritmo encuentra, para una palabra dada, aquella que sigue la misma lógica que la primera relación.

Así, si la primera relación es: si “Italia”, entonces “italiano”; el resultado de ofrecer la palabra “España” nos lleva, siguiendo esa lógica, a “español”.

CAPÍTULO III. NLP EN LEGAL TECH

III.1. ¿QUÉ ES EL LEGAL TECH?

1.1. Definición

LegalTech es, simplemente, la abreviatura del término anglosajón *Legal Technology*. Es por ello que, comúnmente, se define el campo del *LegalTech* como “el uso de la tecnología para proporcionar servicios legales” (LegalTechies, 2020).

Esta definición es, sin duda, muy amplia y puede abarcar desde tareas que ya están sumamente asentadas, como el uso de bases de datos electrónicas o procesadores de texto, hasta otras en las que aún hay un enorme camino por recorrer, entre las que encontramos aquellas tecnologías de las que pretendemos hablar en este trabajo (fundamentalmente, las aplicaciones que el procesamiento de lenguaje natural puede tener para los operadores jurídicos).

De acuerdo con LegalTechies, la crisis financiera de 2008 provocó que los despachos de abogados tuvieran que avanzar rápidamente en su adaptación a estas nuevas tecnologías, que les permitirían ser mucho más eficientes. Esto provocó la aparición de una “nueva” *LegalTech*, que comienza a usarse para referirse a diferentes startups especializadas en estas tecnologías.

Desde 2016 podemos encontrar una definición del concepto que nos permite especificar qué usos de tecnología son englobados bajo el nombre de *LegalTech*. En este sentido, se dice que *LegalTech* es “la utilización de tecnología en servicios legales para crear:

- Software o servicios online que reducen o eliminan la necesidad de consultar con un abogado
- Software o servicios online que aceleran las tareas de un abogado/despacho, reduciendo la mano de obra y/o número de horas necesarias
- Software o servicios online que simplifican y modifican la forma de contactar entre abogados y clientes” (Morell Ramos, 2016)

Podemos observar, en definitiva, que el objetivo fundamental es “eficientar” la prestación de servicios legales. El propio Morell Ramos hace referencia en su artículo de 2016 al principal problema que el *LegalTech* provoca: los juristas aprecian un fenómeno creciente de tecnologías que pueden sustituir su papel, pues hay numerosas tareas que ya no tiene por qué realizar el abogado y la necesidad de contar con uno se reduce.

Considero que este fenómeno es todavía muy lejano y poco plausible. Más bien definiendo una simbiosis entre tecnología y juristas/abogados: el objetivo del empleo de tecnologías debe ser automatizar tareas que consumen gran cantidad de tiempo en los despachos de abogados y que no aportan tanto valor como otras.²

1.2. Conceptos relacionados

1.2.1. *LawTech*

En los últimos años ha surgido un nuevo concepto estrechamente relacionado con el que tratamos de definir: el *LawTech*.

Se trata de un término muy similar, que se define como el conjunto de tecnologías que pretenden apoyar, complementar o sustituir los métodos tradicionales de prestación de servicios jurídicos, o que mejoran el funcionamiento del sistema judicial (The Law Society, 2019).

A priori, podríamos decir que esta definición se parece a lo que hemos expuesto como *LegalTech*. Sin lugar a dudas, es así. The Law Society considera que el único término que debe emplearse es el de *LawTech* y por eso ha dado una definición similar del concepto. A pesar de esta aparente homonimia, existen autores que claramente sí muestran una diferenciación entre ambos conceptos.

Debemos hacer mención aquí a la opinión de dos grandes expertos: Richard Susskind y Richard Trommans (Susskind & Trommans, 2018).

Susskind aboga por el concepto de *LawTech*, citando distintas razones:

- Asocia el concepto antiguo (*LegalTech*) a tareas de *backoffice* (como puede ser la contabilidad) pero no con las nuevas disrupciones tecnológicas que van a cambiar el sector (como el uso de inteligencia artificial).
- Considera que un nuevo término puede servir para relanzar el cambio.
- Cree que los juristas entienden *LegalTech* más para tecnólogos que para operadores del Derecho, por ello un cambio es necesario.

Trommans, por su parte, cree que debe mantenerse el uso de *LegalTech*, alegando que:

- La mayoría de términos similares (*InsurTech*, *PropTech* o *FinTech*) hacen referencia al sector o industria en el que se emplea la tecnología (en términos

² Sobre todo esto se ahondará en distintos apartados del Trabajo-

ingleses, *insurance*, *property* o *finance*). Por ello, debe emplearse el término *legal*, pues generalmente se habla de *legal sector* o *legal industry*. Considero que en España la tendencia es la misma, por lo que es un argumento plenamente aplicable a nuestro caso.

- El término *Law* es más general y tiene otros usos: lo utilizamos para referirnos a la carrera universitaria o a la concepción social del Derecho, pero nunca para referirnos a una industria o sector.

A pesar de que se trata de discusiones terminológicas que, sin duda, se mantendrán en el futuro, lo que realmente importa es que la definición parece siempre orientada a un lugar común: el uso de tecnologías que permitan “eficientar” la práctica jurídica.

1.2.2. *Legal AI*

Utilicemos el concepto de *LegalTech* o el de *LawTech*, las definiciones que podemos encontrar de ambos conceptos son igualmente amplias. Es por ello que han aparecido otros términos, más específicos. Destacamos por encima de todos, debido a la motivación y el objetivo de este Trabajo, el de *Legal AI*. Podemos definirlo como el uso de tecnologías de inteligencia artificial, como el empleo de procesamiento de lenguaje natural o técnicas de *Machine Learning*, en relación con tareas legales (Trommans, 2017). Se trata, por tanto, del empleo de las técnicas que definimos en el Capítulo II de este Trabajo.

Trommans considera que la verdadera limitación actual de estas tecnologías reside más en los usuarios que en las propias tecnologías, indicando que los límites del uso de la Inteligencia Artificial están con frecuencia más relacionados con la imaginación de los usuarios que con las tecnologías en sí.

Adelantándonos ya a lo que se tratará en subsiguientes apartados, debemos exponer aquí las predicciones de Trommans al respecto de la *Legal AI*. En este sentido, indica en su Guía que constantemente están apareciendo nuevas soluciones y tecnologías relacionadas con el NLP o el *Machine Learning*, lo que provoca que hacer un resumen del estado de la cuestión sea complicado. Considera que lo realmente importante es tener clara la estructura de las tendencias actuales en el sector, a pesar de que sea un sector muy dinámico y en continuo cambio.

Sin duda, Trommans tenía razón: elaborar una taxonomía sobre estas tecnologías es no solo complicado, sino arriesgado. El estado de la cuestión depende, más que nunca, del

momento en el que se realice, pues esta cuestión está en continuo cambio. A ello dedicaremos el resto del Trabajo.

III.2. APLICACIONES: ¿CÓMO PODEMOS EMPLEAR ESTAS TÉCNICAS PARA GENERAR VALOR?

Volvemos a centrarnos ahora, en exclusiva, en el procesamiento de textos de lenguaje natural, un campo del conocimiento que junta inteligencia artificial y lenguaje, es decir, se refiere a la interacción entre unos algoritmos y el lenguaje humano (Barry, Daniel Lee, & Jacky, 2020). En este artículo de la Universidad de Sydney se indica que cualquier algoritmo que tenga la capacidad de comprender, manipular y expresar el lenguaje tendrá un enorme impacto en la prestación de servicios jurídicos.

Como hemos venido diciendo, las implicaciones del surgimiento del *LegalTech*, en general, o el NLP, en particular, son gigantescas: se está viviendo una auténtica revolución en la práctica jurídica; una revolución que será aún mayor en el futuro.

El NLP se engloba dentro de lo que se denomina como algoritmos *data-driven*: hacen inferencias y diseñan modelos dinámicos, identificando patrones complejos en los datos y creando sus propias reglas y lógica, a menudo demasiado complejas (Barry, Daniel Lee, & Jacky, 2020).

Para explicar la importancia del surgimiento del procesamiento de lenguaje natural, debemos hacer referencias a las posibilidades que estas técnicas nos ofrecen; es decir, qué es lo que realmente pueden hacer los prestadores de servicios jurídicos gracias al NLP.

2.1. Técnicas empleadas (clasificación de John Nay)

2.1.1. Resumen de contenido

A pesar de que hay numerosas técnicas de resumen de contenido, vamos a limitarnos al análisis de dos: i) resumen textual y ii) identificación de temas.

El resumen textual puede clasificarse en dos tipos de técnicas: resumen extractivo y resumen abstractivo. Mientras que, en la primera, el algoritmo se limita a detectar las partes más importantes de un texto, extraerlas y combinarlas en un nuevo “texto-resumen”; la segunda técnica es mucho más compleja, pues consiste en la elaboración de un texto totalmente nuevo, con palabras y oraciones que no tienen por qué estar en el texto resumido (Nay, 2018, pág. 9). De acuerdo con Nay, una de las herramientas de este estilo de mayor renombre es PageRank, de Google.

Respecto de la identificación de temas, esta es especialmente útil cuando nos encontramos ante un número elevado de documentos que tratan diferentes cuestiones y lo que buscamos es un resumen de los temas que se tratan en cada uno de ellos. Nay (2018) define a este tipo de algoritmos como *topic modeling*.

El algoritmo funciona de la siguiente forma: i) crea temas para un mismo *corpus* (conjunto de documentos), ii) elige una distribución de temas para cada documento, iii) para cada palabra en un documento elige un tema de la distribución anterior. (Nay, 2018, pág. 11).

2.1.2. Extracción de contenido

Se trata de una técnica similar al resumen de contenido. Sin embargo, mientras que el resumen trata de dar una visión holística de un documento o *corpus*, la extracción simplemente se centra en las partes del texto analizado que nos interesan (Nay, 2018, pág. 17).

Destacandistintos tipos de extracción:

- Extracción de atributos: trata de extraer de un texto aquellos atributos que previamente especifiquemos (por ejemplo, cantidades monetarias).
- Extracción relacional: extrae las relaciones existentes entre dos atributos. Siguiendo con el mismo ejemplo, no solo extrae la cantidad monetaria, sino el objeto cuyo valor monetario se indica en el texto.
- Reconocimiento de entidades: predice si un token concreto (recordemos que los tokens son los diferentes elementos que se analizan en las técnicas de NLP)³ es una entidad numérica (si es una fecha, una duración, una moneda, etc.) o una entidad categórica (un nombre, una localización o una organización).
- Correferencias: es habitual en los textos legales que haya numerosas referencias cruzadas dentro de ellos. Por ejemplo, una cláusula en un contrato que se refiere a otra cláusula del mismo o un artículo de la ley que se refiere a uno anterior o posterior. Existen algoritmos que permiten identificar y extraer estas correferencias.

³ Vid. Apartado II.2.2 del Trabajo

2.1.3. Recuperación de información y documentos

Los algoritmos de recuperación de información consisten en una consulta del usuario, aplicada a un conjunto de documentos, que extrae de aquellos la información que se ajusta a la consulta (Nay, 2018, pág. 21).

Este tipo de algoritmos engloban desde algunos muy simples, como los *booleanos*, que simplemente devuelven aquellos documentos que contienen las palabras buscadas hasta otros mucho más complejos que requieren el uso de técnicas de *Machine Learning*. Estos algoritmos más complejos cuentan con la ventaja de poder clasificar la información devuelta según sea más o menos interesante para el usuario.

Los algoritmos que necesitan el uso de *Machine Learning* pueden clasificarse en algoritmos de aprendizaje no supervisado y algoritmos de aprendizaje supervisado. Destacamos especialmente los segundos, entre los que encontramos una herramienta que cada vez se utiliza más en los despachos de abogados, el *e-discovery*.

Este algoritmo consiste en buscar documentos relevantes para un caso, requiriendo un previo entrenamiento del modelo (por eso es de aprendizaje supervisado). El usuario enseña al algoritmo documentos relevantes para un caso; posteriormente, una vez analizado ese conjunto de entrenamiento, el modelo puede analizar documentos ajenos al mismo e ir detectando cuáles son interesantes para la consulta del usuario (Nay, 2018, pág. 24).

Uno de los ejemplos de este tipo de algoritmos es el llamado *Continuous Active Learning*, que, de acuerdo con Nay (2018) se organiza en cuatro pasos:

1. Encontrar al menos un ejemplo de documento relevante.
2. Entrenar un modelo de aprendizaje supervisado para que prediga la relevancia de un documento para un caso, puntuando los restantes documentos según sean de mayor o menor relevancia.
3. Revisar las puntuaciones obtenidas para clasificarlas manualmente como relevantes o no.
4. Repetir los pasos 2 y 3 de manera continuada, para mejorar la capacidad de predicción del modelo supervisado.

2.1.4. Predicción de resultados

Esta técnica consiste en predecir el resultado de un evento de interés, según el texto relacionado con dicho proceso. Por ejemplo, Nay realizó una predicción sobre las leyes norteamericanas para saber si, según su redacción, serían aprobadas o rechazadas por el Congreso (Nay, 2018, pág. 25).

Uno de los resultados de esta investigación fue estudiar cuáles eran los términos utilizados en las leyes sobre medioambiente que eran aprobadas y cuáles en las que eran rechazadas. Así, mientras que términos como “impacto” o “efectos” eran comunes en las primeras, otros como “calentamiento” o “temperatura” eran comunes en las leyes rechazadas. Esto permite predecir el resultado de futuras leyes que se propongan, siempre basándose en los resultados de votaciones anteriores.

2.1.5. Respuesta a preguntas

Se trata de una técnica totalmente revolucionaria, que permite dar a los clientes una primera respuesta sobre sus consultas jurídicas, simplemente a través de un algoritmo. Este tipo de técnicas han permitido el diseño de los primeros *chatbots* legales, que contestan consultas legales gracias a algoritmo previamente entrenados con ejemplos de preguntas y respuestas (Nay, 2018, pág. 28).

Sobre estos *chatbots*, de acuerdo con Nay, debemos diferenciar entre las respuestas más sencillas, de modo que el algoritmo simplemente elige entre el conjunto de preguntas y respuestas para los que ha sido entrenado, o respuestas más complejas, gracias a algoritmos que son capaces de redactar un texto nuevo, recordando así a lo que expusimos sobre el resumen abstractivo.⁴

2.2. ¿Cómo aportan valor a la práctica jurídica estas técnicas?

Hemos analizado ya algunas de las técnicas de procesamiento de lenguaje natural, así como sus fundamentos. Sin embargo, solo hemos pasado de perfil por las soluciones que estas tecnologías ofrecen al sector legal.

David Li (2020) señala que hay tres grandes tipos de análisis de datos que permiten crear valor a los operadores del sector legal: análisis estructurado, análisis conceptual y análisis predictivo.

⁴ Vid Apartado III.2.2.1

2.2.1. Análisis estructurado

Esta solución permite analizar similitudes y diferencias entre textos que conforman un mismo conjunto (Li, 2020, pág. 11). Esto permite establecer, por ejemplo, un porcentaje mínimo de similitud entre documentos, de modo que los buscadores solo extraigan aquellos documentos que superen ese mínimo, para no tener que leer documentos con similar información.

Otra de las utilidades de estas técnicas es el *email threading*, especialmente útil dado que el ejercicio de la abogacía suele implicar mandar muchos correos a clientes, a otros miembros del despacho, etc. Esta técnica permite identificar los correos relevantes en los que se encuentra información de valor y descartar el resto (Li, 2020, pág. 12)

2.2.2. Análisis conceptual

Una vez identificado el conjunto de documentos únicos y distintos, con información que nos interesa analizar, el siguiente paso es el análisis conceptual.

Una de las técnicas típicas es el *clustering*, que divide un conjunto de documentos en grupos que presentan ciertas características únicas. Es especialmente interesante por su carácter de algoritmo de aprendizaje no supervisado; es decir, el usuario no debe entrenar al algoritmo (Li, 2020, pág. 12).

Li identifica como utilidad de esta técnica el análisis exploratorio: si cuento con un conjunto de 100.000 documentos, es difícil analizar todos. Sin embargo, el *clustering* puede agruparlos en diferentes subgrupos y, una vez hecha la agrupación, centrarme en los que realmente me interesan. Es además habitual que la representación de los *clusters* sea muy visual.

Otra de las técnicas habituales es el *sentiment analysis*, que permite identificar patrones de emociones en un documento.

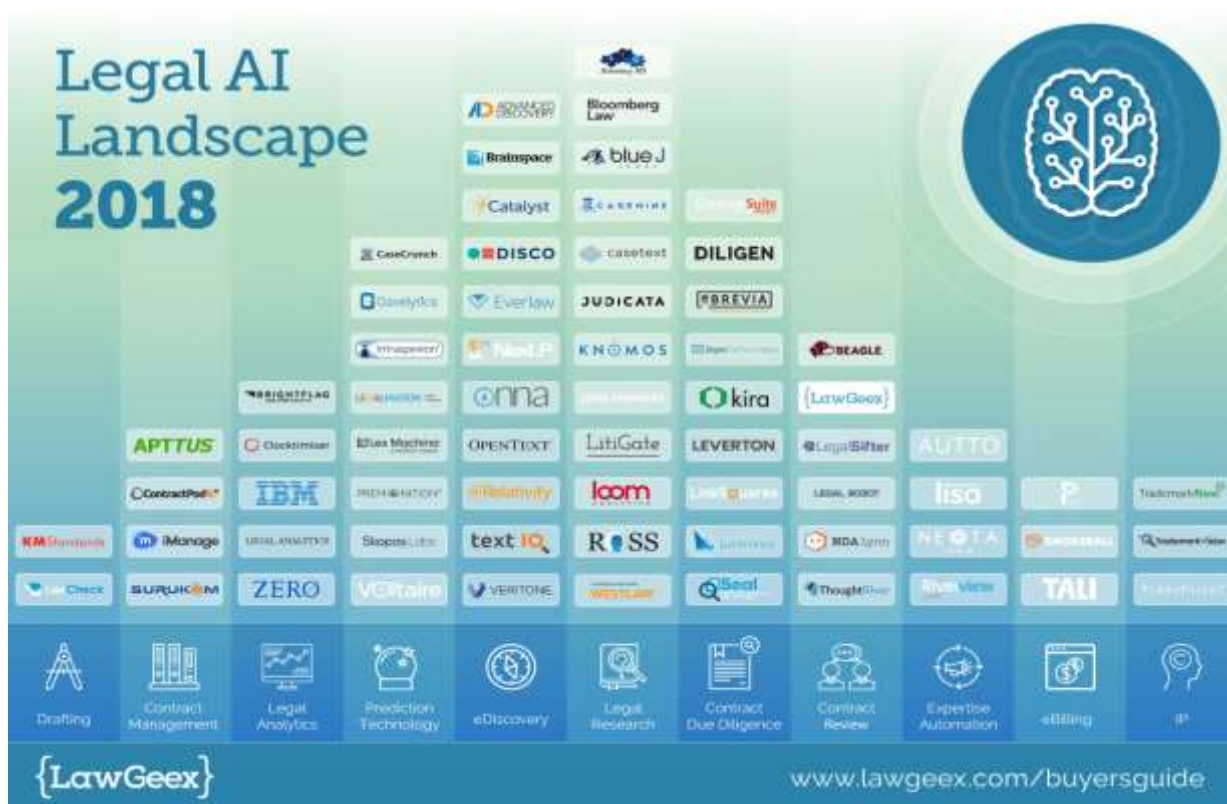
2.2.3. Análisis predictivo

Se trata esta vez de un algoritmo de aprendizaje supervisado que consiste en identificar aquellos documentos que son de interés para el usuario y aquellos que no lo son (Li, 2020, pág. 14). Una vez se haya hecho la clasificación de un número suficientemente grande de documentos, el algoritmo es capaz de identificar cuáles de los demás documentos de una base de datos son de interés para el usuario o no. Esto permite mejorar enormemente la búsqueda de información en bases de datos legales.

2.3. Aplicaciones: empresas especializadas

En la **Figura 4** podemos apreciar algunas de las aplicaciones que se están desarrollando, así como ejemplos de empresas especializadas en dichas aplicaciones.

Figura 4: Resumen de aplicaciones y empresas del sector *LegalTech* (LawGeex, 2018)



Dado que analizarlas todas sobrepasaría los límites de un Trabajo como este, vamos a seleccionar cinco soluciones, que explicaremos partiendo de un ejemplo real de empresa dedicada al desarrollo de esa tecnología:

- *Drafting*: Contract Standards
- Tecnología de predicción: Premonition
- Búsqueda legal: ROSS
- *Due Diligence* contractual: KIRA
- *E-Discovery*: Veritone

2.3.1. *Drafting*: Contract Standards

Contract Standards es una empresa que ofrece servicios muy interesantes para la práctica jurídica. Ofrece una librería de contratos y cláusulas estandarizados, que permite agilizar el proceso de redacción, negociación y conclusión de los contratos. Esto es especialmente

interesante en jurisdicciones anglosajonas, en las que la práctica contractual está mucho más estandarizada.

No solo ofrece este repositorio de contratos y cláusulas, sino que a través de su herramienta llamada *Contract Quick Builder*, permite la redacción casi automática de un contrato. Basta con elegir los elementos clave de la transacción/relación contractual que pretende crearse para que el algoritmo diseñe un contrato totalmente adaptado a las necesidades de los usuarios. Esta solución puede ser utilizada incluso por “usuarios de a pie”; es decir, personas que redactan y cierran contratos más sencillos, como pudiera ser un contrato de arrendamiento entre particulares.

Una solución más compleja y orientada a los despachos de abogados o grandes corporaciones comporta el análisis, organización y estandarización de todo el repositorio de contratos que estas empresas tienen.

En un primer paso, el análisis, el algoritmo estudia e identifica todos los componentes de cada cláusula contractual. Por ejemplo, determina si se trata de una cláusula estándar, negociada o específica para ese contrato, según sea más o menos frecuente en el conjunto de muestra con el que se entrenó al algoritmo. Se trata, sin duda, de un ejemplo de aprendizaje supervisado de clasificación de los que estudiamos en el Capítulo II. Una vez han sido analizadas, las cláusulas se organizan según la estructura interna del repositorio de cláusulas con el que *Contract Standards* cuenta. Realizados estos dos primeros pasos, el algoritmo, por sí mismo, simplifica el lenguaje y los términos empleados en los contratos, facilitando su lectura y comprensión.

2.3.2. Tecnología de predicción: *Premonition*

Haciendo uso de herramientas de predicción de resultado que ya hemos analizado,⁵ la empresa norteamericana *Premonition* se define a sí misma como una “*very, very unfair advantage in litigation*”. El servicio que presta se basa en analizar la mayor cantidad de jurisprudencia. Cuenta con una enorme base de datos, pero no solo pone esta base de datos al servicio de sus clientes, sus servicios van más allá.

Basándose en el análisis de la gran cantidad de datos disponibles, trata de predecir los resultados que determinadas elecciones podrían tener:

⁵ *Vid.* Capítulo III.2.1.4

- Analiza qué acciones legales son las más exitosas o si pedir la recusación de un determinado juez aumenta la probabilidad de éxito.
- Estudia los precedentes que un determinado juez o tribunal dictó anteriormente, lo que permite informar al cliente sobre sus posibilidades en un caso.
- Recoge información sobre los abogados litigadores, lo que permite también conocer de antemano cuáles serán los argumentos de la contraparte.

Ahondando de nuevo en su carácter de “ventaja injusta”, la propia compañía considera que sus servicios son una analogía de lo que se puede ver en la conocida película Moneyball, pionera en la demostración de la utilidad de los datos en casi cualquier campo que se nos ocurra (Premonition, 2017).

Considera que el aprovechamiento de sus servicios, el uso de la estadística avanzada basada en datos, ofrece esta ventaja por dos motivos: i) la falta de transparencia y la resistencia al cambio en el sector legal y ii) el desigual acceso a información en el sector. La tecnología de Premonition trata de dejar a un lado las decisiones basadas en las emociones para que estas pasen a estar basadas en los datos: seleccionar ante qué Tribunal, qué abogado te representará y qué argumentos tienen mayores probabilidades de éxito según el estudio de los precedentes.

2.3.3. *Búsqueda legal: ROSS Intelligence*

ROSS Intelligence es una de las pioneras del sector y una de las empresas más reconocidas en el *LegalTech*. Basada en el uso de inteligencia artificial, utiliza su enorme base de datos para mejorar la eficiencia en la práctica jurídica.

De nuevo, al igual que Premonition, no es un simple repositorio de jurisprudencia. Sin duda estos repositorios supusieron un antes y un después respecto de los grandes libros físicos que entorpecían la búsqueda de casos relevantes, pero la aparición de la Inteligencia Artificial, el NLP o el *Machine Learning*, hacen que una simple base de datos sea ya una tecnología obsoleta.

Por un lado, permite hacer preguntas legales en lenguaje coloquial y, analizando su base de datos interna, el algoritmo ofrece una respuesta basada en la jurisprudencia. Sin embargo, su servicio más interesante es el análisis de documentos.

Se trata este de un servicio muy potente, que permite adjuntar cualquier archivo de texto (por ejemplo, un borrador de una demanda). Posteriormente, permite elegir cualquiera de las partes de nuestro texto (por ejemplo, un argumento concreto del que nos gustaría

conocer la jurisprudencia relevante) y el algoritmo buscará argumentos similares en casos ya resueltos. Esto nos permite conocer, de antemano, cuáles son las posibilidades de nuestras peticiones.

También la lectura de la base de datos interna se ha mejorado: no solo podemos leer una enorme cantidad de jurisprudencia, sino que podemos, al igual que en nuestros propios documentos, seleccionar las partes sobre la que nos interesa ahondar en mayor detalle. El algoritmo, en ese caso, buscará jurisprudencia relacionada con la parte que seleccionamos.

2.3.4. Due Diligence contractual: KIRA

KIRA es, sin duda, una de las referentes del sector *LegalTech*. Haciendo uso de tecnologías de Inteligencia Artificial y *Machine Learning*, ofrece unos servicios difícilmente igualables a sus clientes. Ofrece, además, a través de la herramienta *Quick Study*, la personalización de los algoritmos para el caso concreto que se pretende desarrollar. Esta es la base de un algoritmo de aprendizaje supervisado: aprender con base en unos ejemplos e ir afinando el algoritmo con más y más datos de muestra.

KIRA ofrece servicios relacionados con el análisis de contratos y los procesos de *due diligence* legales. Al fin y al cabo, un proceso de *due diligence* consiste en la revisión de todos los contratos de una empresa, por lo que los servicios más utilizados siempre buscan ese objetivo: simplificar el análisis de los contratos.

El algoritmo analiza el contrato y lo estructura en diferentes partes. Esto permite, entre otras cuestiones, buscar si existen o no determinadas cláusulas en un contrato, acudir rápidamente a la sección que nos interese (por ejemplo, comparar las cláusulas de sumisión a una jurisdicción de distintos contratos). Además, permite organizar los contratos en diferentes grupos y obtener métricas relacionadas con los mismos. Aparece aquí la herramienta llamada KIRA Dashboard, que permite representar de manera muy visual los elementos clave de uno o varios contratos.

Los procesos de *due diligence* consumen una enorme cantidad de tiempo en todos los procesos de M&A, por lo que la utilidad de KIRA es enorme. El algoritmo permite identificar cláusulas relevantes (un ejemplo típico son las cláusulas de control), así como comparar contrato por contrato con un contrato marco, que sirva de ejemplo. Muestra así, como si fuera un mapa de calor, las desviaciones de cada contrato individual respecto del que sirve de ejemplo. Incluso puede extraer el contenido fundamental de diferentes

contratos y organizarlo en tablas de contenido, lo que permite una muy fácil comparación entre los distintos contratos.

2.3.5. *E-Discovery: Veritone*

El *E-Discovery (evidence Discovery)* es una de las soluciones que más se está desarrollando en los últimos años. Consiste en la extracción y análisis de evidencias de textos y otros documentos legales.

La compañía Veritone trabaja en múltiples sectores, pero su tecnología aiWARE es especialmente potente para el sector legal. Esta tecnología se basa en las técnicas que estamos estudiando (procesamiento de lenguaje natural o NLP) para analizar y poder extraer información muy relevante de textos jurídicos. El desarrollo de esta tecnología es tal que incluso permite analizar audios y vídeos como un documento legal más, por lo que los beneficios para las empresas que lo utilizan son enormes.

Destaca especialmente la tecnología Veritone Illuminate, cuyo objetivo es descubrir y analizar grandes volúmenes de información almacenada electrónicamente (ESI, *Electronically Stored Information*) que no necesariamente debe estar ya en formato de texto, pues, como decíamos, es capaz de analizar vídeo y audio. Estos últimos tipos de información son transcritos, para poder ser posteriormente analizadas como textos con las técnicas de NLP.

El objetivo fundamental de esta tecnología es ayudar a los abogados a realizar su trabajo de forma más eficiente: por una parte, porque permite analizar gran cantidad de datos de forma no-manual (esto reduce la cantidad de horas que los abogados dedican a estas tareas) y, por otra, porque permite que los abogados se centren únicamente en los textos donde está la información relevante (generando así un servicio de mayor valor para los clientes).

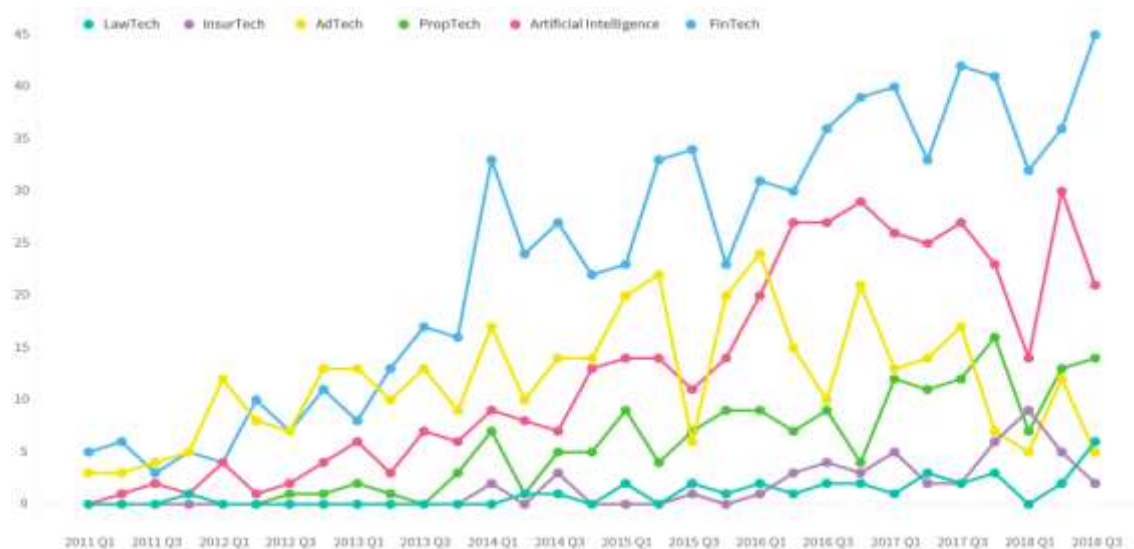
La enorme variedad de documentos que deben revisarse no es problema para esta tecnología. Es capaz de analizar mensajes de texto, emails, contratos, llamadas telefónicas o interrogatorios grabados en vídeo. Permite transcribir, redactar y organizar toda esta información en un mismo formato, que sea utilizable para las técnicas de procesamiento de lenguaje natural. Esto permite centrarse, una vez homogeneizados los formatos, en la extracción del contenido realmente importante.

III.3. ESTADO ACUAL: UNA VISIÓN INTERNACIONAL

3.1. Principales *hubs* del sector

La inversión en el sector *LegalTech* ascendía en agosto de 2019 a 926 millones de libras esterlinas. Se trata, además, de un sector en auge pero menos maduro que otros relacionados con la tecnología, como puede ser la industria *FinTech* (The Law Society, 2019, pág. 3). La **Figura 5** puede servir de ilustración para observar la comparativa de inversión entre diferentes sectores tecnológicos, observando como el *LawTech* (color turquesa) está muy por debajo de otros, pero puede apreciarse una tendencia creciente en los últimos años.

Figura 5: Comparativa inversión (The Law Society, 2019, pág. 5)



Destacamos especialmente cuatro ciudades como líderes del mercado *LegalTech*: Londres, Singapur, Hong Kong y Amsterdam.

La capital británica es uno de los grandes *hubs* tecnológicos a nivel mundial, lo que ha hecho que sea también una referencia en lo referente a las tecnologías aplicadas al sector legal.

Londres destaca especialmente por el ecosistema de tecnologías de Inteligencia Artificial. En esta ciudad podemos encontrar hasta 758 empresas dedicadas a dichas soluciones (The Law Society, 2019, pág. 11). El informe de The Law Society (2019) destaca, entre otras cuestiones, que la inversión en esta tecnología aumentó un 50% desde 2016 a 2017, llegando a los 200 millones de libras.

En el caso de la capital asiática, Singapur, destaca la creación en 1990 de la base de datos de la *Singapore Academy of Law*, llamada *LawNet*. Entre otros avances, la Universidad publicó en 2017 un plan de acción a cinco años que tenía como objetivo fomentar la colaboración público-privada para colocar a Singapur como uno de los principales centros mundiales en el sector (The Law Society, 2019, pág. 15).

En Hong Kong destaca la creación del *The Law & Technology Centre*, también llamado *LawTech Centre*, en 2011. Se trata de una colaboración entre la Facultad de Derecho y el Departamento de *Data Science* de la Universidad de Hong Kong, que trata de promover el crecimiento del *LegalTech* en la ciudad china (The Law Society, 2019, pág. 16). El Informe de The Law Society destaca, además, la creación de una plataforma electrónica llamada eBRAM (*Belt and Road Arbitration and Mediation Centre*). Aprovechando el prestigio de Hong Kong como sede internacional de arbitrajes, se tratan de aprovechar las diferentes soluciones que el *LegalTech* ofrece para mejorar los servicios prestados.

Con respecto al caso de Amsterdam, destaca por encima de todo la *Dutch Legal Tech*, una plataforma de innovación en el sector legal en la que participan emprendedores, abogados, inversores, académicos e instituciones públicas (The Law Society, 2019, pág. 18). Podemos destacar, como herramienta innovadora, la plataforma *Justice42*, que permite a los usuarios trabajar conjuntamente para resolver sus problemas legales, con la ayuda de expertos si fuera necesaria.

El Informe de The Law Society destaca otras ciudades con potencial dentro del sector. Entre ellas nos encontramos a Madrid, la capital española. El Informe pone el foco en este caso en el programa de estudios sobre *LegalTech* del Instituto de Empresa y en la acción de despachos de abogados, como Garrigues (The Law Society, 2019, pág. 23).

3.2. La situación en España

Un reciente Informe del Consejo General de la Abogacía Española, en colaboración con la asociación *Women in a Legal World*, ha realizado un análisis sobre la implantación de la tecnología en el ejercicio de la abogacía. De este informe podemos extraer una serie de conclusiones muy interesantes.

En primer lugar, es importante señalar que el 75% de los encuestados “considera que es importante implantar tecnología como palanca para ganar eficiencia y prestar un mejor servicio”, mientras que solo un 6% expuso que la tecnología no era relevante para el ejercicio de la abogacía (Consejero General de la Abogacía Española, 2021, pág. 11). Más

aún, el 80% (un porcentaje que aumenta hasta casi el 90% para los abogados entre 25 y 34 años) considera que la aplicación de las tecnologías abrirá nuevas oportunidades para el sector legal (Consejero General de la Abogacía Española, 2021, pág. 13).

Podemos hablar de tecnología con un carácter muy general, y en el informe se tratan cuestiones como el teletrabajo o el uso del correo electrónico, que tienen poco que ver con las técnicas de NLP o *Machine Learning* sobre las que versa este Trabajo. Sin embargo, podemos encontrar mucha información que sí es muy relevante. Así, el 50% respondió que “Sería útil para el desempeño de mis tareas disponer de herramientas automáticas de análisis que anticiparan impactos y plantearan alternativas” (Consejero General de la Abogacía Española, 2021, pág. 11).

Las tecnologías que aquí estudiamos están relacionadas con el análisis de información disponible gracias a nuevas soluciones como el *Machine Learning*. En este sentido, el 50% de los encuestados respondió que dedicaba entre el 21 y 40% de su trabajo a procesos relacionados con el “análisis y compilación de información” (Consejero General de la Abogacía Española, 2021, pág. 2). Además, la mitad de encuestados consideró que el 70% de estos trabajos tienen precedentes (Consejero General de la Abogacía Española, 2021, pág. 3). Esto nos permite entender por qué la aplicación de tecnologías en el sector legal puede ayudar a hacer mucho más eficiente la práctica jurídica.

También podríamos destacar los medios de búsqueda de información que los abogados suelen emplear: mientras que el 69% declara utilizar herramientas de búsqueda basadas en palabras clave, solo el 19% emplea herramientas automáticas de búsqueda (Consejero General de la Abogacía Española, 2021, pág. 5). Esto nos recuerda a los sistemas *booleanos* de los que ya hablamos anteriormente.⁶

Por último, es necesario hacer una mención a las barreras que los profesionales declaran percibir para la implantación de las tecnologías en la práctica jurídica. Un 50% de los encuestados declaró no tener ninguna tecnología avanzada para el desarrollo de su actividad, aunque el 31% declaró contar con tecnologías de Big Data y análisis predictivo. Mientras que solo un 6% cuenta con herramientas de Inteligencia Artificial o *Machine Learning*. (Consejero General de la Abogacía Española, 2021, pág. 8).

Debemos destacar, también, la ausencia de conocimientos suficientes, pues solo el 25% considera tener competencias y conocimientos digitales suficientes, así como la falta de

⁶ Vid. Capítulo III.2.1.3

planificación por parte de las empresas y despachos, en tanto que menos de una quinta parte considera que exista en su trabajo un plan de acción para la implementación de la tecnología (Consejero General de la Abogacía Española, 2021, pág. 11).

Existe, además, un consenso abrumador a la hora de señalar cuáles son los tres principales retos para la implantación de tecnologías en el sector legal: casi un 90% de los encuestados considera que el principal reto es el cambio cultural y dos terceras partes creen que los principales retos son: por un lado, el conocimiento de estas tecnologías y, por otro, las inversiones requeridas.

El Informe del Consejo General de la Abogacía Española identifica tres líneas de actuación para mejorar la implantación de la tecnología en la práctica jurídica: i) formación, ii) mayor acercamiento entre el sector legal y el tecnológico y iii) cambio cultural.

i) Formación

Uno de los grandes problemas existentes es la falta de formación. Los abogados no pueden aprovechar el enorme potencial que diferentes softwares como los analizados anteriormente.⁷ Para ello, sería recomendable: i) que los departamentos tecnológicos de las empresas y despachos formaran a los abogados y ii) que los abogados fueran conscientes de los verdaderos casos de uso de estas tecnologías, pues conocer realmente cómo su trabajo será más eficiente gracias al empleo de estas tecnologías les motivará para formarse en ellas (Consejero General de la Abogacía Española, 2021, pág. 15)

ii) Mayor acercamiento entre el sector legal y el tecnológico

El auge del sector *LegalTech* (la unión de ambos sectores) es innegable y cada vez su crecimiento será mayor; sin embargo, lo que se echa en falta es la acogida por parte de los verdaderos operadores jurídicos de las tecnologías disponibles. De nada sirve que existan docenas de softwares especializadas si no se utilizan.

Una de las soluciones que se proponen en el Informe para facilitar este acercamiento es la introducción de expertos tecnológicos en los departamentos legales o en despachos de abogados, que permite a los expertos tecnológicos conocer las verdaderas necesidades de los usuarios, mientras que estos podrán resolver sus reticencias hacia la tecnología, pues podrán conocer mejor sus utilidades.

⁷ Vid. Capítulo III.2.3

iii) Cambio cultural

Debemos recordar que el 90% de los encuestados consideraba que el principal reto para la implantación de la tecnología es el cambio cultural que debe ocurrir en los departamentos jurídicos para que sea posible.

Se propone como solución a esto una mayor movilización de los profesionales del sector legal. No solo es necesaria una mayor formación en tecnología, es necesario que los usuarios sean conscientes de que el cambio que debe producirse en la práctica jurídica es totalmente beneficioso, en primer lugar, para ellos. Para promover estos cambios es necesario que las asociaciones profesionales hagan visibles estos beneficios.

Es difícil imaginar a un abogado que no sueñe con eliminar tareas repetitivas o con trabajar de forma más eficiente.

3.3. Expectativas de futuro: la transformación digital de las profesiones jurídicas

No podemos despreciar la importancia de la aplicación de las tecnologías en el sector legal, al fin y al cabo, la aparición del *LegalTech* “está transformando los valores, los principios, las políticas, los procesos y las prácticas del mundo jurídico” (Barrio Andrés, 2019). De hecho, el mismo Barrio Andrés (2019) considera que el mundo jurídico puede cambiar en las próximas dos décadas más de lo que lo ha hecho en los últimos dos siglos, única y exclusivamente gracias a la intervención e implementación de tecnologías como la inteligencia artificial. En definitiva, no estamos tratando un tema de relevancia menor para los profesionales del sector.

Considera este autor que el objetivo fundamental de la implantación de tecnologías debe ser “conseguir una mayor competitividad y una mejora de la productividad” (Barrio Andrés, 2019, pág. 2), en definitiva, mejorar la calidad del servicio que se presta. Tal y como podrá verse en la entrevista del **Anexo I**, lo que los profesionales del Derecho buscan es reducir (o incluso llegar a eliminar) una serie de tareas concretas, las más rutinarias. Esto permitiría a los profesionales dedicarse a las tareas de mayor valor añadido.

Tal y como expone Barrio Andrés (2019), podemos dividir las tareas legales en dos grandes grupos: i) las que requieren de habilidades humanísticas (como la creatividad y la detección y gestión de emociones) y ii) las repetitivas y rutinarias.

En el primer grupo podemos encontrar ejemplos como “el asesoramiento y la comunicación con los clientes, la preparación y comparecencia ante los tribunales, las entrevistas con los demandantes para obtener información relacionada con los procedimientos judiciales o la investigación académica” (Barrio Andrés, 2019, pág. 3). En estos casos, la automatización es mucho más compleja y el papel del abogado tradicional seguirá siendo elemental.

En el segundo grupo encontramos otras tareas como la gestión y revisión documental o la facturación, procesos que no son verdaderamente jurídicos pero que son indispensables para las tareas de mayor valor añadido. Especialmente destacan los procesos de *due diligence* o el *e-discovery* (de los que hemos hablado ya y seguiremos hablando a lo largo del Trabajo) que serán automatizados, al menos en parte (Barrio Andrés, 2019, pág. 3).

La importancia que estas tecnologías tendrán en el sector legal es incalculable, pero es innegable que los abogados estarán satisfechos de olvidarse de las ocupaciones más rutinarias y repetitivas, con el objetivo de centrarse en las creativas y de alto valor.

Es especialmente importante, tal y como señala Barrio Andrés (2019) que los juristas comiencen a conocer el funcionamiento de estas tecnologías. Al tratarse de herramientas de *Machine Learning*, muchas necesitan un entrenamiento, un aprendizaje. Es por esto que los profesionales jurídicos deben conocer las aplicaciones para “garantizar su adecuación al Derecho y su operatividad real” (Barrio Andrés, 2019, pág. 4).

Una conclusión importante que debemos extraer es que la implantación de la digitalización y la Inteligencia Artificial en la profesión jurídica no debe ser vista como una “batalla entre máquinas y personas” (Barrio Andrés, 2019, pág. 4). Más bien debe entenderse que la aparición del *LegalTech* es un facilitador de las tareas más repetitivas y rutinarias que permite mejorar la eficiencia y productividad de los operadores jurídicos. Es por ello que el conocimiento de estas tecnologías y técnicas pasará a ser, en un futuro muy cercano (si no lo es ya) una ventaja competitiva que los profesionales del sector no pueden desdeñar.⁸

⁸ Por esto podemos apreciar como despachos de referencia como Cuatrecasas están apostando por mejorar el conocimiento de estas tecnologías (*vid.* Anexo I)

CAPÍTULO IV. CASOS DE USO

IV.1. TAXY.IO, DERECHO TRIBUTARIO ALEMÁN

Taxy.io es una startup del sector *LegalTech* fundada en 2018 por Daniel Kirch y Sven Weber. El objetivo de este caso era realizar un análisis de las diferentes fuentes que conforman el Derecho Tributario alemán (leyes, recomendaciones o sentencias), tratando de desarrollar en el algoritmo utilizado el entendimiento de estos textos para poder dar, después, respuestas ajustadas a preguntas individuales (Kirch, 2019).

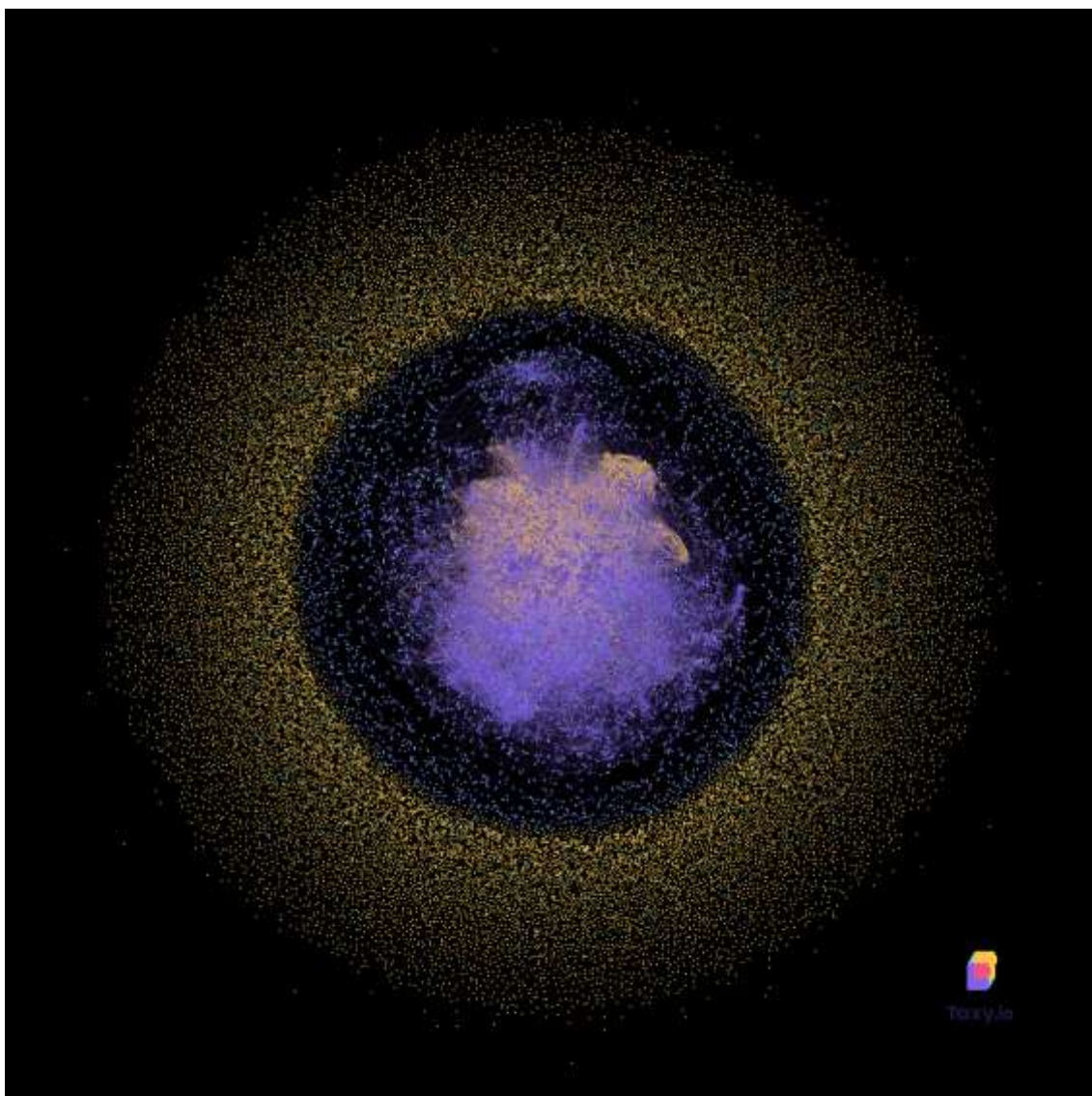
Para comenzar a desarrollar el algoritmo, los fundadores empezaron por analizar como decenas de miles de textos legales estaban relacionados a través de distintas referencias.⁹ Para ello, de acuerdo con Kirch (2019), se analizaron aproximadamente 6.500 leyes y 45.000 sentencias.

Uno de los principales problemas a los que se enfrentaron es que había diferentes formas de referenciar o de referirse a los mismos conceptos o leyes -en el artículo encontramos el siguiente ejemplo: “§3 paragraph 4 sentence 1 Einkommensteuergesetz (Income Tax Act)” or “§3 p. 4 s. 1 EstG” (Kirch, 2019)-. Por ello, tuvieron que desarrollar una gramática inteligente que reconociera las referencias de unos documentos a otros. Finalmente, optaron por analizar cada párrafo o artículo de los textos legales como documentos separados, creando un gráfico en el que cada elemento era uno de estos párrafos o artículos.

La primera representación de esta amalgama de artículos y documentos consistía de 137.000 documentos y hasta 173.000 conexiones entre ellos. Podemos observar esta representación en la **Figura 6**: los puntos amarillos representan las sentencias, los verdes a las leyes y los morados sus diferentes artículos y párrafos.

⁹ Esto puede recordarnos al análisis de correferencias que vimos en el Capítulo III.2.1.2

Figura 6: Representación del Derecho Tributario alemán y sus correferencias (Kirch, 2019)

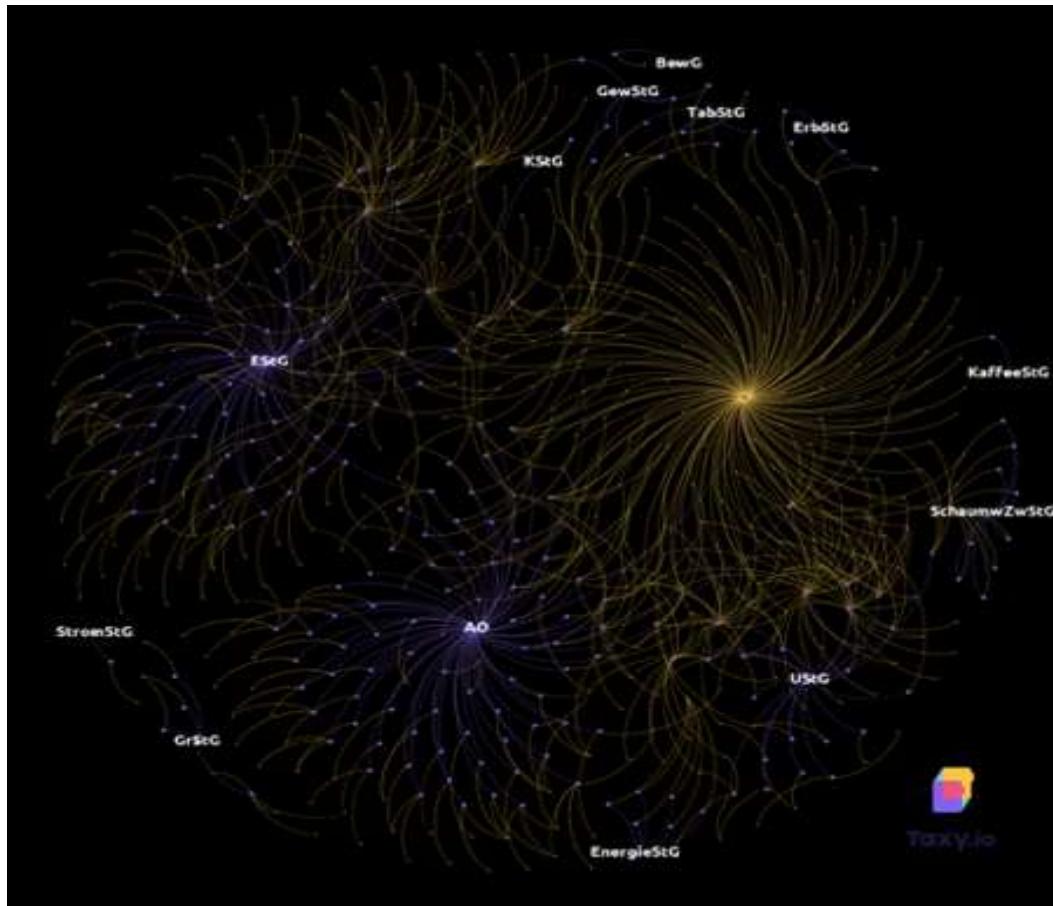


Lo primero que apreciaron es que había dos clases muy diferenciadas de documentos analizados: por un lado, los que no estaban rodeados de líneas/conexiones y que, por tanto, no hacían referencia ni eran referenciados; por otro, una gran masa de puntos muy interrelacionados. El número de conexiones (es decir, de referencias) que recibía cada documento fue utilizado como criterio para medir la importancia de cada uno de ellos, lo que permite trabajar al algoritmo más fácilmente (Kirch, 2019).

Ahondando en el gráfico, el estudio permitió apreciar diferentes *clusters*, grupos de documentos que partían de un núcleo y que también mostraban un alto grado de conexión (puede apreciarse en la **Figura 7**). Entre otras conclusiones, esto permite apreciar que un artículo del AO (Código Tributario Alemán) era citado en numerosas sentencias. Es por

esto que podemos apreciar numerosas líneas amarillas (sentencias) que terminan en un punto morado (artículo o párrafo de una ley).

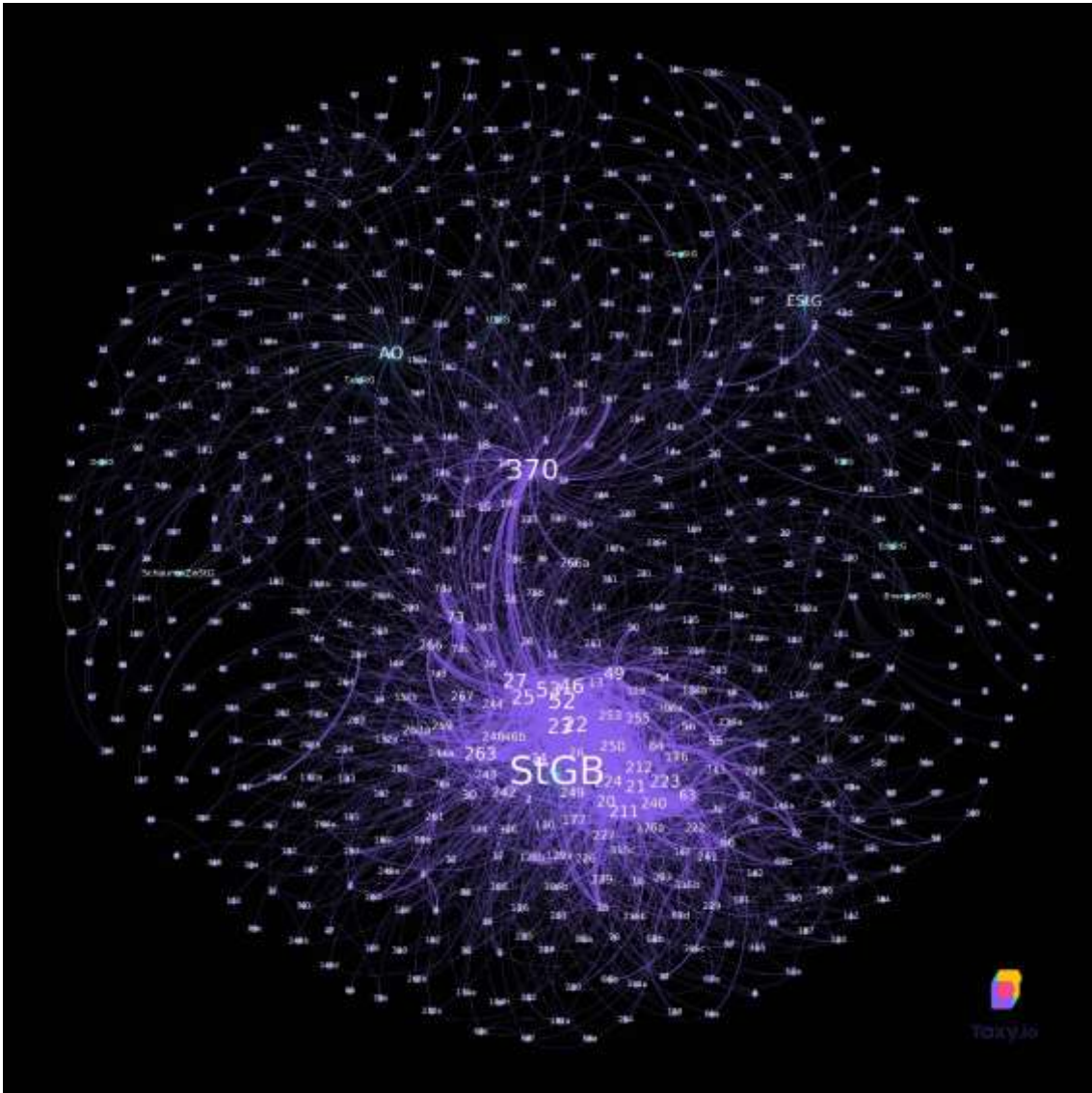
Figura 7: Clusters en el Derecho Tributario alemán (Kirch, 2019)



Todas las técnicas empleadas para realizar este análisis son técnicas de procesamiento de lenguaje natural (Maschek, Natural Language Processing as the automation driver in legal knowledge management, 2020).

En la **Figura 8** podemos ahondar en la conexión de ese artículo 370 AO, apreciando además que muchas de estas conexiones proceden del StGG (el Código Penal alemán). Esta conexión es fácil de entender, pues el art. 370 AO versa sobre la evasión fiscal.

Figura 8: conexiones en Derecho alemán (Maschek, Natural Language Processing as the automation driver in legal knowledge management, 2020).



Otra de las funcionalidades de Taxy.io es la **clasificación de textos** en diferentes subgrupos. El algoritmo de aprendizaje supervisado va relacionando ciertas palabras con subcategorías del Derecho Tributario, así, por ejemplo, agrupa los textos analizados según afectan al IVA o al derecho procesal (Maschek, Natural Language Processing as the automation driver in legal knowledge management, 2020).

Las técnicas de procesamiento de lenguaje natural de Taxy.io van más allá: reconocimiento de entidades o modelación del lenguaje son otros de los ejemplos (Maschek, Natural Language Processing in Legal Tech — Next level applications, 2020).

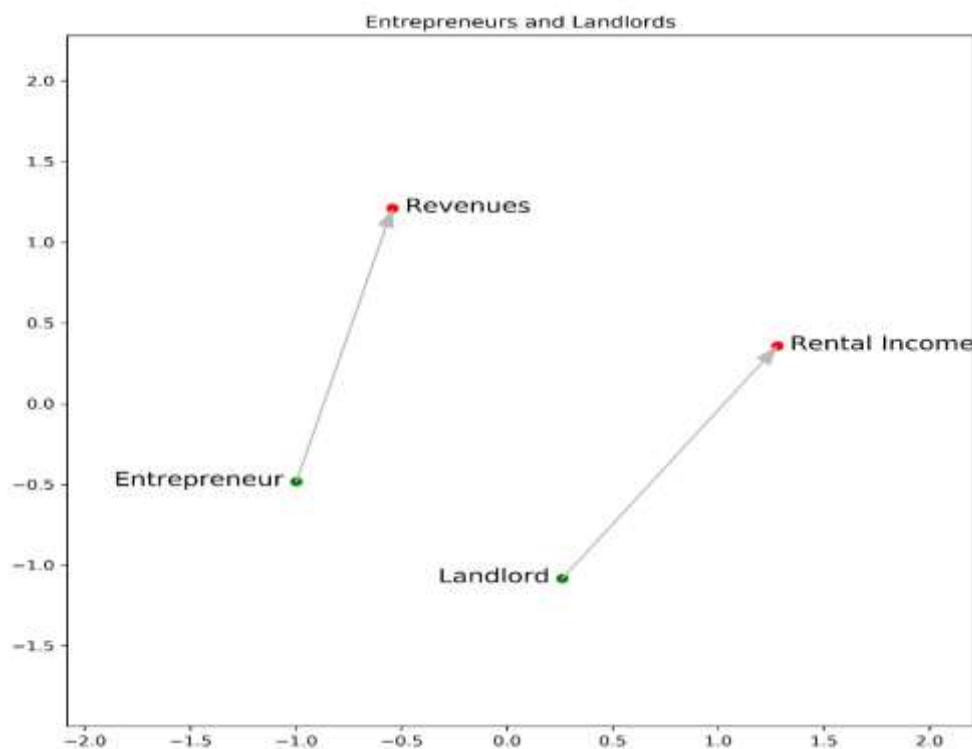
Ya estudiamos en un apartado anterior el reconocimiento de entidades,¹⁰ pero el caso de uso de Taxy.io nos permite apreciar su relevancia práctica.

Esta técnica divide el texto en unidades muy pequeñas (entidades) y esta atomización permite que grupos de términos sean reconocidos, por ejemplo: personas, fechas, lugares, términos técnicos, etc. Una de las utilidades de esta técnica es reconocer términos correferenciados pero que significan lo mismo (Maschek, Natural Language Processing in Legal Tech — Next level applications, 2020).

La modelación del lenguaje es una técnica muy novedosa, en continuo desarrollo. Consiste en la representación vectorial de un grupo de palabras. Esto permite, por ejemplo, establecer relaciones semánticas entre palabras que suelen aparecer juntas en diferentes textos. Para entrenar este tipo de algoritmos, es necesario comenzar por un lenguaje general y luego ir especificando según el sector en el que vayamos a aplicarlo.

Uno de los ejemplos que pudo aprender el algoritmo es que la relación entre empresario e ingreso era similar a la de propietario y renta (puede apreciarse en la **Figura 9**).¹¹

Figura 9: representación vectorial de palabras (Maschek, Natural Language Processing in Legal Tech — Next level applications, 2020)



¹⁰ Vid. Capítulo III.2.1.2

¹¹ Vid. Capítulo II.2.4 y II.2.5

IV.2. ANÁLISIS PREDICTIVO: TRIBUNAL EUROPEO DE DERECHOS HUMANOS

En este apartado vamos a explicar un caso real realizado en 2016 que consiste en un análisis predictivo de las sentencias del Tribunal Europeo de Derechos Humanos.

Se trata de un problema de clasificación binario: lo que este Tribunal determina es si un determinado artículo ha sido o no violado, es decir, si una violación de Derechos Humanos. Esto es lo que debe predecir el algoritmo.

El modelo de *Machine Learning* supervisado nos permite, habiéndole enseñado una suficiente muestra de datos de entrenamiento, determinar cuál será el resultado más probable. Se utiliza un modelo SVM (Support Vector Machine)¹², que una vez entrenado pudo llegar a predecir el resultado (violación o no) con un 79% de precisión, simplemente utilizando información textual, los hechos que rodean al caso y los argumentos de las partes. (Aletras, Tsarapatsanis, Preoțiuc-Pietro, & Lampos, 2016, pág. 2).

Es especialmente interesante para este caso que todas las sentencias tienen la misma estructura¹³ (Aletras, Tsarapatsanis, Preoțiuc-Pietro, & Lampos, 2016, pág. 4):

- Procedimiento: indica los pasos que se siguieron hasta llegar a la sentencia.
- Hechos: comprende todos los puntos relacionados con las circunstancias que rodean al caso.
- Derecho: agrupa los argumentos jurídicos de las partes y el razonamiento del Tribunal, examinando las violaciones de los artículos que se hayan alegado.
- Resultado del caso: se anuncia aquí si se ha decidido que la violación alegada ha ocurrido o no.

El estudio comienza con la creación de un conjunto de datos compuesto por todos los casos relacionados con los Artículos 3, 6 y 8 de la Convención Europea de Derechos Humanos. Se eligieron estos por haber suficiente información disponible para entrenar el modelo (Aletras, Tsarapatsanis, Preoțiuc-Pietro, & Lampos, 2016, pág. 6).

Se eligieron los casos en inglés y se dividió cada texto según la estructura en cuatro partes que indicamos más arriba. Se recogieron 250 casos relacionados con el artículo 3º, 80 con el 6º y 254 con el 8º.

¹² Los algoritmos SVM permiten realizar tareas de regresión y de clasificación. Permite realizar modelos más complejos, pues busca el hiperplano que mejor permita clasificar los datos (Yadav, 2018)

¹³ La hemos simplificado un poco para que el ejemplo quede más fácilmente explicado.

Es importante decir que no se extrajeron para el conjunto de datos las partes relacionadas con el resultado/fallo del caso, de otra forma se estaría pervirtiendo el modelo. Además, se hizo un procesamiento del texto en el que se utilizó la técnica de las *stop words*, eliminando aquellas palabras que no aportaban información semántica relevante (Aletras, Tsarapatsanis, Preoțiuc-Pietro, & Lampos, 2016, pág. 8).

Para desarrollar el modelo se utilizaron dos tipos de elementos: N-gramas y *topics* (clustering de palabras).

Los N-gramas parten del modelo de *Bag-of-Words* que explicamos en el Capítulo II. Así, se eligieron los 2.000 términos más repetidos para cada caso, en cada una de las secciones relevantes. (Aletras, Tsarapatsanis, Preoțiuc-Pietro, & Lampos, 2016, pág. 8)

Los *topics* se crean aplicando técnicas de clusterización a estos N-gramas, permiten agrupar aquellas palabras que son semánticamente similares. La similitud se calculó con la métrica del coseno¹⁴. Finalmente, se obtuvieron 30 clusters de N-gramas (Aletras, Tsarapatsanis, Preoțiuc-Pietro, & Lampos, 2016, pág. 9)

Gracias al uso de clusters, el tamaño de los datos de cada caso se reduce, limitando el posible *overfitting*¹⁵ del conjunto de entrenamiento.

Conseguidos ya los datos que realmente servirían para entrenar el modelo, se decidió emplear el algoritmo SVM para la tarea de clasificación binaria. Así, se clasificó con +1 todos los casos que habían resultado en violación del artículo alegado y con -1 todos los que consideraron que no había violación. Por ello, una vez aplicado el modelo, los pesos positivos serían para aquellos que estuvieron relacionados con admitir la violación y los pesos negativos para los de la no violación.

3.1. Resultados del modelo

La precisión del modelo se calcula con la siguiente forma (casos bien clasificados/total de casos).

Se pudo comprobar que el conjunto de n-gramas obtenidos de la sección “circunstancias del caso” obtuvo los mejores resultados, con un 82% y 77% de precisión prediciendo los casos de los artículos 6 y 8. Respecto del artículo 3, usar el caso completo resultaba en un 70% de precisión, mientras que usar solo las circunstancias resulta en un 68%. Podemos

¹⁴ Vid. Capítulo II.2.5

¹⁵ Sobreajuste de un modelo: efecto de sobreentrenar un algoritmo de aprendizaje supervisado, que se ajusta demasiado al conjunto de entrenamiento y no sabe predecir en el conjunto de testeo.

concluir, por tanto, que las circunstancias del caso nos permiten predecir mejor el resultado de un determinado caso utilizando este modelo (Aletras, Tsarapatsanis, Preoțiuc-Pietro, & Lampos, 2016, pág. 10)

Es especialmente destacable que utilizar los *topics* nos da una muy buena capacidad de predicción en todos los casos (78%, 81% y 76% para cada artículo). Combinando, por tanto, tanto los N-gramas de la sección “circunstancias” y los *topics* seleccionados, podemos obtener un 79% de precisión media para todos los casos (75%, 84% y 78% para caso artículo) (Aletras, Tsarapatsanis, Preoțiuc-Pietro, & Lampos, 2016, pág. 11)

Otro de los análisis que este modelo nos permite hacer es el llamado “*topic analysis*”, que permite obtener cuáles son los términos más frecuentes para cada resultado (violación o no del artículo), una vez obtenidos los pesos de los *topics*. Así, por ejemplo, pueden encontrarse cuáles son los conjuntos de palabras (*topics*) más frecuentes en los casos que aprecian la violación (pesos positivos) y en los que no aprecian violación (pesos negativos) (Aletras, Tsarapatsanis, Preoțiuc-Pietro, & Lampos, 2016, pág. 13)

IV.3. CUATRE CASAS: REFERENTE EN ESPAÑA

3.1. Visión general

Cuatrecasas es uno de los más importantes despachos de abogados de nuestro país. Cuenta con más 1.000 abogados, presencia en una docena de países y aparece siempre reconocido como uno de los líderes del sector legal español. Además, ha destacado en los últimos años por su carácter innovador, habiendo recibido del Financial Times el premio a despacho de abogados más innovador de Europa continental (excluido Reino Unido) en 2018.

Las áreas en las que el Despacho está aplicando Inteligencia Artificial son las siguientes: (Boil Ballesteros, 2019, pág. 3)

- Revisión de documentos para encontrar información relevante
- Búsqueda legal en repositorios de jurisprudencia y/o doctrina
- Análisis de contratos y otros documentos legales
- Revisión de errores y organización documental

Asimismo, se reconocen algunos de los beneficios que la aplicación de estas tecnologías puede implicar para el Despacho: (Boil Ballesteros, 2019, pág. 3)

- Ahorra tiempo
- Produce trabajo de mayor calidad
- Mejora la estructura organizativa
- Permite la mejora del análisis e identificación de precedentes relevantes
- Reduce el estrés en los abogados y mejora las relaciones con los clientes

Cuatrecasas está actualmente enfocado en el desarrollo de dos grandes bloques de trabajo relacionados con la Inteligencia Artificial, el NLP y el *Machine Learning*: i) automatización de tareas de escaso valor y ii) resolución de pequeños problemas legales (Boil Ballesteros, 2019, pág. 4).

A finales de 2015, tras haber rechazado el uso de IBM Watson, Cuatrecasas optó por adquirir una tecnología llamada SEMANTI-K, especializada en el procesamiento de lenguaje natural (Boil Ballesteros, 2019, pág. 4)

Al mismo tiempo, de acuerdo con el Informe ya citado, Cuatrecasas optó por trabajar con KIRA en lo referente a los procesos de *due diligence*, uno de los campos en los que el potencial beneficio procedente del uso de tecnologías es más grande.

Finalmente, en 2017 se empezó a contar con la tecnología de Doxray, basada en tecnología de *Deep Learning* y especializada en revisión de documentos y contratos. Esta tecnología permitía la personalización de la tecnología, adaptándola al caso concreto. Entre otras, se desarrollaron soluciones relacionadas con transacciones inmobiliarias o con procesos de M&A (compraventa de empresas) (Boil Ballesteros, 2019, pág. 5)

En la actualidad, Cuatrecasas está estudiando otras soluciones en vías de desarrollo. Por ejemplo, Cuatrecasas Word Intelligence, una tecnología que pretende ayudar a los abogados en la redacción de documentos, así como otras tecnologías más disruptivas como el diseño de *bots*, relacionadas con el concepto de abogado virtual.

3.2. Entrevista a Idoia Fernández Elorza y Marta Hernández Alvarado, Consejeras de Cuatrecasas, Área de Conocimiento e Innovación

En el **Anexo I** se recogen las respuestas las dos Consejeras del despacho Cuatrecasas dieron a las preguntas que les formulamos. A continuación, expondremos algunas de las principales conclusiones que podemos extraer de ellas, si bien recomendamos encarecidamente su lectura completa.

2.2.1. *¿Cuáles son los objetivos de un equipo de gestión del conocimiento e innovación en una firma de abogados como Cuatrecasas?*

Este equipo está formado por 40 personas de diversa formación (abogados, académicos y técnicos) que refuerzan la innovación y la gestión de los recursos del Despacho. El principal objetivo es el desarrollo de “soluciones tecnológicas innovadoras” que permitan “eliminar tareas o procesos que no generen aportación de valor sostenible”.

Cabe destacar que Cuatrecasas está intentando desarrollar un nuevo perfil de profesional, los llamados *legal engineers*, abogados con formación tecnológica.

2.2.2. *Como responsables del Área de Conocimiento e Innovación de un Despacho como Cuatrecasas, ¿cuál es vuestra percepción del estado de tecnologías de Machine Learning e Inteligencia Artificial en el sector legal, en términos generales?*

Uno de los principales problemas que aprecian es que la mayoría de las tecnologías están mucho más desarrolladas en lengua inglesa. Enseñar un idioma a uno de estos programas requiere de una gran inversión y tiempo, lo que dificulta su implantación.

Destacan cinco fases necesarias para implantar un proyecto de *machine learning* en un despacho de abogados:

- Seguimiento de tecnologías disponibles
- Identificación de casos de uso
- Medición de la necesidad de entrenamiento previo de la tecnología
- Formación de los equipos de entrenamiento
- Puesta en marcha de pruebas piloto

Fernández Elorza y Hernández Alvarado vuelven a señalar que las tareas de entrenamiento son especialmente importantes. Además, exigen la participación de juristas, pues es necesario conocer los conceptos jurídicos que tratan de enseñarse. No solo esto, sino que se vuelve a señalar el problema que supone trabajar con un idioma que no sea el inglés. Una de las causas de esto es que la documentación legal español está menos estructurada/estandarizada, mientras que en el mundo jurídico anglosajón esa estandarización (que es especialmente útil para la tecnología) es mucho mayor.

Se destacan distintos casos de uso en el despacho, destacando por encima de todos aquellos que impliquen un amplio volumen de documentación, ya que es aquí donde la tecnología puede aportar valor. Por ejemplo, se habla del análisis de notas simples del

Registro de la Propiedad, que permite mejorar la práctica en Derecho Inmobiliario o de los procesos de *due diligence*, que exigen un análisis de grandes volúmenes de cláusulas contractuales que pueden ser clave para las operaciones (como las de cambio de control).

Otros ejemplos de casos de uso son: por un lado, las herramientas de análisis estadístico y predictivo que dan información sobre porcentaje de éxito en procesos judiciales (además de la duración o la diferencia entre tribunales) y, por otro, anonimizar documentos, lo que que permite eliminar datos personales y, así, cumplir con la normativa de protección de datos.

2.2.3. ¿Qué proyectos está desarrollando actualmente Cuatrecasas respecto al uso de tecnologías en la práctica jurídica?

Además de las recogidas ya en el apartado anterior, se destacan algunas iniciativas. Por ejemplo, el programa Cuatrecasas Acelera, que busca la identificación de proyectos *Legal Tech* maduros, que permitan desarrollar un piloto e implantar en el Despacho lo que sea más útil.

Señalan, además, una serie de tecnologías que se consideran imprescindibles en el sector legal:

- Programas de gestión documental (con buscadores, archivo de versiones, etc.)
- Sistemas de gestión de casos litigiosos o arbitrales
- Tecnologías que garanticen la seguridad de los datos
- Bibliotecas y repositorios digitales de legislación y jurisprudencia

La implantación de tecnología basada en Inteligencia Artificial tiene un objetivo: mejorar la eficiencia. En esa línea destacan los buscadores que emplean Inteligencia Artificial o programas de revisión documental que analizan, entre otras, las referencias cruzadas (que son especialmente importantes en los textos legales).

Es necesario destacar el desarrollo del Proyecto Cuatrecasas Automated Knowledge Experience, dedicado a la automatización de documentos. Una de las principales ventajas es reducir el tiempo destinado al diseño de los primeros borradores de contratos. El objetivo siempre es el mismo, ser más eficiente y dedicar más tiempo a las tareas de mayor valor.

*2.2.4. ¿En qué áreas del Despacho genera verdadero valor el uso de tecnologías?
¿Cuáles son los beneficios que se buscan?*

De acuerdo con las entrevistadas, las tecnologías más extendidas son “aquellas que automatizan procesos repetitivos, las herramientas para trabajo colaborativo y las que permiten la gestión de grandes volúmenes de documentación”.

El beneficio que se busca siempre es el mismo: ganar eficiencia, pero manteniendo siempre la calidad del servicio.

2.2.5. ¿Cómo ve el futuro de las tecnologías, de aquí a 5 años?

El futuro a corto plazo se ve como un proceso de mayor consolidación de la tecnología, lo que ayudará a producir grandes cambios en el sector; aunque, no consideran que en estos cinco años pueda verse un cambio totalmente disruptivo.

Respecto de la Inteligencia Artificial, creen que el principal reto es que estas tecnologías mejoren en el análisis semántico (es aquí donde un desarrollo mayor de las técnicas de procesamiento de lenguaje natural ayudaría al sector legal), lo que permitiría que fueran capaces de “entender” los textos.

2.2.6. ¿Se imagina un supuesto en el que las tecnologías avanzaran tanto que acabarían sustituyendo a los abogados?

La visión respecto de esta cuestión es muy clara: la tecnología no puede sustituir a los abogados, en general, pero sí a algunos abogados, algunos servicios.

El objetivo que se busca es la sustitución de tareas repetitivas en las que el valor solo está en el volumen (revisión de gran cantidad de documentación). No solo la tecnología puede llegar a ser mejor que los abogados en esto, sino que además permitirá a los abogados dedicarse a las tareas de mayor valor.

Un punto a destacar es que la aplicación de estas tecnologías origina dudas legales; es por esto que el papel de los abogados y resto de juristas no podrá ser sustituido por la Inteligencia Artificial u otras tecnologías. Un ejemplo que nos brindan es el de las tecnologías predictivas, que podrían infringir los derechos fundamentales por su uso: el papel del jurista siempre será indispensable para ello.

Consideran, en conclusión, que la tecnología debe ser una herramienta más para el abogado y que, por tanto, le ayude a prestar un mejor servicio.

CAPÍTULO V. CONCLUSIONES

Cuando pensamos en conceptos como *Big Data*, análisis de datos o *Machine Learning*, nuestra mente, intuitivamente, se imagina números, operaciones matemáticas, tablas interrelacionadas de 0s y 1s.

La evolución de las tecnologías de Inteligencia Artificial y *Machine Learning* nos han permitido aplicar estas técnicas, aparentemente reservadas para los números, a documentos de puro texto. Se trata, sin duda, de aplicaciones menos intuitivas y que requieren de una formación específica en la materia, pero cuyo desarrollo y avance en los últimos años está siendo impresionante.

Este desarrollo ha permitido la llegada de las técnicas de procesamiento de lenguaje natural (o NLP) al sector legal. Las posibilidades que estas técnicas ofrecen a los operadores jurídicos son enormes: revisión de documentos (e incluso redacción automática), búsqueda inteligente de jurisprudencia y legislación, gestión de la información, análisis predictivo, ... y esto solo con la actual madurez del NLP. No podemos imaginar un futuro que no implique que estas posibilidades se expandan y sean cada vez más y más beneficiosas para el mundo jurídico.

Los beneficios son claros: los abogados podrán dedicar mucho más tiempo a las tareas en las que son realmente diferenciales (las puramente jurídicas y que requieren de una capacidad de análisis creativo que las tecnologías no pueden ofrecer), gracias a que las tecnologías que hemos analizado serán capaces de llevar a cabo las tareas más rutinarias, que en algunos tipos de operaciones (como las *due diligence*) comportan una enorme cantidad de horas.

Como ha sido habitual a lo largo de la historia (podemos recordar a los luditas en la Revolución Industrial), la aparición de nuevas tecnologías es vista como una amenaza por los operadores ya asentados en el sector. Nada más lejos de la realidad. Estas tecnologías no pueden sustituir a los abogados, en todo caso podrán sustituir a aquellos que se dediquen a las tareas más repetitivas. No puede costar imaginar que la automatización de estas últimas permita a los abogados centrarse en la creación de valor, ofreciendo un servicio mejor al cliente. El ahorro de tiempo en una profesión como la abogacía comporta un ahorro de costes que permitirá: por un lado, a los clientes, conseguir un servicio más barato y, por otro, a los abogados, encargarse de más casos en el mismo periodo de tiempo.

Respecto de esta amenaza que pueden representar las tecnologías, me gustaría concluir este trabajo con una cita del libro “Así empieza todo: La guerra oculta del siglo XXI”:

Probablemente esta sea la capacidad real de la tecnología, ofrecer un instrumento que ayude al ser humano a tomar mejores decisiones, del mismo modo que el coche autónomo puede ayudar a la conducción humana, pero de momento no tiene la capacidad de sustituirlo. Muy a menudo, la tecnología se convierte en una fantasía dibujada en el papel según la cual las máquinas podrían tomar mejores decisiones que nosotros, basadas en criterios puramente objetivos, incluso sobre asuntos tan personales como nuestra vida privada. Hemos de ser conscientes de que la tecnología, los algoritmos y la inteligencia artificial son meros instrumentos que no deben ser sobredimensionados. Su peligro no consiste sólo en que puedan emplearse de modos torcidos; atribuir a esas técnicas una capacidad de la que carecen también es un riesgo importante.

Se trata de que los programas y la tecnología se ajusten a los usuarios, no al revés; deben ampliar nuestras posibilidades, no restringirlas. Y la idea de fondo, la posición subordinada del ser humano y su conocimiento a una máquina que produciría un saber objetivo, supone construir un mundo paralelo, ineficiente, rígido y autoritario que dista mucho de ser positivo. Muchas de las críticas a los sistemas totalitarios, como el nazismo y el comunismo, se producían por el temor que causaba un sistema poderoso que obligaba a los individuos a doblegarse ante él. Las fantasías tecnológicas no son tan diferentes: también tienen ribetes muy perturbadores. (Hernández Jiménez, 2020)

De los testimonios de referentes del mundo jurídico (como Barrio Andrés o Fernández Elorza)¹⁶ podemos extraer precisamente esta misma conclusión. Las tecnologías, el *Legal Tech*, son una realidad (y sobre todo representan un futuro prometedor) que no deben ser entendidas como una amenaza para el mundo jurídico, sino como una palanca más de creación de valor y de mejora del servicio.

Aquellos que sean capaces de, gracias a su formación tecnológica y su mente abierta a la innovación, entender y aplicar con acierto estas técnicas de procesamiento de lenguaje natural, obtendrán una ventaja competitiva sin parangón respecto al resto de operadores jurídicos.

¹⁶ Vid. Capítulo III.3.3 y Capítulo IV.3

BIBLIOGRAFÍA

- Aletras, N., Tsarapatsanis, D., Preoțiuc-Pietro, D., & Lampos, V. (24 de Octubre de 2016). *Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective*. Obtenido de peerj.com: https://peerj.com/articles/cs-93/?mod=article_inline
- Barrio Andrés, M. (2019). *Hacia la transformación digital de las profesiones jurídicas*. Real Instituto Elcano.
- Barry, W., Daniel Lee, A., & Jacky, Z. (25 de 09 de 2020). *The Immediate Future of AI in Law: An Overview of Natural Language Processing Algorithms*. Obtenido de suls.org.au: <https://www.suls.org.au/citations-blog/2020/9/25/natural-language-processing>
- Boil Ballesteros, P. (2019). *Artificial Intelligence initiatives in a Spanish*.
- Brownlee, J. (7 de Agosto de 2019). *What Are Word Embeddings for Text?* Obtenido de machinelearningmastery: <https://machinelearningmastery.com/what-are-word-embeddings/>
- Consejero General de la Abogacía Española. (2021). *EL SECTOR LEGAL CREE EN LA TECNOLOGÍA*.
- Enzyme Advising Group. (16 de Agosto de 2019). *Natural Language Processing: ¿cómo es la técnica Word Embeddings?* Obtenido de [blog.enzymeadvisinggroup.com](https://blog.enzymeadvisinggroup.com/natural-language-processing): <https://blog.enzymeadvisinggroup.com/natural-language-processing>
- Ghatak, A. (2019). *Deep Learning with R*. Calcuta: Springer.
- Hernández Jiménez, E. (2020). *Así empieza todo: La guerra oculta del siglo XXI*. Madrid: Ariel.
- Hernández, M., & Gómez, J. (2013). Aplicaciones de Procesamiento de Lenguaje Natural. *Revista Politécnica*, 87-96.
- Kirch, D. (17 de Febrero de 2019). *On galaxies, satellites and black holes*. Obtenido de medium.com: <https://medium.com/taxy-io/on-galaxies-satellites-and-black-holes-c88b0be84fb0>

- LawGeex. (14 de Mayo de 2018). *LawGeex Launches 2018 LegalTech Buyer's Guide for In-House Counsel*. Obtenido de LawGeex: <https://blog.lawgeex.com/lawgeex-launches-2018-legal-technology-buyers-guide-for-in-house-counsel>
- LegalTechies. (10 de 1 de 2020). *El concepto de Legaltech y sus variantes*. Obtenido de legaltechies.es: <https://legaltechies.es/2020/01/10/el-concepto-de-legaltech-y-sus-variantes/>
- Li, D. (2020). Improve Business Decision-Making with Data Analytics. *Data Science and Analytics, ILTA*, 10-14.
- López Yse, D. (15 de Enero de 2019). *Your Guide to Natural Language Processing (NLP)*. Obtenido de towardsdatascience: <https://towardsdatascience.com/your-guide-to-natural-language-processing-nlp-48ea2511f6e1>
- Maschek, S. (29 de Enero de 2020). *Natural Language Processing as the automation driver in legal knowledge management*. Obtenido de medium.com: <https://medium.com/taxy-io/natural-language-processing-as-the-automation-driver-in-knowledge-management-f287ebbaeac0>
- Maschek, S. (10 de Febrero de 2020). *Natural Language Processing in Legal Tech — Next level applications*. Obtenido de medium.com: <https://medium.com/taxy-io/natural-language-processing-in-legal-tech-some-more-tools-a8506cc9f3d8>
- Morell Ramos, J. (20 de 7 de 2016). *Legal Tech en España, muy primeros pasos*. Obtenido de terminosycondiciones.es: <https://terminosycondiciones.es/2016/07/20/legaltech-espana-mucho/>
- Nay, J. (2018). *Natural Language Processing and Machine Learning for Law and Policy Texts*. NYU.
- Premonition. (26 de Octubre de 2017). *Moneyball for Law: An Unfair Advantage?* Obtenido de premonition.ai: <https://premonition.ai/moneyball-law-unfair-advantage/>
- Susskind, R., & Trommans, R. (11 de 04 de 2018). *#LegalTech v #LawTech – WTF?* Obtenido de legalgeek.co: <https://www.legalgeek.co/learn/lawtech-legaltech-wtf/>
- The Law Society. (2019). *Lawtech: a comparative analysis of legal technology in the UK and in other jurisdictions*.

The Law Society. (5 de 6 de 2019). *What is lawtech?* Obtenido de lawsociety.org.uk:
<https://www.lawsociety.org.uk/campaigns/lawtech/guides/what-is-lawtech>

Trommans, R. (2017). *Legal AI: a beginner's guide*. Thomson Reuters.

Yadav, A. (20 de Octubre de 2018). *SUPPORT VECTOR MACHINES(SVM)*. Obtenido de towardsdatascience: <https://towardsdatascience.com/support-vector-machines-svm-c9ef22815589>

ANEXOS

ANEXO I. ENTREVISTA SOBRE *LEGALTECH* e INTELIGENCIA ARTIFICIAL EN LOS DESPACHOS DE ABOGADOS

Entrevista realizada a Idoya Fernández Elorza y Marta Hernández Alvarado, Consejeras de Cuatrecasas, Área de Conocimiento e Innovación, en abril de 2021

1. Pregunta: ¿Cuáles son los objetivos de un equipo de gestión del conocimiento e innovación en una firma de abogados como Cuatrecasas?

Respuesta:

El equipo de Conocimiento e Innovación de Cuatrecasas es un equipo multidisciplinar de 40 personas (abogados, académicos y técnicos), dedicados a la I+D+i, con el objetivo de garantizar un asesoramiento jurídico a los clientes de la máxima calidad, gestionar eficientemente los recursos y el conocimiento colectivo, apoyar el desarrollo de negocio de la firma y promover la mejora continua de sus procesos internos, así como la innovación, digitalización y los servicios de valor añadido para clientes.

Con estos objetivos, trabajamos, junto con nuestro equipo de tecnologías, en implantar soluciones tecnológicas innovadoras para eliminar tareas o procesos que no generen aportación de valor sostenible, dentro de nuestra estrategia de centrarnos exclusivamente en servicios de valor añadido.

Junto a los abogados y técnicos de conocimiento de nuestra Área de Conocimiento e Innovación y los ingenieros informáticos del equipo de Tecnologías, estamos desarrollando un nuevo perfil profesional para la implantación de la tecnología en nuestro trabajo y el desarrollo de proyectos de legaltech a los que llamamos “legal engineers”, que son abogados con formación y conocimientos tecnológicos.

2. Pregunta: Como responsables del Área de Conocimiento e Innovación de un Despacho como Cuatrecasas, ¿cuál es vuestra percepción del estado de tecnologías de Machine Learning e Inteligencia Artificial en el sector legal, en términos generales?

Respuesta:

Creemos que el desarrollo es desigual, es desigual dependiendo de las jurisdicciones y desigual dependiendo de los casos de uso. En nuestra experiencia, incluso si las

herramientas son agnósticas desde el punto de vista del idioma, la fiabilidad en idiomas diferentes del inglés todavía no está lo suficientemente avanzada por lo que requiere aún de inversiones y tiempo y dedicación al entrenamiento.

La práctica más desarrollada sería la de análisis de contratos (especialmente para el asesoramiento en due diligence) y el e-discovery.

Para la revisión de documentación legal con tecnologías de machine learning o inteligencia artificial, hemos trabajado en Cuatrecasas en diferentes proyectos en las fases de: (i) seguimiento y *scouting* de nuevas tecnologías aplicables al mercado legal; (ii) identificación de casos de uso (para lo que, a nuestro juicio, es necesario conocer hasta dónde llega la tecnología para gestionar adecuadamente las expectativas); (iii) medición de la necesidad de entrenamiento previo de la máquina para determinar el volumen de recursos necesarios en su entrenamiento; (iv) formación y tutoría de los equipos de entrenamiento de la máquina; y (v) puesta en marcha de pruebas piloto de uso en la firma.

Tras haber trabajado con distintas tecnologías de revisión de documentación legal, mis conclusiones preliminares serían que:

- Todas las tecnologías analizadas requieren una actividad previa de identificación de conceptos legales (cláusulas o contingencias) y un entrenamiento del software en cómo reconocerlos en diferentes tipos de documentos. El entrenamiento del software es un trabajo mecánico de identificación de los conceptos en distintos tipos de documentos, que requiere tiempo y formación jurídica.
- En todos los casos el “starter pack” de reconocimiento de conceptos legales se ha desarrollado solo en inglés, aunque todas ellas pueden aprender en cualquier lengua. Nuestros ensayos han estado dirigidos a la enseñanza en español.
- Teniendo en cuenta la inversión necesaria en la formación de la tecnología en otro idioma, hemos tratado de identificar casos de uso en el que el volumen de la documentación justificara el uso de tecnología para su revisión. Así, hemos realizado un entrenamiento previo de la máquina con conceptos legales en español en el sector inmobiliario (registro de la propiedad español – titularidad y cargas- y cláusulas típicas de contratos de arrendamiento) y contratación mercantil (cláusulas contractuales típicas que son objeto de revisión en un proceso de due diligence; entre ellas, objeto, duración, terminación, penalizaciones, renunciaciones, cesión, cambio de control, ley aplicable o jurisdicción). Para no duplicar las tareas

de formación, hemos trasladado el entrenamiento inicial a las distintas plataformas tecnológicas testadas. Con esa importante inversión en formación, hemos conseguido, en determinados casos de uso, obtener información estructurada a partir de un amplio volumen de información desestructurada (por ejemplo, en el análisis de notas simples del Registro de la Propiedad)

Tras este trabajo, la principal conclusión sería que es necesario un número considerablemente mayor de ejemplos de cada concepto legal de lo inicialmente indicado por el proveedor para llegar a una fiabilidad del 90% o más en nuestro idioma. Frente a los modelos anglosajones, la formación en documentos en idiomas latinos es más compleja. Ello puede ser debido a diversas razones; una de ellas sería, a nuestro juicio, que la documentación legal en español habitualmente es menos estructurada (los ejemplos del mismo concepto legal son a menudo diversos sin seguir un patrón o modelo), lo que dificulta su identificación por parte de la tecnología y la necesidad de muchos más ejemplos para llegar a una fiabilidad que resulte mínimamente adecuada (nunca un 100%, algo que -apuntamos- es difícil de “vender” a un abogado).

Aparte de la tecnología de revisión de documentación legal, también tenemos disponibles herramientas de análisis estadístico y predictivo que ofrecen las editoriales jurídicas. Se trata de bases de datos por suscripción que te dan información estadística sobre porcentaje de éxito de un proceso judicial, duración, diferencias entre tribunales. Antes de empezar un litigio, es posible hacer un análisis con esta tecnología, que te da información estadística sobre qué ha pasado en pleitos similares en los últimos años.

Por último, utilizamos también tecnologías de machine learning para “anonimizar” documentos; para eliminar datos personales. Esto es especialmente relevante para cumplir con la normativa de protección de datos.

3. Pregunta: ¿Qué proyectos está desarrollando actualmente Cuatrecasas respecto al uso de tecnologías en la práctica jurídica?

Respuesta:

En Cuatrecasas tenemos un programa de aceleración de start ups con elevada complejidad jurídica, sobre todo regulatoria de cualquier ámbito, y de tecnología legal, Cuatrecasas Acelera, que acaba de terminar su 5ª edición. Ello nos permite, entre otras cosas, estar al día del estado real de desarrollo en este campo. En los proyectos de legal tech más maduros y que nos parece que son aplicables para nuestra práctica, desarrollamos un

piloto e implantamos aquello que nos resulta útil, mayoritariamente en tecnologías dirigidas a la automatización y mejora de procesos tanto de gestión interna como procesos legales.

Internamente, investigamos soluciones de diversos casos de uso con tecnología de inteligencia artificial y otras. Por supuesto, aplicamos todas aquellas tecnologías cuyo uso ya está extendido en el sector (cloud, gestión documental, automatización de documentos y otras).

Hay alguna tecnología que es necesaria hoy sí o sí en cualquier empresa de servicios jurídicos o despacho de abogados, otras que tenerlas marca la diferencia y otras que hoy están creciendo y desarrollándose. Así son tecnologías necesarias, entre otras:

- Los programas de gestión documental que son herramientas necesarias para gestionar todo el volumen de documentación que manejamos los abogados y que incorporan buscadores, sistemas para archivo de versiones, histórico de documentación, etc.
- Los sistemas de gestión de casos litigiosos o arbitrales.
- Tecnología para garantizar la seguridad de los datos, esto es necesario para cualquier sector, pero en la abogacía es imprescindible por la importancia de la confidencialidad de los datos de nuestros clientes.
- Por supuesto, como cualquier empresa, sistemas de gestión administrativa y de la base de datos de clientes.
- Para llevar a cabo nuestro trabajo también es imprescindible el acceso a información jurídica (leyes y jurisprudencia) y bibliotecas digitales.

Disfrutamos, además, de otras tecnologías que, como apuntábamos, marcan la diferencia ya que nos permiten ser mucho más eficientes. La tecnología lo hace más rápido y mejor que nosotros. Entre estas nos referimos, por ejemplo, a un buscador de última generación que incorpore inteligencia artificial, o a los programas de revisión documental que permiten en pocos minutos editar un contrato complejo antes de enviárselo al cliente: que identifica, por ejemplo, incoherencias, errores en las referencias cruzadas o en las definiciones. Con ellos, hemos olvidado las largas horas que invertíamos en buscar información o en editar documentos complejos.

Por último, nuestro proyecto CAKE (Cuatrecasas Automated Knowledge Experience) está dedicado a la automatización de documentos y nos permite ganar en eficiencia en la

redacción de documentación legal y asegurar la máxima calidad. Con la automatización de nuestros modelos y documentos reducimos significativamente el tiempo que dedicamos para crear un primer esqueleto de contrato y, por tanto, podemos dedicar ese tiempo a una tarea de mucho más valor: afinar, adaptar ese primer borrador a las necesidades de nuestro cliente en ese caso concreto.

4. Pregunta: ¿En qué áreas del Despacho genera verdadero valor el uso de tecnologías? ¿Cuáles son los beneficios que se buscan?

Respuesta:

Hoy por hoy, las tecnologías más extendidas son aquellas que automatizan procesos repetitivos, las herramientas para trabajo colaborativo y la que permiten la gestión de grandes volúmenes de documentación, que incorporan herramientas de gestión documental y de búsqueda de información y datos. Fuera de los procesos exclusivamente legales, también nos parecen indispensables las herramientas de análisis de datos para tomar decisiones más informadas.

En cuanto a los beneficios que buscamos, fundamentalmente ganar en eficiencia manteniendo en todo caso la calidad de nuestros servicios, incorporando a cada paso cualquier herramienta que nos permita hacer las cosas de la mejor manera posible. Buscamos también estrechar la relación con nuestros clientes mediante la comunicación y la colaboración constante, algo para lo que la tecnología es especialmente útil.

5. Pregunta: ¿Cómo ve el futuro de las tecnologías, de aquí a 5 años?

Respuesta:

Salvo que aparezca una tecnología totalmente nueva que ahora no se conoce, no prevemos que a 5 años ocurra algo totalmente disruptivo, pero sí creemos que el uso de la tecnología legal se consolidará y extenderá en nuestro sector, lo que puede producir cambios importantes en los modelos de negocio vigentes en la actualidad.

En particular, en relación con la inteligencia artificial, a nuestro juicio esta tecnología dará un salto cualitativo importante cuando sea capaz de hacer análisis semántico con fiabilidad, cuando sea capaz de “entender” textos, de identificar cuáles son las partes relevantes y de saber que un texto y otro dicen lo mismo, aunque lo hagan con distintas palabras. Una tecnología capaz de ello sí traería consigo una revolución en el sector legal.

6. Pregunta: ¿Se imagina un supuesto en el que las tecnologías avanzaran tanto que acabarían sustituyendo a los abogados?

Respuesta:

La verdad es que no, pero que no haya una tecnología que pueda sustituir “a los abogados” no quiere decir que no pueda sustituir a “algunos abogados” y a algunos servicios. La tecnología sí sustituirá aquellas tareas repetitivas de los abogados en las que el valor solo está en el volumen (por ejemplo, la tecnología sustituirá muchas horas de trabajo de revisión de grandes cantidades de documentación para una auditoría de compra de empresa o para preparar la prueba para un arbitraje). Integrando la tecnología podremos dedicar tiempo a tareas de mucho más valor. Nuestro trabajo será mejor, más interesante.

No obstante, creemos que, con independencia de algunas tareas, en las que las personas somos fácilmente sustituibles, la aplicación de la tecnología, por sí misma, origina tantas dudas legales en su aplicación que requerirán siempre a un abogado o a un operador legal para interpretar o regular los efectos jurídicos y los límites del uso, por ejemplo, de las tecnologías predictivas, que pueden infringir seriamente nuestros derechos fundamentales. Lo que yo veo es que, cuanto más tecnología hay, aumentan más las voces que requieren que su uso se regule más, y no menos, por lo que siempre se necesitarán abogados que contribuyan a esa regulación y la apliquen.

En definitiva, creemos que seguiremos siendo un negocio de personas, dónde las personas seguirán siendo lo más importante y se ayudarán de la tecnología para dar un mejor servicio.