



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Escola Tècnica
Superior d'Enginyeria
Informàtica

Escola Tècnica Superior d'Enginyeria Informàtica
Universitat Politècnica de València

Análisis de sentimiento a nivel aspectual en el dominio de la experiencia de usuario en aprendizaje virtual

TRABAJO FIN DE GRADO

Grado en Ingeniería Informática

Autor: Luis López Cuerva

Tutora: María José Castro Bleda

Directora Experimental: Rosario Sanchis Font

Curso 2020-2021

Resum

Les plataformes d'aprenentatge virtual són una eina popular per a realitzar cursos que permeten ampliar els coneixements del seu alumnat i la seua avaluació està en un estat incipient d'investigació. Així mateix l'anàlisi de sentiment és una disciplina del processament del llenguatge natural que resulta de gran utilitat per a conèixer l'opinió dels usuaris. És per això que en aquest treball es proposa utilitzar l'anàlisi de sentiment per a avaluar l'experiència d'usuari en l'aprenentatge virtual.

Per a realitzar aquesta anàlisi s'analitzen totes les fases del procés, començant per l'elecció d'una plataforma d'etiquetatge per a anotar el corpus, estudiant les eines de processament de llenguatge natural existents en el mercat, i continuant amb l'experimentació que abasta des de la creació d'una ontologia pròpia per a abordar el problema fins a la detecció automàtica dels aspectes i la seua polaritat. El treball finalitza amb la presentació de dues dimensions per a l'anàlisi de l'experiència d'usuari i una anàlisi utilitzant aquestes dimensions.

Paraules clau: Processament del Llenguatge Natural, anàlisi de sentiment, polaritat, experiència d'usuari, ferramentes d'etiquetat, aspectes, UX, plataformes d'aprenentatge virtual

Resumen

Las plataformas de aprendizaje virtual son una herramienta popular para realizar cursos que permitan ampliar los conocimientos de su alumnado y su evaluación está en un estado incipiente de investigación. A su vez el análisis de sentimiento es una disciplina del procesamiento del lenguaje natural que resulta de gran utilidad para conocer la opinión de los usuarios. Es por esto que en este trabajo se propone utilizar el análisis de sentimiento para evaluar la experiencia de usuario en el aprendizaje virtual.

Para realizar este análisis se analizan todas las fases del proceso, comenzando por la elección de una plataforma de etiquetado para anotar el corpus, estudiando las herramientas de procesamiento de lenguaje natural existentes en el mercado, y continuando con la experimentación que abarca desde la creación de una ontología propia para abordar el problema hasta la detección automática de los aspectos y su polaridad. El trabajo finaliza con la presentación de dos dimensiones para el análisis de la experiencia de usuario y un análisis utilizando dichas dimensiones.

Palabras clave: Procesamiento de Lenguaje Natural, análisis de sentimiento, polaridad, experiencia de usuario, herramientas de etiquetado, aspectos, UX, plataformas de aprendizaje virtual

Abstract

Virtual learning platforms are a popular tool for conducting courses that allow students to expand their knowledge and their assessment is in an incipient state of research. Likewise, sentiment analysis is a discipline of natural language processing that is very useful to know the user opinion. That is why in this work it is proposed to use sentiment analysis to evaluate the user experience in virtual learning.

To carry out this analysis, all the phases of the process are analyzed, starting with the choice of a labeling platform to annotate the corpus, studying the natural language processing tools existing in the market, and continuing with the experimentation that

ranges from the creation of an own ontology to approach the problem until the automatic detection of the aspects and their polarity. The work ends with the presentation of two dimensions for the analysis of the user experience and an analysis using these dimensions.

Key words: Natural Language Processing, sentimental analysis, polarity, User eXperience, annotation tools, aspects, UX, virtual learning environments

Índice general

Índice general	V
Índice de figuras	VII
Índice de tablas	VIII
<hr/>	
1 Introducción	1
1.1 Motivación	1
1.2 Objetivos	1
1.3 Estructura de la memoria	3
2 Estado del arte	5
2.1 Análisis de sentimiento	5
2.2 Extracción de aspectos	6
2.3 Experiencia de usuario en el aprendizaje virtual	7
3 Herramientas de procesamiento del lenguaje natural	9
3.1 Meaning Cloud	9
3.2 Google Cloud	11
3.3 Azure Microsoft Text Analytics	12
3.4 Comparativa herramientas NLP	13
4 Herramientas de etiquetado	15
4.1 brat	15
4.2 Doccano	16
4.3 INCEpTION	17
4.4 Comparativa de herramientas	17
5 Análisis de sentimiento a nivel de aspecto	21
5.1 Introducción	21
5.2 Corpus para la experimentación	21
5.3 Creación de ontología	22
5.4 Extracción de aspectos	23
5.5 Etiquetado	25
5.5.1 Creación de etiquetas	25
5.5.2 Preprocesado	26
5.5.3 Proceso de etiquetado	26
5.5.4 Postprocesado	27
5.5.5 Evaluación del etiquetado	28
5.6 Métricas de evaluación de la polaridad a nivel de aspecto	28
6 Experimentación	31
6.1 Herramientas comerciales	31
6.1.1 Creación ontología	31
6.1.1.1 Meaning Cloud	31
6.1.1.2 Google Cloud	31
6.1.2 Uso de herramientas comerciales	32
6.1.2.1 Meaning Cloud	32
6.1.2.2 Google Cloud	33

6.1.3	Procesado de los datos de las herramientas de NLP	34
6.1.3.1	Meaning Cloud	35
6.1.3.2	Google Cloud	35
6.1.4	Experimentación: Polaridad a nivel de aspecto	35
6.1.4.1	Meaning Cloud	36
6.1.4.2	Google Cloud	37
6.2	<i>Cross-domain models</i>	38
6.2.1	<i>Cross-domain models</i>	38
6.2.2	Experimentación: Polaridad a nivel de aspecto	38
6.3	Resultados y análisis	39
6.3.1	Comparativa ontología de etiquetado y ontologías comerciales	39
6.3.2	Distribución por polaridad	42
6.3.3	Distribución por categorías de herramientas comerciales	43
6.3.4	Distribución por aspectos	45
6.3.5	Comparativa evaluación de herramientas	46
7	Experiencia de usuario	49
7.1	Introducción	49
7.2	Herramientas de medición para plataformas interactivas	49
7.2.1	User Experience Questionnaire	49
7.2.2	Modelo de Zaharias y Pappas	50
7.2.3	Requisitos de Ovesleová	51
7.3	Relación entre dimensiones de la UX y nuestra ontología	52
7.4	Estadísticas y análisis de la experiencia de usuario	53
7.5	Conclusiones del análisis de la UX en entornos de aprendizaje virtual con herramientas de análisis de sentimiento para el NLP	55
8	Conclusiones y trabajos futuros	57
8.1	Conclusiones	57
8.2	Evaluación de objetivos	58
8.3	Trabajos futuros	58
	Bibliografía	59
<hr/>		
Apéndices		
A	Guía etiquetado Doccano	63
A.1	Primeros pasos	63
A.2	Anotación	64
A.3	Comentarios	64
A.4	Guía auxiliar	65
B	Aspectos por categoría	67

Índice de figuras

3.1	Interfaz web de Meaning Cloud.	9
3.2	Ontología creada mediante Meaning Cloud.	10
3.3	Demo de Google Cloud.	11
3.4	Demo de Azure Microsoft Text Analytys.	12
4.1	Ejemplo de brat.	15
4.2	Ejemplo de Doccano.	16
4.3	Ejemplo de INCEPTION.	17
4.4	Pantalla de administración de Doccano.	18
4.5	Ejemplo de ayuda visual en brat.	18
6.1	Interfaz web de Meaning Cloud.	32
6.2	Ejemplo de respuesta de Meaning Cloud.	33
6.3	Ejemplo de respuesta de Google Cloud.	34
6.4	Matriz de confusión por polaridad Meaning Cloud.	36
6.5	Matriz de confusión por polaridad Google Cloud.	37
6.6	Matriz de confusión por polaridad <i>cross-domain model</i>	39
6.7	Matriz de confusión por aspectos Meaning Cloud.	40
6.8	Matriz de confusión por aspectos Google Cloud.	40
6.9	Frecuencia de aparición por polaridad en Meaning Cloud.	42
6.10	Frecuencia de aparición por polaridad en Google Cloud.	42
6.11	Frecuencia de aparición por polaridad en <i>cross-domain model</i>	43
6.12	Frecuencia de aparición por categorías en Meaning Cloud.	44
6.13	Frecuencia de aparición por categorías en Google Cloud.	44
6.14	Frecuencia de aparición por entidad en Meaning Cloud.	45
6.15	Frecuencia de aparición por entidad en Google Cloud.	46
6.16	Frecuencia de aparición por entidad en <i>cross-domain model</i>	46
7.1	Estructura general del UEQ.	50
7.2	Dimensiones para evaluar un sistema de aprendizaje virtual.	51
7.3	Distribución relativa de dimensiones.	54
7.4	Distribución relativa de las categorías de la dimensión "Conexiones sociales".	54
7.5	Distribución relativa de las categorías de la dimensión "Material de aprendizaje".	55
A.1	Login.	63
A.2	Elección del proyecto.	64
A.3	Comenzar anotación.	64
A.4	Crear comentarios.	65
A.5	Ver guía auxiliar.	65
A.6	Guía auxiliar.	66

Índice de tablas

1.1	Rúbrica de evaluación de objetivos.	2
3.1	Comparativa herramientas NLP.	13
4.1	Comparativa herramientas etiquetado.	19
5.1	Ejemplos de corpus.	22
5.2	Aparición entidades Meaning Cloud.	23
5.3	Aparición entidades Google Cloud.	24
5.4	Aparición entidades en inglés Meaning Cloud.	24
5.5	Aparición entidades en inglés Google Cloud.	25
5.6	Aparición entidades en inglés Google Cloud.	26
5.7	Distribución del etiquetado con valores absolutos.	27
5.8	Distribución del etiquetado con valores relativos.	27
5.9	Índices de Kappa de Cohen entre etiquetadores.	28
6.1	Relación escala de polaridad etiquetado - escala de polaridad Meaning Cloud.	35
6.2	Relación escala de polaridad etiquetado - escala de polaridad Google Cloud.	35
6.3	Resultados evaluación Meaning Cloud.	36
6.4	Resultados evaluación Google Cloud.	37
6.5	Resultados evaluación <i>cross-domain model</i>	38
6.6	Comparativa frecuencia de aparición por polaridad.	43
6.7	Comparativa evaluación herramientas comerciales.	47
7.1	Relación entre diversas herramientas de medición UX y las categorías de ontología para el análisis de sentimiento en plataformas de aprendizaje online.	53
8.1	Evaluación de objetivos.	58

CAPÍTULO 1

Introducción

En las siguientes subsecciones se va a presentar la motivación para haber elegido realizar un trabajo de final de grado con esta temática, los objetivos de este trabajo y cómo se va a evaluar si estos se han cumplido o no y finalmente se expondrá cuál es la estructura de esta memoria.

1.1 Motivación

El aprendizaje virtual es actualmente una de las modas a la hora de realizar cursos que permitan ampliar los conocimientos formales de sus participantes y los cursos masivos online (más conocidos por sus siglas en inglés, MOOC, provenientes de *massive open online course*) representan las mayores plataformas de este tipo de aprendizaje. MOOC es un término acuñado en 2008 por Stephen Downes y George Siemens para hablar de redes *peer to peer* en las que se distribuía material de aprendizaje [1]. Pero desde entonces el término se ha resignificado para referirse a cursos online gratuitos abiertos a todo el público.

Con la pandemia iniciada en 2020 la importancia de estos cursos ha aumentado exponencialmente, ya que una de sus características principales es que se pueden realizar a distancia, y además los alumnos que optan por este tipo de educación cada vez la valoran más positivamente, llegando incluso a apreciar en mayor medida estos cursos que los estudios tradicionales [2]. Pero pese a esto muchas plataformas no han invertido en el desarrollo de una experiencia de usuario satisfactoria para sus alumnos, es por esto que este trabajo busca una forma de analizar las opiniones de los usuarios de este tipo de cursos con la finalidad de analizar qué aspectos de la plataforma es prioritario mejorar.

1.2 Objetivos

El trabajo aquí presentado consta de cuatro objetivos complementarios. El primero de ellos es la investigación y elección de una herramienta que permita el etiquetado manual de grandes corpus de datos, el segundo es la investigación, análisis y uso de diversas herramientas de procesamiento de lenguaje natural que permiten realizar análisis de polaridad a nivel aspectual, el tercero es etiquetar el conjunto de datos, el cuarto es realizar un análisis de la experiencia de usuario mediante la información obtenida de la experimentación realizada.

Para evaluar el grado de cumplimiento de estos objetivos se plantea la siguiente rúbrica 1.1, la cual se utilizará al final del trabajo para ver hasta qué punto se han cumplido los objetivos.

Objetivos	Objetivo no cumplido	Objetivo trabajado insuficientemente	Objetivo alcanzado suficientemente	Objetivo cumplido
Investigar y elegir de plataforma de etiquetado	No se han investigado plataformas de etiquetado.	Se ha elegido una plataforma de etiquetado.	Se han investigado varias plataformas de etiquetado.	Se han investigado y comparado varias plataformas de etiquetado para finalmente utilizar una durante este trabajo.
Investigar, analizar y utilizar diversas plataformas de NLP	No se ha investigado ninguna plataforma.	Se han investigado y analizado múltiples herramientas.	Se han investigado y analizando diversas herramientas. Se han utilizado varias herramientas de NLP.	Se han investigado, analizado, usado y evaluado varias herramientas de NLP.
Etiquetar el conjunto de datos	No se han etiquetado los datos.	Los datos han sido etiquetados por una sola persona.	Los datos han sido etiquetados por varias personas.	Los datos han sido etiquetados por varias personas y se ha validado el etiquetado estadísticamente.
Realizar un análisis de la experiencia de usuario	No se ha realizado un análisis de la experiencia de usuario	Se han empleado los datos etiquetados para realizar un análisis estadístico de la UX.	Se ha realizado un análisis de la UX a partir de las categorías de aspectos.	Se ha realizado un análisis de la UX relacionando los datos obtenidos en este trabajo con otros trabajos punteros del campo.

Tabla 1.1: Rúbrica de evaluación de objetivos.

1.3 Estructura de la memoria

El presente trabajo se divide en ocho capítulos, bibliografía y dos anexos. En cada uno de ellos se presenta el estado actual de la tecnología, un conjunto de herramientas, los análisis realizados, la experimentación o un análisis final de la experiencia de usuario. Concretamente los capítulos presentes son los siguientes:

1. **Introducción:** presenta este trabajo y sus objetivos.
2. **Estado del arte:** expone el estado actual de los diversos estudios relacionados con este trabajo.
3. **Herramientas de procesamiento del lenguaje natural:** introduce diversas herramientas comerciales de procesamiento de lenguaje natural que incorporan la capacidad de realizar análisis de polaridad a nivel aspectual.
4. **Herramientas de etiquetado:** ilustra un conjunto de herramientas que permiten el etiquetado humano de corpus de datos.
5. **Análisis de sentimiento a nivel de aspecto:** describe el corpus de datos utilizado, el proceso de etiquetado y las métricas utilizadas en la evaluación.
6. **Experimentación:** detalla el conjunto de experimentos llevados a cabo.
7. **Experiencia de usuario:** desarrolla los análisis basados en la experiencia de usuario realizados con los datos obtenidos mediante la experimentación.
8. **Conclusiones y trabajos futuros:** pormenoriza las conclusiones obtenidas y los posibles trabajos futuros que ampliarían la investigación realizada.

Después de estos capítulos se listan las referencias bibliográficas que se han utilizado en el trabajo. La memoria finaliza con dos anexos, el anexo **A** que presenta la guía que hemos desarrollado para la herramienta de etiquetado utilizada, y el anexo **B** que describe todas las entradas de nuestra ontología.

CAPÍTULO 2

Estado del arte

En este capítulo presentaremos el estado de la cuestión en el procesamiento del lenguaje natural, conocido por sus siglas en inglés NLP, provenientes de *Natural Language Porcessing*, prestando especial atención al análisis de sentimiento y a la extracción de entidades o aspectos en textos no estructurados. Finalmente, también estudiaremos el estado actual de las investigaciones en experiencia de usuario, especialmente aquellas centradas en la usabilidad y la experiencia de usuario en plataformas de aprendizaje virtual.

2.1 Análisis de sentimiento

Desde principios de 2000 las tareas de análisis de sentimiento han visto aumentada su importancia. Esto se debe a que cada vez existe una mayor cantidad de texto no estructurado que es necesario analizar para poder realizar tareas como el filtrado de mensajes, sistemas de recomendación o sistemas inteligentes. En las primeras aproximaciones se emplearon reglas basadas en conocimiento previo para detectar la polaridad, cabe destacar el uso de lexicones estáticos que recopilan palabras que indican subjetividad. No obstante, el gran esfuerzo que requiere la creación de estas reglas creadas de manera manual y la efectividad de los métodos basados en datos han hecho que se desarrollen en mayor medida los métodos de aprendizaje automático. En esta última aproximación destacaron el uso de máquinas vector soporte [3], un modelo basado en aprendizaje supervisado que busca llevar el conjunto de datos proporcionados a un hiperplano donde las muestras sean separables, lexicones con palabras que indicaban subjetividad [4] y etiquetadores basados en modelos de Markov junto con ventanas de búsqueda para encontrar la subjetividad sobre un término concreto [5].

A partir del segundo lustro de los 2000 se amplía el conjunto de técnicas empleadas, empezándose a utilizar clasificadores basados en reglas y aproximaciones estadísticas [6], conjuntos de binomios sentimiento [7] y grupos de evaluación [8]. Entre estos últimos destaca el gran hincapié puesto en la combinación de las técnicas de análisis de sentimiento junto con otras técnicas de procesamiento de lenguaje natural, utilizándose herramientas como los n-gramas, analizadores sintácticos y etiquetados gramaticales.

Entre 2010 y 2015, coincidiendo con el auge de las redes sociales [9], el abanico de herramientas empleadas para realizar el análisis de sentimiento aumenta en gran medida [10]. Se empiezan a analizar contenidos no lingüísticos como los emoticonos presentes en el texto y se hacen escalas psicométricas para analizar las fluctuaciones de emociones en estos. Además de estas técnicas innovadoras se sigue avanzando en el estudio de técnicas con un largo desarrollo como las estadísticas basadas en Bayes o los lexicones que almacenan la polaridad de una palabra.

A partir de 2015 se comienza a usar aproximaciones basadas tanto en lógica difusa como el análisis de sentimiento dirigido por características, clasificadores de máxima entropía y sistemas multilingües [11]. Estos últimos surgen debido a que hoy en día cuando a los usuarios de un sistema se les pide retroalimentación se les permite dar dicha opinión en el idioma en el que se sientan cómodos, por lo que de una misma herramienta de evaluación de experiencia pueden salir comentarios en diversos idiomas. Para lidiar con esto generalmente los sistemas multilingües traducen toda la información de entrada al inglés y después realizan el análisis.

En la actualidad existen dos aproximaciones en las tareas de análisis de sentimiento, la primera de ella es mediante lexicones y la segunda mediante métodos de aprendizaje profundo [12]. La primera de ellas consiste en la creación de amplios lexicones que contengan tantas palabras referidas a sentimientos como sea posible, la segunda aproximación se basa en el entrenamiento de modelos de aprendizaje para clasificar los sentimientos presentes en el texto objetivo. Los modelos más populares de esta última aproximación son las redes neuronales convolucionales (CNN), redes memoria largo-corto plazo (LSTM) y la combinación de ambas. Concretamente para la tarea española TASS 2019 de análisis de sentimiento [13] predominaban los sistemas de aprendizaje profundo y ganó un sistema centrado en una codificación preentrenamiento de palabras empotradas mediante un transformador de codificación para realizar la codificación.

Además de estas técnicas un gran número de soluciones comerciales han proliferado con la finalidad de facilitar la puesta en marcha de sistemas basados en aprendizaje automático que realicen tareas de análisis de sentimiento. Algunas de estas soluciones son Google Cloud [14], IBM Watson [15], Azure Microsoft Text Analytics [16] y Meaning Cloud[17]. Este tipo de productos permiten realizar tareas de análisis de textos como análisis de sentimiento en un amplio abanico de idiomas y dominios, siendo una forma sencilla de obtener resultados competitivos.

2.2 Extracción de aspectos

La extracción de aspectos o entidades es básica para múltiples tareas de procesamiento de lenguaje natural como el análisis de sentimiento, el reconocimiento de los aspectos o entidades o la extracción de relaciones. Esta tarea generalmente se divide en dos sub-tareas: reconocimiento de aspectos y su desambiguación [18, 19].

La primera de estas dos tareas consiste en la búsqueda de los conceptos de los que se habla en el texto. Las tres aproximaciones habituales para esta búsqueda son:

- **Basadas en conocimiento previo:** mediante el conocimiento previo de la tarea se crean reglas y lexicones. Estos lexicones pueden ser tanto estáticos como dinámicos, los primeros tienen puntuaciones de polaridad más fiables, ya que son creados y revisados por personas, mientras tanto los lexicones dinámicos tienen la ventaja de que se pueden adaptar a los cambios y capturar los cambios que se producen de manera natural en los idiomas [20].
- **Aprendizaje automático:** mediante un conjunto de entrenamiento se entrena un modelo que permite extraer los aspectos. Este tipo de aproximaciones requieren una gran cantidad de datos para entrenar el modelo, a cambio tienen mejor rendimiento que las aproximaciones basadas en conocimiento previo [21].
- **Redes neuronales:** utilizan aprendizaje automático, pero en vez de utilizar las reglas de clasificación infieren el modelo de clasificación. Esta aproximación es especialmente eficaz al utilizar características de clasificación ocultas y al no necesitar

ontologías o lexicones específicos de un dominio, a cambio tienen un gran coste computacional y necesitan una gran cantidad de datos [19].

La desambiguación de entidades o aspectos consiste en distinguir si una palabra, generalmente polisémica, pertenece a una entidad o no. Existen dos aproximaciones para realizar esta desambiguación:

- Aproximación tradicional: son el conjunto de aproximaciones que utilizan características diseñadas a mano. Se puede subdividir en:
 - Aproximaciones independientes: cada mención de un aspecto se desambigua independientemente de las otras ocurrencias. Esta aproximación es incapaz de relacionar menciones a lo largo del documento a analizar, pero permite desambiguar términos en documentos que traten sobre diversos temas [22].
 - Aproximaciones colectivas: a la hora de desambiguar los diferentes aspectos que aparecen en un texto se basan en que el aspecto que se está tratando estará relacionado con el tema de todo el documento y no solamente de su contexto propio [23].
- Aprendizaje automático: se entrena un modelo inteligente que extrae las características necesarias para desambiguar los diferentes aspectos en el texto a analizar. Esta técnica permite desambiguar los términos correctamente pese a la existencia de ruido en la frase analizada o que su contexto sea inesperado, a cambio requiere una gran cantidad de datos para entrenar el modelo [24].

2.3 Experiencia de usuario en el aprendizaje virtual

Los sistemas de aprendizaje virtual son el medio principal para el aprendizaje a distancia a través de internet, tanto para empresas y gobiernos como para estudiantes vocacionales [25]. La posibilidad de ofrecer y cursar estudios desde cualquier punto del planeta. Esta posibilidad de que en un mismo curso haya estudiantes de todo el globo plantea un nuevo desafío: crear una experiencia de aprendizaje óptima para estudiantes con experiencias vitales y bagajes socioeconómicos y culturales muy diferentes.

Para poder evaluar de manera estándar estos sistemas se crea el concepto de la experiencia de usuario, más conocido por sus siglas en inglés UX, provenientes de User eXperience; el cual es aplicable tanto a sistemas de aprendizaje virtual como en cualquier otra aplicación se creó la normativa ISO 9241-210:2010 [26], la cual analiza desde la perspectiva de la interacción persona-computador la experiencia de usuario para definirla como todas las emociones, creencias, preferencias, percepciones, respuestas físicas y psicológicas, comportamientos y logros que ocurren antes, durante y después del uso [27]. Dicha norma ISO se actualizó en 2019, creando así la ISO 9241-210:2019 [28].

Los métodos tradicionales alternativos a la ISO para analizar estas plataformas y otros productos interactivos utilizaban el concepto de usabilidad, el cual analiza la facilidad y satisfacción de uso además de la eficiencia y eficacia de empleo [29]. Sin embargo, para poder afrontar los nuevos retos surgidos de los sistemas de aprendizaje virtual se realizaron numerosos estudios con la finalidad de ampliar el concepto de experiencia de usuario (UX) [30].

Gracias a esta labor de investigación sobre el concepto de experiencia de usuario aplicado a los sistemas de aprendizaje virtual han surgido diversas formas de medir la calidad del sistema, por ejemplo, en [25] se evalúan 4 características principales: atributo

pragmático, aprendizaje real, autonomía y relacionados y finalmente motivación y compromiso. Mientras tanto en [31] el número de categorías principales aumenta a nueve: las necesidades de los docentes, las habilidades de los docentes, la actitud de los docentes, la calidad pragmática: usabilidad del sistema, la adecuación pedagógica de la plataforma, la calidad hedonista de la plataforma, la organización, la estructura administrativa y el soporte técnico y tecnologías disponibles. Además de estas métricas existen cuestionarios como el *User Experience Questionnaire*, más conocido como UEQ [32] creados para evaluar automáticamente la experiencia de usuario en los entornos de aprendizaje virtual y variaciones de este como la propuesta en [33], en las que al cuestionario UEQ se le añaden preguntas sobre datos sociodemográficos y un campo de texto abierto.

CAPÍTULO 3

Herramientas de procesamiento del lenguaje natural

Actualmente en el mercado existen diversas herramientas de Procesamiento de Lenguaje Natural que permiten realizar análisis de sentimiento sobre textos. Estas herramientas emplean ontologías y taxonomías genéricas para extraer información de los textos y clasifican su polaridad mediante distintas métricas.

Para este trabajo hemos investigado tres herramientas comerciales [34]: Meaning Cloud [17], Google Cloud [14] y Azure Microsoft Text Analytics [16], y hemos procedido a su posterior comparación.

3.1 Meaning Cloud

Meaning Cloud [17] es una empresa especializada en la extracción de conocimiento a partir de textos de diversas características. Sus productos se ofrecen siguiendo una filosofía *Software As a Service* mediante un conjunto de interfaces de programación de aplicaciones, diversas integraciones con aplicaciones de terceros y una interfaz web similar a la que se puede observar en la imagen 3.1. Concretamente hemos estudiado el uso de la API de análisis de sentimiento prestando especial atención al análisis de polaridad a nivel de aspecto.

The screenshot displays the Meaning Cloud web interface. At the top, there is a 'Demo' header. Below it, a language dropdown is set to 'Español', with 'Limpiar' and 'Enviar' buttons. The input text is 'En general, la plataforma me resulta bonita, fácil de manejar y accesible.' The 'Results' section shows a summary: 'Este es un resumen del análisis de sentimiento obtenido utilizando Sentiment Analysis.' Under 'Global sentiment', it states 'Este texto es Positive con una seguridad del 100 por cien. Las polaridades detectadas muestran agreement. El texto es Subjective y without irony.' Under 'Feature-level sentiment', there is a table with columns for 'Concept', 'Type', 'Subtype', and 'Sentimiento'. The table contains one row: 'plataforma' (Concept), 'Location' (Type), 'Facility' (Subtype), and 'Positive' (Sentimiento).

Concept	Type	Subtype	Sentimiento
plataforma	Location	Facility	Positive

Figura 3.1: Interfaz web de Meaning Cloud.

Sus características principales son las siguientes:

1. Soporte multilingüe.
2. Análisis de sentimiento global.
3. Análisis de sentimiento a nivel aspectual.
4. Identificación de opiniones y hechos.
5. Detección de ironía.
6. Detección de acuerdo y desacuerdo.
7. Creación de modelos personalizados.
8. Creación de ontologías personalizadas mediante diccionarios.
9. Escala de polaridad graduada.

Esta plataforma concretamente utiliza una escala de polaridad básica y simétrica de 5 sentimientos con el añadido de la categoría *None*, que indica ausencia de opinión. La escala es la siguiente:

- **P+**: muy positivo.
- **P**: positivo.
- **NEU**: neutral.
- **N**: negativo.
- **N+**: muy negativo.
- **NONE**: sin polaridad.

La posibilidad de crear ontologías propias mediante diccionarios permite realizar análisis de sentimiento mucho más especializado, ya que permite aplicar el conocimiento previo existente sobre la tarea para la búsqueda de aspectos concretos que los modelos generalistas no detectan. En la imagen 3.2 se muestra la ontología creada mediante esta herramienta para la realización de este trabajo.

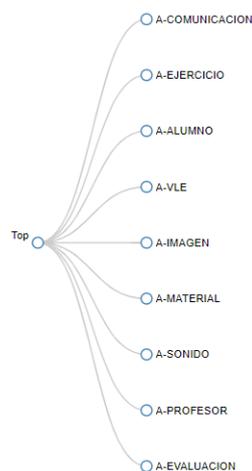


Figura 3.2: Ontología creada mediante Meaning Cloud.

3.2 Google Cloud

Google Cloud [14] es un conjunto de herramientas desarrolladas por Google para desarrollar aplicaciones en menos tiempo y ofrecer un conjunto de soluciones a problemas empresariales complejos. Una de estas herramientas es *Natural Language* [35]. Esta API permite aplicar un amplio abanico de técnicas de procesamiento de lenguaje natural sobre los textos que queramos. Una de dichas técnicas es el análisis de polaridad a nivel de aspectos, de la cual hay una versión de prueba que se puede observar en la figura 3.3.

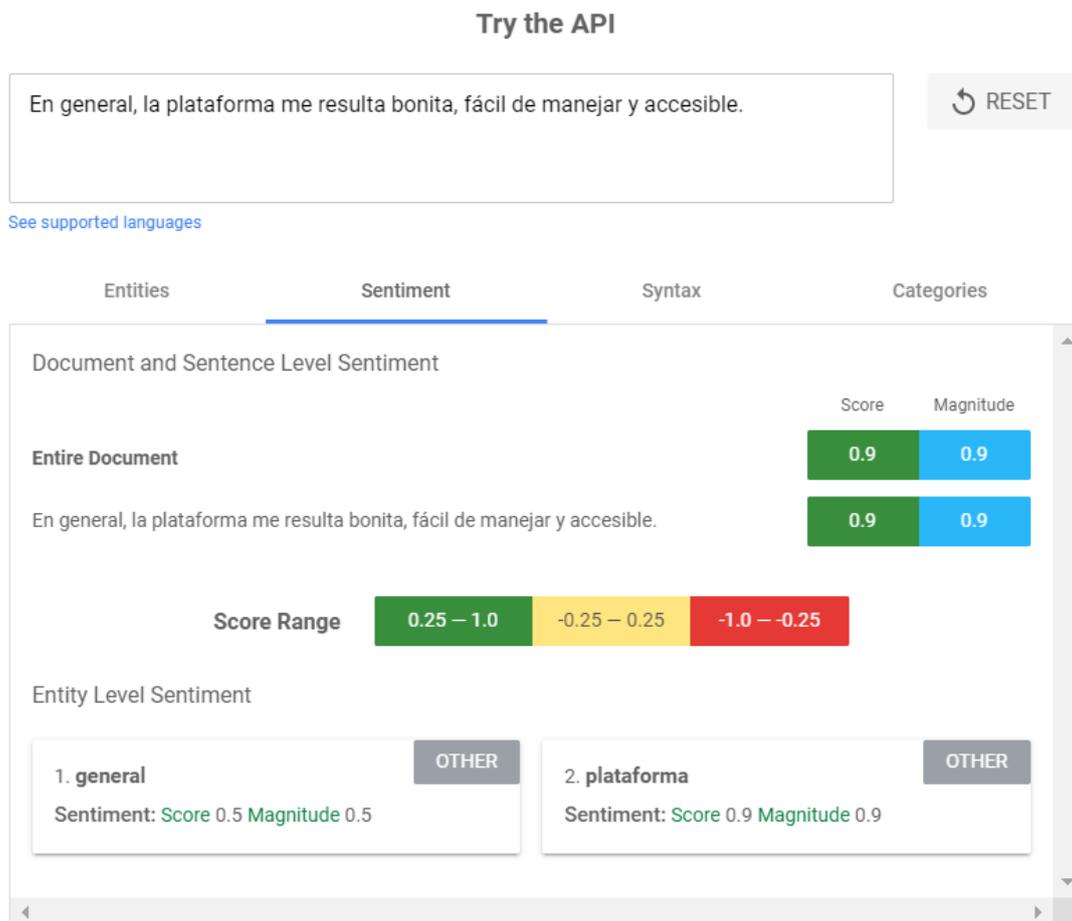


Figura 3.3: Demo de Google Cloud.

Las características principales de esta herramienta son:

1. Análisis de opiniones.
2. Extracción de aspectos.
3. Análisis de opiniones de aspectos.
4. Análisis sintáctico.
5. Clasificación de contenido.
6. Detección automática de idioma.
7. Creación de modelos personalizados.

Para realizar el análisis de sentimientos Google Cloud emplea dos escalas numéricas para describir la opinión presente en el texto. La primera escala es *score*, la cual oscila entre -1.0 (negativo) y 1.0 (positivo) y describe la emoción presente en el texto y la segunda es *magnitude*, que define la intensidad de las emociones presentes y su rango de valores es desde 0.0 hasta $+\infty$.

3.3 Azure Microsoft Text Analytics

Azure Microsoft Text Analytics [16] es un servicio de inteligencia artificial para minería de textos que detecta información, como análisis de sentimiento, aspectos, relaciones y frases clave, en texto no estructurado proporcionado por Microsoft. Está especialmente diseñado para ser integrado con otros servicios de la misma empresa para proporcionar soluciones empresariales. En la imagen 3.4 se puede ver una prueba en línea de este servicio.



Figura 3.4: Demo de Azure Microsoft Text Analytics.

Sus características principales son:

1. Extracción de aspectos.
2. Detección automática de idioma.
3. Procesamiento de datos no estructurados.
4. Extracción de frases claves.
5. Análisis de sentimiento.
6. Minería de opinión.
7. Creación de modelos personalizados.

Concretamente la característica deseable para este trabajo, análisis de polaridad a nivel aspectual, se denomina minería de opinión en este servicio, la cual a fecha de realización de este trabajo está solamente disponible en la versión *v3.1-preview.5* y para textos en inglés.

A nivel de graduación de polaridad Azure Microsoft Text Analytics utiliza una escala de polaridad básica y simétrica que consta de tres etiquetas:

- **Positive:** positivo.

- **Negative:** negativo.
- **Neutral:** neutral.

3.4 Comparativa herramientas NLP

Como se ha visto en los últimos tres apartados cada herramienta tiene unas características y funcionalidades diferentes. Destaca especialmente: la capacidad de personalización de Meaning Cloud gracias a la posibilidad de crear una ontología propia mediante un diccionario y el entrenamiento de modelos propios y la utilización de dos escalas numéricas para caracterizar las opiniones de Google Cloud. En la tabla 3.1 se resumen las principales características de las tres herramientas analizadas. Tal y como se desprende de la comparativa, no hemos podido usar Azure Microsoft porque no soporta modelos en castellano.

Característica	Meaning Cloud	Google Cloud	Azure Microsoft
Creación de diccionarios personalizados	✓	✗	✗
Detección automática de idioma	✓	✓	✓
Detección de ironía	✓	✗	✗
Análisis sintáctico a nivel aspectual en castellano	✓	✓	✗
Escala de polaridad básica	✓	✗	✓
Acceso mediante API	✓	✓	✓
Creación de modelos personalizados	✓	✓	✗
Extracción de aspectos	✓	✓	✓

Tabla 3.1: Comparativa herramientas NLP.

CAPÍTULO 4

Herramientas de etiquetado

Para realizar el etiquetado del corpus con el que se va a trabajar se han investigado diversas posibilidades que permiten una anotación sencilla, para lo cual, en primer lugar se ha realizado una documentación de diversas herramientas de etiquetado [36, 37].

Del conjunto de herramientas descubiertas hemos procedido a probar y analizar en mayor profundidad las siguientes: brat, Doccano e INCEpTION.

4.1 brat

Brat [38] es un entorno online para el etiquetado manual de textos que puede ejecutarse en un servidor local y ser usado desde la web. Está especialmente diseñado para la anotación estructurada y por lo tanto las anotaciones no son de texto libre, sino que se deben prefijar antes de empezar el etiquetado y deben tener una forma fija. En la figura 4.1 se puede muestra un ejemplo de etiquetado mediante esta herramienta.



Figura 4.1: Ejemplo de brat.

Las principales características de Brat son:

- Integración con fuentes externas.
- Soporte multilingüe.
- Interfaz complicada.
- Soporte visual avanzado.
- Anotación estructurada.
- Permite crear relaciones entre etiquetas.
- Manual de usuario.

Para la realización de pruebas con Brat existe una versión de prueba online <http://weaver.nlplab.org/~brat/demo/latest> que es accesible con el usuario "crunchy" y la

contraseña “frog”. Las conclusiones obtenidas de las pruebas realizadas es que el etiquetado en esta plataforma resulta muy complejo y cuenta con numerosas características extras que no son interesantes para este trabajo.

Respecto a la puesta en marcha de un servidor online para realizar el etiquetado se observó un problema, la aplicación está pensada para ser ejecutada en un servidor propio, el cual no disponemos.

4.2 Doccano

Doccano [39] es una herramienta de código abierto que permite la anotación de textos realizada por humanos. Permite realizar anotaciones para tareas de clasificación de texto, etiquetado de secuencias, etiquetado libre, creación de un conjunto de etiquetas para tareas de análisis de sentimiento, reconocimiento de aspectos, etc. Además, Doccano permite un despliegue rápido en múltiples plataformas como Amazon Web Services, Google Cloud y Heroku. En la imagen 4.2 se puede ver un ejemplo de la interfaz de usuario de esta plataforma.

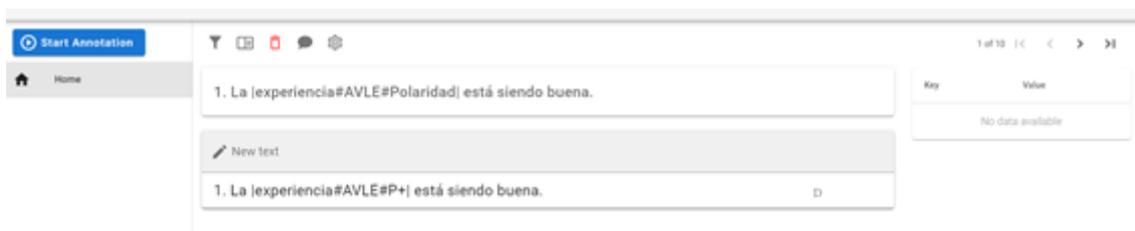


Figura 4.2: Ejemplo de Doccano.

Las principales características de Doccano son:

- Herramienta de anotación colaborativa.
- Soporte multilingüe.
- Soporte web.
- Despliegue sencillo.
- RESTful API.
- Uso sencillo.
- Estadísticas de etiquetado.
- Permite creación de conjuntos de etiquetas.
- Permite múltiples tipos de etiquetados.

En las pruebas realizadas con esta herramienta, la demo se puede encontrar en <https://doccano.herokuapp.com/demo/translation>, hemos podido realizar el proceso de creación de servidor, importación de corpus, etiquetado y exportación del etiquetado en diversos formatos de una manera ágil y sencilla. Además de esto, su puesta en marcha online es sencilla, ya que los propios autores de la herramienta proporcionan despliegues rápidos en plataformas como Amazon Web Services, Google Cloud y Heroku.

4.3 INCEpTION

INCEpTION [40] es una aplicación web que permite a varios etiquetadores trabajar en un mismo proyecto a la vez. Utiliza el navegador como método de acceso a la plataforma de etiquetado y permite acciones tales como etiquetado de archivos pdf, mostrar el análisis estadístico del etiquetado y la exportación del corpus en un amplio abanico de formatos. En la imagen 4.3 se puede ver un ejemplo de la interfaz de usuario.

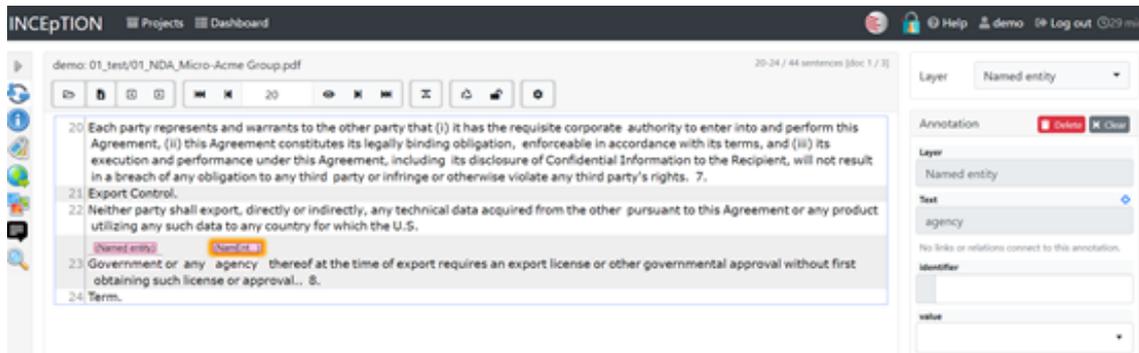


Figura 4.3: Ejemplo de INCEpTION.

INCEpTion, al igual que brat proporciona la posibilidad de probar la plataforma de manera online <https://morbo.ukp.informatik.tu-darmstadt.de/login.html> con las credenciales de usuario: "demo", contraseña: "demo".

Después de probar la plataforma de manera online hemos llegado a la conclusión de que el etiquetado resulta muy complejo a causa de la existencia de capas y referencias. Además, a la hora de realizar el despliegue online de la aplicación observamos el mismo problema que con el despliegue de brat, que requiere de un servidor propio.

4.4 Comparativa de herramientas

Una vez analizadas estas herramientas de etiquetado vamos a proceder a compararlas para elegir la que utilizaremos en este trabajo. Para esta elección analizaremos un conjunto de características deseables y diversos aspectos que resultan interesantes para la interacción con la plataforma como son: la facilidad de uso, puesta en marcha y administración del servidor donde se aloje la herramienta, la escalabilidad o la existencia de diversas ayudas al etiquetado.

Facilidad de uso: para analizar la facilidad de uso de cada plataforma de etiquetado se han empleado las herramientas de pruebas online. De las tres herramientas analizadas la más sencilla de usar es Doccano, seguida por brat y finalmente por INCEpTION.

Facilidad de escalabilidad: si bien el tamaño del corpus de datos actualmente es limitado, ver sección 5.2, de cara a futuros trabajos, ver sección 8, donde exista un mayor conjunto de datos se quiere elegir una plataforma que nos permita escalar el total de los mismos. Todas las opciones cumplen satisfactoriamente con esta característica.

Facilidad de puesta en marcha y administración del servidor: ya que el etiquetado del corpus debe poder ser simultáneo por parte de todos los etiquetadores que participan en el proyecto, la herramienta debe poder alojarse en la web para que exista la posibilidad de acceder a esta sin importar donde esté el usuario. A causa de esta necesidad resulta importante que la herramienta elegida permita poner en marcha y solucionar cualquier

error del arranque fácilmente. Doccano vuelve a ser la opción que presenta una mejor adaptación a esta necesidad.

Figura 4.4: Pantalla de administración de Doccano.

En la figura 4.4 se puede observar la pantalla de administración de Doccano, desde donde se puede configurar distintos tipos de proyectos de diversas maneras, administrar todas las cuentas de los etiquetadores, aunque pertenezcan a diversos proyectos de etiquetado, predefinir conjuntos de etiquetas para distintos proyectos y la configuración y visualización de los comentarios realizados durante la anotación.

Existencia de ayudas visuales: la tarea de etiquetado es una tarea que puede resultar más sencilla gracias a diversas ayudas visuales como la creación de un código de colores asociados a las etiquetas, de manera que la lectura del etiquetado y la visualización de los datos resulta más sencilla e intuitiva. Este hecho se puede apreciar en la imagen 4.5, donde se muestra un ejemplo de las ayudas visuales al etiquetado ofrecidas por brat.

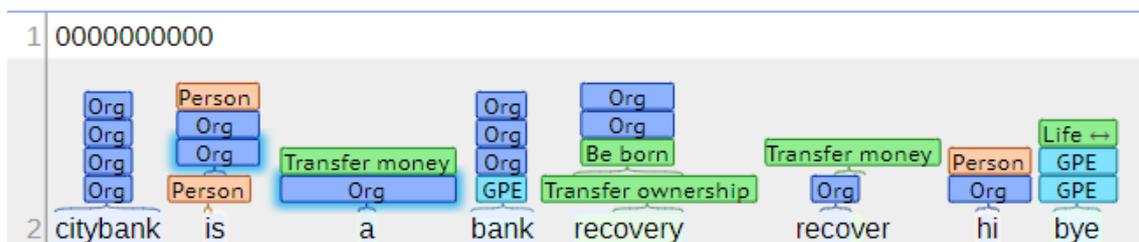


Figura 4.5: Ejemplo de ayuda visual en brat.

Guardado automático del etiquetado: la tarea de etiquetado de un corpus puede llegar a ser tediosa, por lo tanto, resulta interesante que la herramienta empleada tenga algún tipo de ayuda visual para etiquetar. Las tres herramientas analizadas poseen esta característica, no obstante, Doccano no la ofrece para el tipo de etiquetado que se va a emplear.

En la tabla 4.1 se puede ver una comparación de diversas características deseables en la herramienta de etiquetado.

Característica	brat	Doccano	INCEpTION
Uso online	✓	✓	✓
Uso concurrente	✓	✓	✓
Origen de datos .xlsm	✗	✗	✗
Etiquetado nativo intra-frase	✗	✗	✗
Exportación en formato JSON	✓	✓	✓
Estadísticas de etiquetado	✗	✓	✗
Comparación de etiquetados	✓	✗	✗

Tabla 4.1: Comparativa herramientas etiquetado.

Una vez comparadas las tres herramientas analizadas decidimos utilizar Doccano para etiquetar el corpus de este proyecto. Esta elección viene propiciada por la sencillez de anotación, puesta en marcha y administración del servidor donde se ejecuta la aplicación. Además, diversas características como la posibilidad de ver estadísticas de etiquetado como porcentaje de frases etiquetadas le dan un valor añadido muy interesante cualquier proyecto de etiquetado. Finalmente cabe destacar que con la finalidad de facilitar el aprendizaje de esta herramienta y de esta manera agilizar el proceso de anotación se ha creado una guía de etiquetado para este trabajo, ver [Guía etiquetado Doccano](#).

A raíz de esta investigación con Doccano, esta herramienta se está utilizando en otro proyecto del grupo de investigación VRAIN: GUAITA: MONITORIZACION Y ANALISIS DE REDES SOCIALES PARA LA AYUDA A LA TOMA DE DECISIONES, 01/01/2020-31/12/2021. Financiado por: AGENCIA VALENCIANA DE LA INNOVACION, en colaboración con À Punt Mèdia.

CAPÍTULO 5

Análisis de sentimiento a nivel de aspecto

El análisis de sentimiento es una disciplina del procesamiento del lenguaje natural que permite analizar la opinión que tiene el autor de texto. En función del nivel al que se quiera tratar el texto se puede extraer la polaridad de todo el documento [41], la polaridad de cada frase [33] o la polaridad de cada aspecto que aparece en el texto. En este trabajo se aborda el tercer reto.

5.1 Introducción

En las tareas de procesamiento del lenguaje natural la existencia de un corpus de datos amplio y de calidad ayuda a mejorar los resultados en gran medida al habilitar la utilización de un conjunto de técnicas no supervisadas que requieren un gran número de datos, sin embargo el corpus de datos del que disponemos para la realización de este trabajo tiene una extensión limitada, se explica más profundamente en el siguiente subapartado, por lo que para conseguir resultados satisfactorios se debe extraer toda la información posible de este.

Para conseguir expresar al máximo el corpus se han realizado diferentes tipos de procesados tanto antes de realizar la anotación del corpus como después para ampliar la fiabilidad de resultados en todas las fases del trabajo.

En los siguientes cuatro subapartados se presenta el corpus de datos disponible, las diversas técnicas de procesamiento de lenguaje natural y las fases en las que estas han sido aplicadas, concretamente estas fases son: creación de una ontología para la clasificación de aspectos, tareas realizadas relacionadas con la extracción de aspectos, proceso de etiquetado dividido en sus diversas fases. Finalmente, en el quinto subapartado se presentan las métricas de evaluación que se utilizarán durante la experimentación.

5.2 Corpus para la experimentación

El corpus de datos a utilizar se ha obtenido mediante la evaluación de la experiencia de usuario en dos sistemas de gestión de aprendizaje. El primero de ellos es un sistema llamado “Conecto” (tanto en español como en inglés)¹ y el segundo es un curso Moodle

¹https://postgrado.adeituv.es/es/cursos/salud-7/assisted-reproduction/datos_generales.htm

abierto (en español)². Para conseguir los datos concretamente se ha evaluado la UX mediante el cuestionario de experiencia de usuario (UEQ) [42], este formulario consiste en una lista de preguntas de respuesta cerrada que recopilan información sociodemográficas (edad, sexo, etc.) y un campo de respuesta abierta llamado “Otros comentarios”. Este último campo es el lugar perfecto para que los participantes del curso den una opinión más desarrollada y precisa sobre el curso.

Concretamente se han utilizado los datos recopilados en [33], es decir, se han recopilado datos en diferentes ediciones: 2016-17, 2017-18, enero de 2018, abril de 2018 y tanto a mitad como a final de los cursos. Este cuestionario fue dirigido a 559 usuarios, pero solamente 133 personas (37 en inglés y 96 en español) rellenaron el campo “Otros comentarios”. Para este trabajo únicamente utilizaremos los comentarios escritos en español, los cuales forman un corpus consistente en 599 enunciados (entre frases y oraciones).

En la tabla 5.1 se pueden observar varios ejemplos del corpus de datos y sus características según las propiedades que resultan interesantes en este trabajo. Se puede obtener más información sobre la categoría y la polaridad que aparecen en la tabla en la sección 5.5

Frase	Aspecto	Categoría	Polaridad
La experiencia está siendo buena.	experiencia	VLE	Positiva
Los vídeos necesitan que se mejore el audio no se pueden escuchar bien.	vídeos audio escuchar	IMAGEN SONIDO SONIDO	Neutral Negativa Negativa
La valoración general es satisfactoria.	-	VLE	Positiva

Tabla 5.1: Ejemplos de corpus.

En la tabla 5.1 se puede observar cómo algunas frases del corpus contienen aspectos eludidos, los cuales tienen una polaridad que se desea capturar.

5.3 Creación de ontología

Con la finalidad de poder categorizar los aspectos presentes en el texto se ha creado una ontología, basada en el conocimiento previo de la tarea, que consiste en diez categorías que engloban las principales peculiaridades de un sistema de aprendizaje virtual.

- **VLE:** es la categoría más general, dentro podemos encontrar aspectos como curso, entorno, experiencia o interfaz.
- **Alumno:** agrupa los aspectos relacionados con el alumnado.
- **Comunicación:** incluye todas las formas de comunicación, desde el foro, los mensajes, anuncios pasando por las interacciones o el correo electrónico.
- **Ejercicio:** esta categoría está formada por las diversas actividades que se realizan en clase, pero no se evalúan, dentro encontraríamos ejercicios, encuestas, preguntas, tareas o entregables.
- **Evaluación:** agrupa los aspectos relacionados con la realización de la evaluación, la corrección de esta y los elementos obtenidos de esta.
- **Imagen:** engloba los aspectos relacionados con los elementos que permiten visualizar las clases o que señalan algún elemento de la clase. En esta categoría podemos

²<https://medicinagenomica.com/eugmygo/>

encontrar verbos como enfocar o señalar, elementos como puede ser un apuntador o un láser, pantallas o diapositivas.

- **Material:** incluye los recursos que se emplean a la hora de dar las clases, podemos observar en esta categoría material, archivo, módulo, tema y otros aspectos.
- **Profesor:** categoría formada por todas las palabras que hacen referencia al profesorado del curso.
- **Sonido:** agrupa todos los aspectos relacionados con oír las clases, por ejemplo: audio, entender, escuchar.

5.4 Extracción de aspectos

Una vez creada la ontología el siguiente paso fue comprobar que esta cubría todos los aspectos presentes en el texto. Para esto se realizaron dos aproximaciones, la primera fue realizar una extracción de entidades mediante Google Cloud y Meaning Cloud con la finalidad doble de comprobar en que categorías generalistas clasificaban estas herramientas los aspectos detectados y si había algún aspecto detectado que no se pudiera clasificar en ninguna categoría de la ontología creada.

Una vez recuperadas las entidades mediante las dos herramientas observamos que ninguna categoría empleada por estas nos resultaba de utilidad, ya que eran demasiado generalistas y nosotros trabajamos en el dominio de la experiencia de usuario. El siguiente paso fue agrupar por herramienta empleada todos los aspectos que compartieran lema y ordenarlos según frecuencia absoluta de aparición. En las tablas 5.2, 5.3 se pueden ver las diez entidades con mayor frecuencia de aparición según cada herramienta acompañadas de las puntuaciones de relevancia que acumulada y promedio que le otorgan las herramientas.

Entidad	Frecuencia	Relevancia total	Relevancia promedio
curso	180	2792.0	15.51
plataforma	82	1360.0	16.59
clase	63	1020.0	16.19
error	12	120.0	10.00
contenido	11	169.0	15.36
foro	10	200.0	20.00
en línea	9	182.0	20.22
caso	8	96.0	12.00
preguntar	7	105.0	15.00
examen	7	123.0	17.57
estudiante	7	105.0	15.00
docente	7	117.0	16.71

Tabla 5.2: Aparición entidades Meaning Cloud.

En las tablas 5.2 y 5.3 se puede ver como las dos herramientas extraen conjuntos de entidades diferentes pese a tener los mismos datos de entrada. Llama especialmente la atención la diferencia entre el número de veces detectadas las entidades que detectan las dos herramientas.

Como muchas veces las herramientas de procesamiento del lenguaje natural tienen resultados más precisos al trabajar en inglés, debido a la mayor cantidad de datos exis-

Entidad	Frecuencia	Relevancia total	Relevancia promedio
video	44	16.2	0.37
vez	43	10.2	0.24
todo	37	51.8	1.40
experiencia	35	25.1	0.72
general	24	10.2	0.42
nada	22	10.4	0.47
curso	18	22.2	1.23
contenido	14	8.2	0.59
forma	12	6.2	0.52
tema	12	4.6	0.38

Tabla 5.3: Aparición entidades Google Cloud.

tentes en este idioma, se decidió realizar el mismo proceso que el seguido para generar las tablas 5.2 y 5.3 pero traduciendo el corpus al inglés. Los resultados de esta prueba se pueden observar en las tablas 5.4, 5.5.

Entidad	Frecuencia	Relevancia total	Relevancia promedio
platform	243	4725.0	19.44
course	85	1527.0	17.96
stratum	42	794.0	18.90
video	40	727.0	18.18
teacher	31	672.0	21.68
class	21	397.0	18.90
forum	10	174.0	17.40
online	10	198.0	19.80
exam	7	136.0	19.43
opinion	6	118.0	19.67
job	6	146.0	24.33
kind	6	123.0	20.50

Tabla 5.4: Aparición entidades en inglés Meaning Cloud.

En las tablas 5.2, 5.3, 5.4 y 5.5 se puede ver que a priori conceptos existentes en nuestra ontología como curso tienen una gran frecuencia de aparición, sin embargo, palabras como plataforma tienen una importancia significativa, pero no estaban presentes en la ontología original, de manera que se han añadido.

Entidad	Frecuencia	Relevancia total	Relevancia promedio
course	85	36.7	0.43
platform	78	46.6	0.60
everything	50	35.7	0.71
video	39	12.0	0.31
experience	29	25.6	0.88
class	21	4.4	0.21
lot	19	9.0	0.47
content	19	7.3	0.38
teacher	19	4.8	0.25
way	12	5.1	0.43
question	12	2.3	0.19

Tabla 5.5: Aparición entidades en inglés Google Cloud.

5.5 Etiquetado

El etiquetado de datos es un paso imprescindible en los proyectos de clasificación y análisis, ya que es la manera de generar datos que permiten entrenar y validar los sistemas inteligentes que llevan a cabo las clasificaciones necesarias. A continuación, se presenta el proceso completo realizado desde la obtención de los datos hasta los tratamientos efectuados para permitir un posterior uso óptimo de estos.

5.5.1. Creación de etiquetas

Para el etiquetado de este proyecto se han agrupado los aspectos que se quieren detectar en las categorías que forman la ontología creada en el punto 5.3, pero esta clasificación de aspectos en las categorías de la ontología genera algunos problemas que trataremos a continuación.

El primer problema a tratar es la existencia de palabras polisémicas, la problemática con estas es que una misma palabra puede pertenecer a dos o más categorías según que acepción se esté utilizando o incluso que un significado no caiga en ninguna categoría. Un ejemplo de este hecho es ella palabra “experiencia”, la cual puede referirse a la experiencia de usuario y cómo se percibe esta o puede ser utilizado como referencia a la destreza y vivencias de una persona, “Mi experiencia con este tipo de plataformas es limitada”, sin indicar nada referido a la experiencia de usuario. En estas situaciones se ha decidido que dichas palabras se clasificarían en la categoría que sea más propensa a aparecer.

La segunda decisión trata los aspectos eludidos en el texto. Para estos se ha decidido que se tendrán en cuenta en aquellas frases en las que no se perciba ningún otro aspecto y que siempre se clasificarán en la categoría más general, que es VLE.

La tercera cuestión es referente a las palabras que pertenecen a múltiples categorías, por ejemplo “audiovisual”, palabra que inicialmente se encontraba tanto en la categoría Imagen como en la categoría Sonido. Para abordar esta decisión se ha decidido que cada palabra puede aparecer en un máximo de una categoría.

Una vez decididas las medidas adoptadas ante las problemáticas mencionadas anteriormente se definió la versión final de que palabra pertenecía a cada categoría. En el

anexo B se puede ver la lista completa de categorías y que aspectos pertenecen a cada Categoría.

Respecto a la escala de polaridad empleada se ha optado por emplear una escala de polaridad básica y simétrica compuesta por 3 niveles similar a la empleada en [33]. Al igual que en dicho artículo si en una frase, “La interfaz es bonita pero caótica”, aparecen aspectos positivos y negativos sobre un mismo aspecto se etiquetará como neutral. La escala es la siguiente:

- **P:** positivo.
- **NEU:** neutral.
- **N:** negativo.

5.5.2. Preprocesado

Para facilitar el etiquetado del corpus se han detectado de manera semisupervisada todos los aspectos presentes y se han clasificado en su categoría correspondiente, de manera que únicamente sea necesario clasificar la polaridad del aspecto. Concretamente los aspectos detectados se han sustituido por la siguiente estructura: |Aspecto#Categoría#Polaridad|. Por ejemplo:

Frase original: La plataforma está muy bien, es fácilmente entendible y tiene una interfaz bonita.

Frase preprocesada: La |plataforma#AVLE#Polaridad| está muy bien, es fácilmente entendible y tiene una |interfaz#AVLE#Polaridad| bonita.

Frase etiquetada: La |plataforma#AVLE#P| está muy bien, es fácilmente entendible y tiene una |interfaz#AVLE#P| bonita.

Con la finalidad de mantener una coherencia entre los aspectos y el etiquetado en el preprocesamiento que se acaba de describir se ha prestado especial atención al tratamiento de signos de puntuación, de manera que estos no aparecen en las etiquetas.

5.5.3. Proceso de etiquetado

El proceso de etiquetado ha sido realizado por cuatro etiquetadores de manera independiente utilizando **Doccano**. La caja de comentarios ha sido utilizada para indicar cualquier error, duda o información extra sobre el etiquetado.

En la tabla 5.6 se puede ver el tipo de problemas detectados en los comentarios y su frecuencia de aparición.

Problema detectado	Frecuencia
El texto a etiquetar es ambiguo	5
No existe espacio suficiente en la caja de etiquetado para la etiqueta	4
Es aspecto detectado utiliza una palabra polisémica y no debería estar en ese aspecto	2
No se ha detectado un aspecto	2
No se ha detectado un aspecto global debido a la existencia de un aspecto local	1

Tabla 5.6: Aparición entidades en inglés Google Cloud.

Durante el proceso de etiquetado se solucionó el problema de que un etiquetado no cupiera en la caja de etiquetado acordando que delante de estas situaciones se utilizaría una versión reducida de la anotación, en la cual no se guarda la posición de la etiqueta

en la frase, aunque sí se respeta el orden de aparición de etiquetas, y en el postprocesado de la anotación se convertiría en el etiquetado estándar.

5.5.4. Postprocesado

El primer paso del postprocesado ha sido la conversión de etiquetas reducidas al etiquetado estándar. A continuación, se ha analizado el corpus etiquetado y se ha observado la existencia de errores de etiquetado como la existencia de etiquetas incorrectas o el etiquetado doble de un mismo enunciado por una sola persona.

Para solucionar el segundo problema se ha eliminado la repetición y se ha procedido a unificar los tres etiquetados propuestos. La etiqueta final de cada aspecto se elige como la etiqueta más repetida y en caso de empate se reúnen los etiquetadores para discutir la etiqueta final.

Una vez postprocesado todo el etiquetado se han detectado 894 aspectos, la distribución por aspectos y polaridad se puede encontrar en la tabla 5.7 utilizando valores absolutos y en la tabla 5.8 con valores relativos.

Categoría	Positivo	Neutral	Negativo	Total
VLE	402	44	72	521
Alumno	3	7	3	13
Comunicación	1	1	33	35
Ejercicio	0	2	22	24
Evaluación	2	1	19	22
Imagen	15	8	99	122
Material	29	8	68	105
Profesor	5	9	18	32
Sonido	4	1	18	23
Total	461	352	81	894

Tabla 5.7: Distribución del etiquetado con valores absolutos.

Categoría	Positivo	Neutral	Negativo	Total
VLE	44,97 %	4,92 %	8,05 %	57,94 %
Alumno	0,34 %	0,78 %	0,34 %	1,45 %
Comunicación	0,11 %	0,11 %	3,69 %	3,91 %
Ejercicio	0,00 %	0,22 %	2,46 %	2,68 %
Evaluación	0,22 %	0,11 %	2,13 %	2,46 %
Imagen	1,68 %	0,89 %	11,07 %	13,65 %
Material	3,24 %	0,89 %	7,61 %	11,74 %
Profesor	0,56 %	1,01 %	2,01 %	3,58 %
Sonido	0,45 %	0,11 %	2,01 %	2,57 %
Total	51,57 %	9,06 %	39,37 %	100,00 %

Tabla 5.8: Distribución del etiquetado con valores relativos.

En las tablas 5.7 y 5.8 se puede ver como la mayor parte de los aspectos etiquetados son positivos sobre la categoría VLE. Esta categoría acumula el 57,94 % de las etiquetas, siendo la categoría más importante y teniendo una valoración positiva. Las siguientes dos categorías más importantes son Imagen y Sonido, cada una con más de un 10 % de las apariciones y con valoraciones negativas.

5.5.5. Evaluación del etiquetado

Para evaluar la calidad del etiquetado se han calculado las siguientes medidas de concordancia inter-etiquetadores: índice de Kappa de Cohen (κ) [43], el *alpha* de Krippendorff (α) [44] y el valor pi (π) de Scott [45]. Los resultados muestran $\alpha = \pi = 0.663$ y los índices de Kappa de Cohen se muestran en la tabla 5.9.

	Etiquetador A	Etiquetador B	Etiquetador C
Etiquetador A	-	0.758	0.614
Etiquetador B	0.758	-	0.634
Etiquetador C	0.614	0.634	-

Tabla 5.9: Índices de Kappa de Cohen entre etiquetadores.

Para valorar el índice de Kappa y valor de π de Scott se va a seguir la interpretación clásica de Cohen, donde valores ≤ 0 indican ausencia de concordancia, 0 - 0,20 concordancia escasa, 0,21 - 0,4 concordancia media, 0,41 - 0,60 concordancia moderada, 0,61 - 0,8 concordancia sustancial y 0,81 - 1 concordancia completa [46]. De acuerdo con esta interpretación todos nuestros etiquetados han estado sustancialmente de acuerdo mediante la medida del índice de Kappa y moderadamente de acuerdo según el valor de π de Scott.

Respecto al *alpha* de Krippendorff su autor defiende que valores a partir de 0,667 son aceptables [47]. El valor obtenido de *alpha* no llega a ser aceptable, ya que es inferior, pero se encuentra muy cerca, de manera que pese al valor del *alpha* de Krippendorff se valida nuestro etiquetado porque el índice de Cohen Kappa y el valor de π de Scott son satisfactorios.

5.6 Métricas de evaluación de la polaridad a nivel de aspecto

Con la finalidad de evaluar los diversos sistemas de procesamiento de lenguaje natural empleados en esta memoria se van a utilizar diversas medidas para comprobar la calidad de la clasificación. En estas medidas de evaluación Ω representa el conjunto de muestras, Ω_c las muestras de la clase c presentes en Ω , $y(x)$ es la predicción del modelo f para una muestra x , C es el conjunto de clases (positiva, negativa, neutral) y $[\cdot]$ representa los corchetes de Iverson.

Las cinco medidas empleadas en la evaluación son [48]:

1. *Accuracy*: mide la cantidad de casos en los que el modelo ha acertado su predicción.

$$\mathbf{Acc} = \frac{\sum_{c \in C} \sum_{x \in \Omega_c} [f(x) = c]}{|\Omega|} \quad (5.1)$$

2. *Precision*: permite medir cómo de correctas son las predicciones del modelo.

$$\mathbf{P}_c = \frac{\sum_{x \in \Omega_c} [f(x) = c]}{\sum_{x \in \Omega} [y(x) = c]} \quad (5.2)$$

3. *Recall*: mide cómo de completa es la información recuperada.

$$\mathbf{R}_c = \frac{\sum_{x \in \Omega_c} [f(x) = c]}{|\Omega_c|} \quad (5.3)$$

4. *Macro F1*: valor que combina las medidas de precisión y exhaustividad para el conjunto de clases.

$$\mathbf{MF}_1 = \frac{1}{|C|} \sum_{c \in C} F_1^c \quad (5.4)$$

5. *F1 class*: valor que combina las medidas de precisión y exhaustividad para cada clase.

$$\mathbf{F}_1^c = \frac{2 \cdot P_c \cdot R_c}{P_c + R_c} \quad (5.5)$$

CAPÍTULO 6

Experimentación

6.1 Herramientas comerciales

Se ha llevado a cabo un proceso de experimentación con Meaning Cloud y Google Cloud. Dicho proceso ha consistido en realizar un análisis de sentimiento a nivel aspectual con las herramientas analizadas en [Herramientas de procesamiento del lenguaje natural](#) y su posterior evaluación. Las fases en las que se divide dicho análisis son: creación de la ontología, petición, procesado de datos y evaluación de herramientas.

6.1.1. Creación ontología

Meaning Cloud es la única herramienta comercial analizada que permite la creación de ontologías propias, pero tanto esta herramienta como Google Cloud poseen un conjunto de ontologías propias generales.

6.1.1.1. Meaning Cloud

El primer paso para la experimentación con Meaning Cloud ha sido la creación de nuestra ontología, la cual se puede observar en la figura [3.2](#), y añadir a esta las diversas palabras que forman cada categoría, la lista completa de la ontología creada y qué palabras incluye cada categoría se puede encontrar en el anexo [B](#).

Para hacer este paso se ha utilizado la herramienta “Diccionario” de Meaning Cloud, cuya interfaz se puede ver en la figura [6.1](#).

Además de estas categorías personalizadas Meaning Cloud utiliza una ontología general, la cual se puede ver más detalle en [\[49\]](#).

6.1.1.2. Google Cloud

Google Cloud no permite la creación de ontologías personalizadas, su ontología general [\[50\]](#) consta de las siguientes categorías:

- UNKNOWN
- PERSON
- LOCATION
- ORGANIZATION

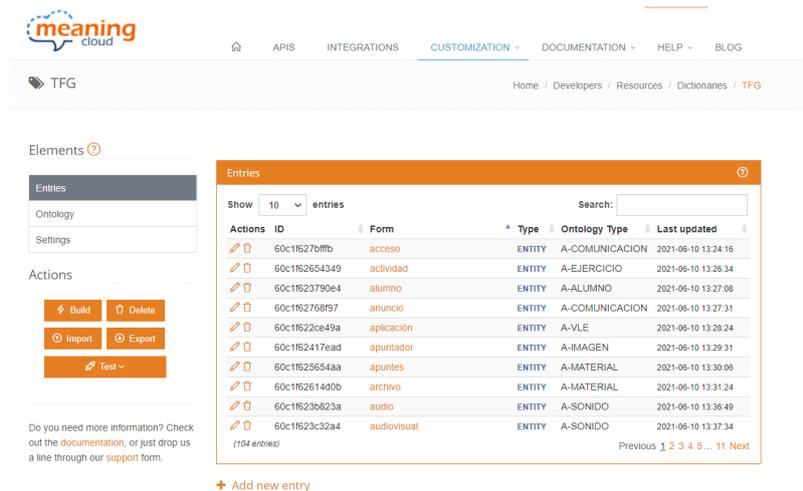


Figura 6.1: Interfaz web de Meaning Cloud.

- EVENT
- WORK_OF_ART
- CONSUMER_GOOD
- OTHER
- PHONE_NUMBER
- ADDRESS
- DATE
- NUMBER
- PRICE

6.1.2. Uso de herramientas comerciales

6.1.2.1. Meaning Cloud

Para recuperar el análisis de sentimiento a nivel aspectual que lleva a cabo Meaning Cloud se ha utilizado su API, pasándole como parámetros el nombre del diccionario creado en la subsección 6.1.1.1 y el corpus de datos codificado mediante UTF-8.

La respuesta recibida consiste en un fichero JSON que contiene información de polaridad a nivel global, de frase y aspectual. Es importante destacar que esta herramienta ha detectado aspectos presentes en el diccionario creado y aspectos no presentes en este, detectados por el modelo de análisis de sentimiento genérico de la herramienta. En la imagen 6.2 se puede ver parte de una respuesta ofrecida por Meaning Cloud.

```
101     },
102     {
103         "text": "es fácilmente entendible",
104         "segment_type": "main",
105         "inip": "63",
106         "endp": "86",
107         "confidence": "100",
108         "score_tag": "P",
109         "agreement": "AGREEMENT",
110         "polarity_term_list": [
111             {
112                 "text": "fácil",
113                 "inip": "66",
114                 "endp": "75",
115                 "confidence": "100",
116                 "score_tag": "P",
117                 "sentimented_entity_list": [
118                     {
119                         "form": "plataforma",
120                         "id": "60c1f622b30ae",
121                         "variant": "plataforma",
122                         "inip": "37",
123                         "endp": "46",
124                         "type": "Top>A-VLE",
125                         "score_tag": "P"
126                     }
127                 ]
128             }
129         ]
130     },
```

Figura 6.2: Ejemplo de respuesta de Meaning Cloud.

6.1.2.2. Google Cloud

Para realizar en análisis de sentimiento de entidades mediante Google Cloud en primer lugar se ha almacenado el corpus de datos en Cloud Storage, ya que nuestro corpus de datos consta de 599 enunciados, pero Google Cloud solamente permite subir junto a cada petición de análisis de sentimiento un documento con un máximo de diez entidades. Una vez almacenado el corpus de datos que se quiere analizar se ha procedido a hacer una única petición mediante la SDK (siglas inglesas de herramientas estándar de desarrollo) de Google Cloud.

La respuesta obtenida del uso de dichas herramienta es un fichero JSON que contiene una lista de las entidades detectadas en el texto, con sus puntuaciones de *score* y *magnitude* y su posición absoluta respecto al documento. En la imagen 6.3 se puede observar parte de una respuesta ofrecida por Google Cloud.

```
1  {
2    "entities": [
3      {
4        "mentions": [
5          {
6            "sentiment": {
7              "magnitude": 0.9,
8              "score": 0.9
9            },
10           "text": {
11             "beginOffset": 38,
12             "content": "plataforma"
13           },
14           "type": "COMMON"
15         },
16         {
17           "sentiment": {
18             "magnitude": 0.9,
19             "score": 0.9
20           },
21           "text": {
22             "beginOffset": 753,
23             "content": "plataforma"
24           },
25           "type": "COMMON"
26         },
27         {
28           "sentiment": {
29             "magnitude": 0.1,
30             "score": -0.1
31           },
```

Figura 6.3: Ejemplo de respuesta de Google Cloud.

6.1.3. Procesado de los datos de las herramientas de NLP

Las dos herramientas utilizadas devuelven ficheros JSON, pero su estructura interna es muy diferente, de manera que se han procesado los datos obtenidos tanto por Google Cloud como por Meaning Cloud para almacenar todos los datos recuperados con el mismo formato y la misma estructura. Es importante destacar que todos los procesos descritos a continuación, se han hecho de manera independiente para cada herramienta, por lo que en ningún momento se han mezclado los datos obtenidos del uso de las diversas herramientas de procesamiento de lenguaje natural.

Para la unificación del formato de los datos se ha decidido separar los datos en función de si los aspectos han sido detectados tanto por la herramienta como en el etiquetado o si no han sido detectados en los dos procesos. Para separarlos se han creado dos ficheros CSV, uno que almacena los aspectos detectados por las dos herramientas y guarda la siguiente información: aspecto, categoría asignada durante el etiquetado, polaridad asignada durante el etiquetado, categoría asignada por la herramienta y polaridad asignada por herramienta. El segundo fichero CSV almacena los aspectos detectados únicamente

por la herramienta cuyos datos se estén procesando y guarda la siguiente información: aspecto, categoría asignada por la herramienta y polaridad asignada por herramienta.

Una vez realizados estos pasos se ha representado de forma gráfica las estadísticas de los datos recuperados. Dicha información se puede ver en el apartado 6.3, concretamente en las figuras 6.4, 6.5, 6.7, 6.8, 6.9, 6.10, 6.12, 6.13, 6.14 y 6.15.

Además de estos pasos generales cada herramienta utilizada ha tenido ciertas peculiaridades en procesamiento, las cuales se explican a continuación.

6.1.3.1. Meaning Cloud

Meaning Cloud utiliza una escala de polaridad que consta de seis elementos, pero nuestro etiquetado contempla solamente tres etiquetas, así que se ha comprimido la escala de polaridad de Meaning Cloud tal y como muestra la tabla 6.1.

Polaridad final	Polaridad de Meaning Cloud
N	N+, N
NEU	NEU, NONE
P	P+, P

Tabla 6.1: Relación escala de polaridad etiquetado - escala de polaridad Meaning Cloud.

6.1.3.2. Google Cloud

Google Cloud utiliza una escala de polaridad numérica doble, donde el valor *score* indica la emoción presente en el texto. Dicha escala numérica se convierte en la escala utilizada en el etiquetado tal y como se muestra en la tabla 6.2.

Polaridad final	Valor inferior de <i>score</i>	Valor superior de <i>score</i>
N	-1	0,1
NEU	0,1	0,3
P	0,3	1

Tabla 6.2: Relación escala de polaridad etiquetado - escala de polaridad Google Cloud.

Estos valores se han calculado de manera empírica, tal y como se recomienda en [35]. Inicialmente los valores de conversión eran $N < -0,25 \leq NEU < 0,25 \leq P$. A partir de estos valores se han variado el valor inferior y superior maximizando el *Accuracy* obtenido hasta conseguir el óptimo alcanzado con los límites de la tabla 6.2.

6.1.4. Experimentación: Polaridad a nivel de aspecto

Para evaluar el análisis de sentimiento a nivel de aspecto en nuestro corpus, se van a emplear las medidas presentadas en el apartado **Métricas de evaluación de la polaridad a nivel de aspecto**. Es importante destacar que, debido a que no podemos evaluar la corrección de una etiqueta que no se ha clasificado en el etiquetado, en este apartado se va a trabajar con un subconjunto de las respuestas obtenidas de las diversas herramientas de NLP utilizadas. Concretamente, se van a utilizar aquellos aspectos detectados por la herramienta que se esté analizando en cada momento y que a la vez se hayan etiquetado en el proceso de anotación.

6.1.4.1. Meaning Cloud

Meaning Cloud ha detectado un total de 806 aspectos, de los cuales 248 han sido detectados también en el etiquetado y 558 solamente han sido detectados por Meaning Cloud.

Utilizando esos 248 aspectos detectados, los cuales forman un 27,74 % de los aspectos etiquetados, obtenemos un *Accuracy* de 0,548 y una *Macro F1* con valor 0,494. De estos 248 aspectos durante el etiquetado 105 fueron clasificados como positivos, 28 como neutrales y 115 como negativos. La matriz de confusión se puede ver en la figura 6.4. El valor del resto de medidas calculadas se puede observar en la tabla 6.3.

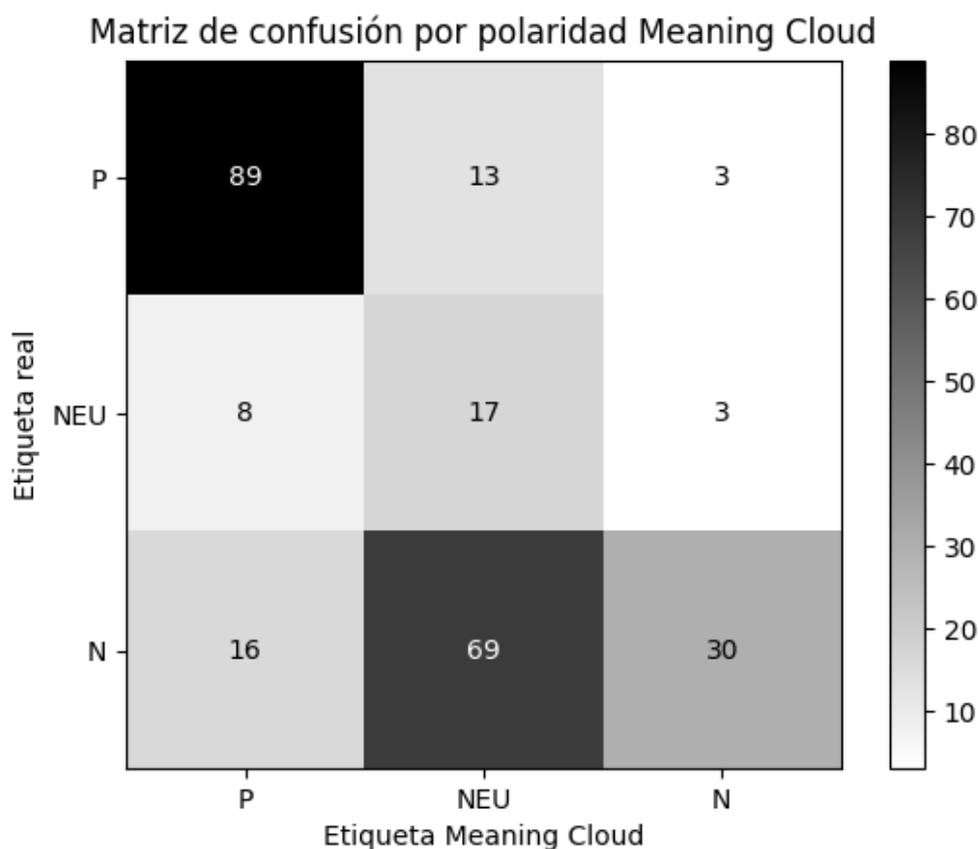


Figura 6.4: Matriz de confusión por polaridad Meaning Cloud.

Clase	Precision	Recall	$F1_c$
N	0,833	0,261	0,397
NEU	0,172	0,607	0,268
P	0,788	0,848	0,817

Tabla 6.3: Resultados evaluación Meaning Cloud.

Meaning Cloud presenta la mayor precisión a la hora de clasificar aspectos negativos, sin embargo, su *recall* es el más bajo de todas las clases, es decir, los aspectos detectados como negativos son negativos en la mayoría de los casos, pero hay un número considerable de aspectos negativos no detectados como negativos que hacen bajar el *recall*. Estos hechos se pueden apreciar de forma gráfica en la matriz de confusión de la figura 6.4.

En la figura 6.4 se puede observar como existen 69 comentarios negativos que se detectan como neutros, motivo que hace bajar la medida del *recall*. Sin embargo, de los 36 aspectos que Meaning Cloud detecta como negativos solamente seis están equivocados, lo que explica el alto nivel de *accuracy*.

6.1.4.2. Google Cloud

Google Cloud ha detectado un total de 1375 aspectos, de los cuales 491 han sido detectados también en el etiquetado y 884 solamente han sido detectados por Google Cloud.

Utilizando los 491 aspectos detectados, los cuales forman un 54,92 % de los aspectos etiquetados obtenemos un *Accuracy* de 0,815 y una *Macro F1* con valor 0,616. De estos 491 aspectos durante el etiquetado 178 fueron clasificados como positivos, 28 como neutrales y 263 como negativos (la matriz de confusión se muestra en la figura 6.5. El valor del resto de medidas calculadas se puede observar en la tabla 6.4.

Clase	Precision	Recall	$F1_c$
N	0,799	0,951	0,868
NEU	0,235	0,080	0,119
P	0,901	0,820	0,861

Tabla 6.4: Resultados evaluación Google Cloud.

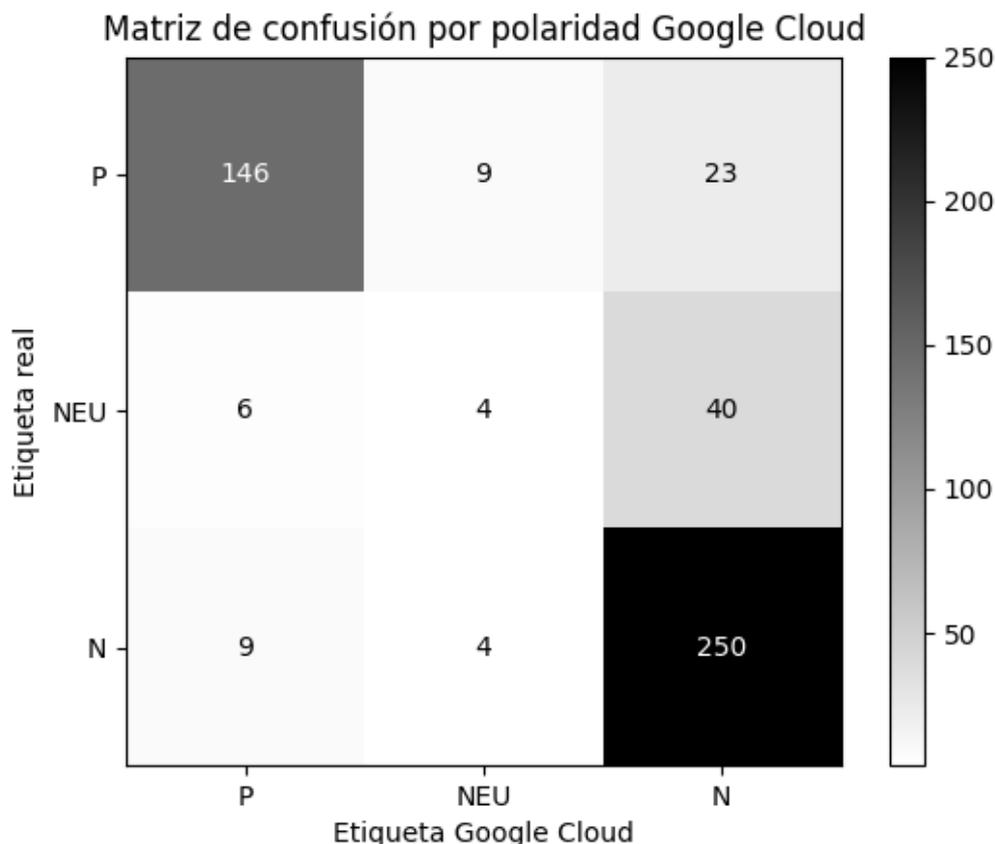


Figura 6.5: Matriz de confusión por polaridad Google Cloud.

Google Cloud presenta un mejor rendimiento con la clase positiva, que posee el mejor valor de *accuracy* a la par que un valor tanto en *recall*. Esto se debe a que, tal y como se

puede observar en la matriz de confusión de la figura 6.5, solamente hay 15 casos de un total de 161 en los que Google Cloud asigna una etiqueta positiva y se equivoca.

También es importante destacar que la clase negativa es la que acapara un mayor número de etiquetas, absorbiendo incluso 40 etiquetas marcadas como neutrales en el proceso de anotación y 23 positivas. Este es el hecho que hace que su *precision* sea inferior a la *precision* de la clase P.

Por último, recalcar que esta herramienta apenas detecta la polaridad neutral, y que cuando lo hace generalmente está equivocada, de hecho solamente ha detectado 17 aspectos negativos, de los cuales 13 han sido identificados erróneamente.

6.2 Cross-domain models

En esta sección se procede a llevar a cabo un análisis de sentimiento utilizando *Cross-domain models*, modelos entrenados en un dominio diferente, para abordar el problema del análisis de polaridad de aspectos en las plataformas de aprendizaje online.

6.2.1. Cross-domain models

Para la experimentación hemos utilizado un modelo entrenado con datos de una tarea de polaridad en Twitter, usando Redes Neuronales Profundas, en particular, *Convolutional Neural Networks*, *Attentional Bidirectional Long Short Term Memory*, y *Transformer Encoders* [51]. Los modelos se han testeado en varias competiciones internacionales [13, 52] y su resultado se pueden consultar en [51].

Concretamente la aproximación propuesta en este trabajo consiste en la determinación de la polaridad de cada aspecto mediante una ventana de longitud fija definida a izquierda y derecha del aspecto. La longitud de dicha ventana se ha determinado mediante los resultados experimentales descritos en el trabajo previo presentados en [18].

6.2.2. Experimentación: Polaridad a nivel de aspecto

El modelo utilizado ha analizado todos los aspectos detectados en el proceso de etiquetado, siendo estos un total de 894 (veánse figuras 5.7 y 5.8). Es importante recalcar que durante el proceso de análisis los aspectos no se clasifican en ninguna ontología.

Utilizando los 894 aspectos detectados obtenemos un *Accuracy* de 0,433 y una *Macro F1* con valor 0,424. De los 894 aspectos analizados 238 fueron catalogados como positivos, 517 como neutrales y 139 como negativos. El valor del resto de medidas calculadas se pueden observar en la tabla 6.5.

Clase	Precision	Recall	F1 _c
N	0,813	0,321	0,460
NEU	0,112	0,716	0,194
P	0,908	0,469	0,618

Tabla 6.5: Resultados evaluación *cross-domain model*.

Este modelo presenta valores de *recall* bajos para la clase negativa y neutral, junto con valores bajos en *precision* para la clase neutral. Estos valores se deben, tal y como se puede observar en la figura 6.6, a que hay una tendencia en el modelo *cross-domain* a clasificar los aspectos como neutrales: 232 aspectos positivos se clasifican como neutrales y 227 negativos se clasifican como neutrales.

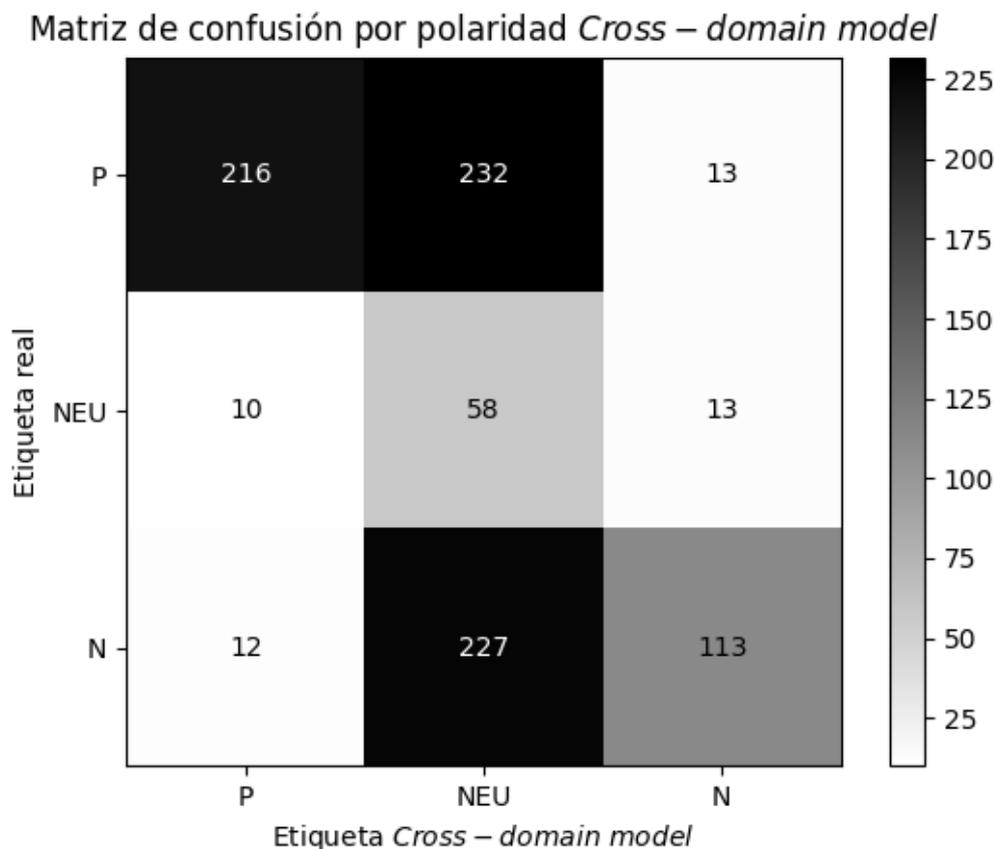


Figura 6.6: Matriz de confusión por polaridad *cross-domain model*.

6.3 Resultados y análisis

A continuación, se presentarán los resultados obtenidos de la experimentación y se analizarán, prestando atención a la polaridad, a las categorías de nuestra ontología y a las palabras detectadas.

6.3.1. Comparativa ontología de etiquetado y ontologías comerciales

En primer lugar se van a comparar las categorías devueltas por cada herramienta comercial con las categorías creadas en nuestra ontología.

En la imagen 6.7 se puede observar como todos los aspectos se han clasificado correctamente en la categoría marcada durante el etiquetado. Esto se debe a que al utilizar nuestra propia ontología en Meaning Cloud, ver sección 6.1.1.1, los aspectos se han clasificado de la misma manera que en el proceso de etiquetado. Este hecho también implica que Meaning Cloud no ha detectado ningún aspecto que apareciera en el etiquetado y que no estuviera dentro de nuestra ontología.

En la figura 6.8 se observa como Google Cloud no utiliza ninguna de las categorías que hemos definido en nuestra ontología. No obstante, se pueden percibir relaciones entre las categorías propuestas y las categorías utilizadas por Google Cloud.

Por ejemplo, la categoría *PERSON* de Google Cloud agrupa las categorías *PROFESOR* y *ALUMNO* de nuestra ontología. Esta agrupación se puede observar en la figura 6.8, ya que, con la excepción de una muestra, todos los aspectos que Google Cloud ha clasificado

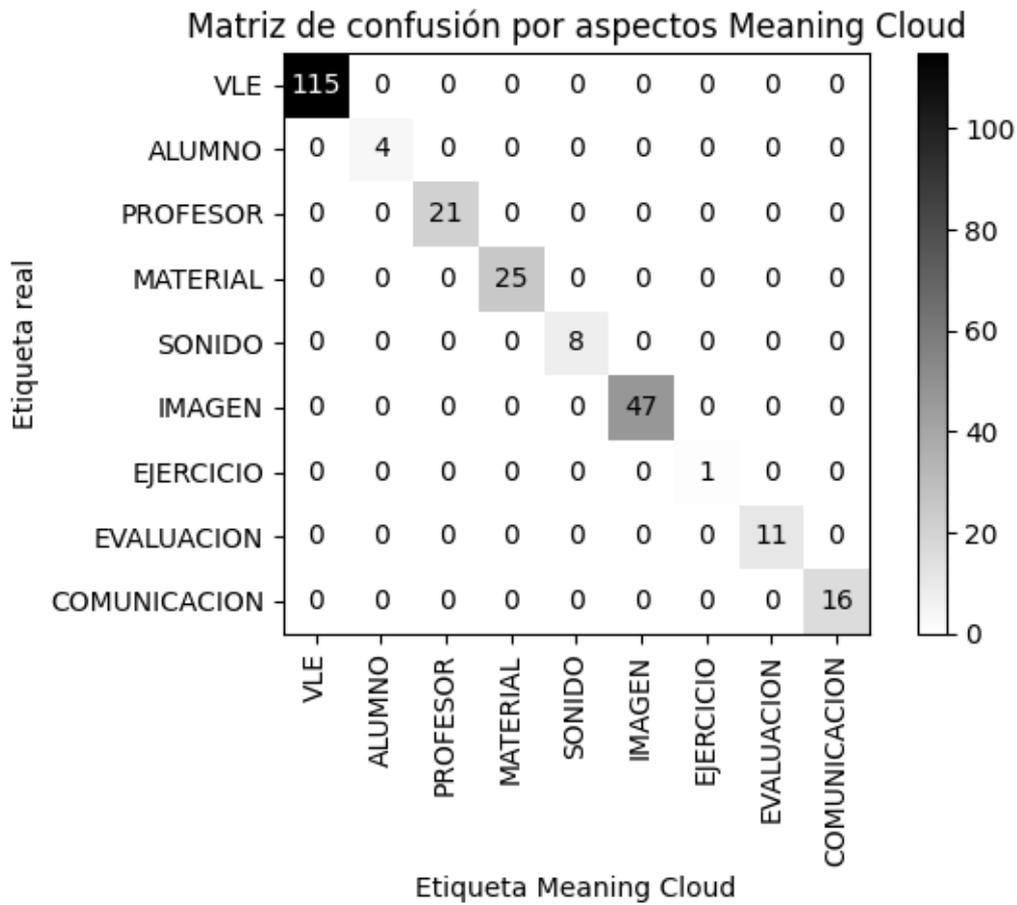


Figura 6.7: Matriz de confusión por aspectos Meaning Cloud.

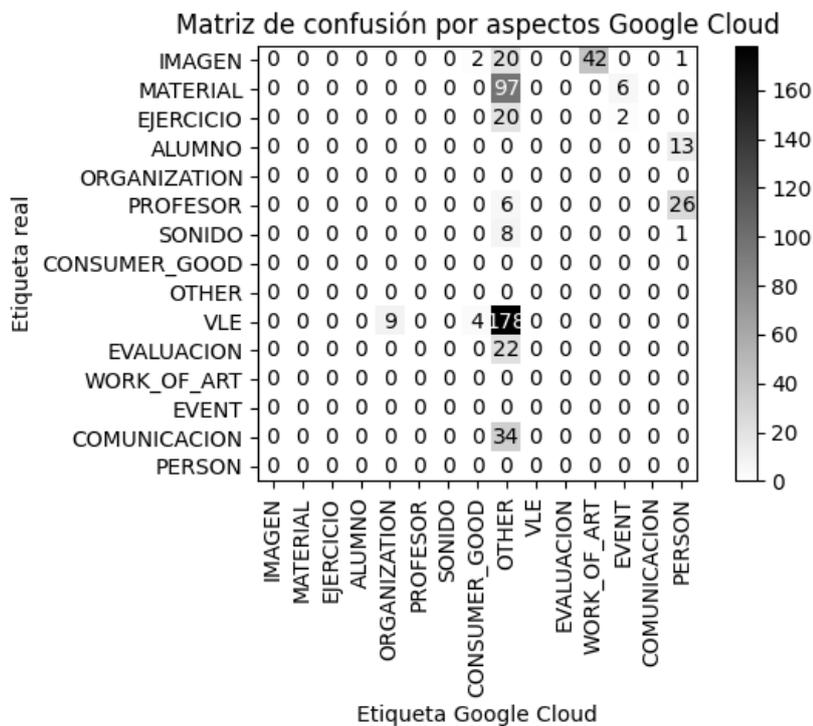


Figura 6.8: Matriz de confusión por aspectos Google Cloud.

en *PERSON* han sido etiquetados como parte de *PROFESOR* o *ALUMNO*. Además, la categoría *WORK_OF_ART* agrupa el 64% de los aspectos pertenecientes a la categoría *IMAGEN*, por lo que también se ve una relación entre estas dos categorías.

También es destacable que la mayor parte de aspectos Google Cloud los clasifica en la categoría *OTHER*, de manera que agrupa aspectos de todas las categorías.

6.3.2. Distribución por polaridad

Si bien hemos transformado la respuesta de las dos herramientas, Meaning Cloud y Google Cloud, así como los modelos *cross-domain*, para que utilicen la misma escala de polaridad, estas no tienen por qué valorar un aspecto de la misma manera.

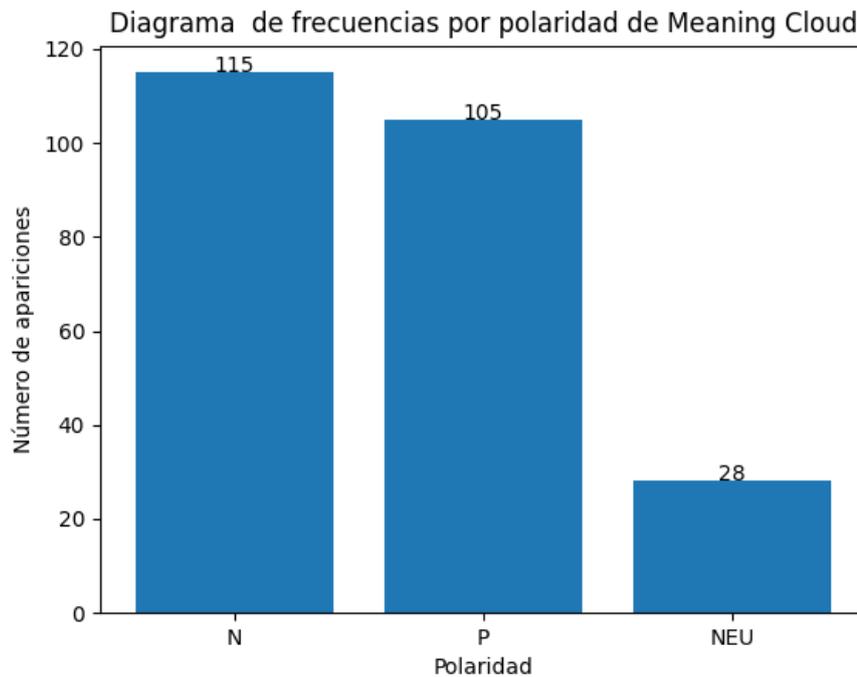


Figura 6.9: Frecuencia de aparición por polaridad en Meaning Cloud.

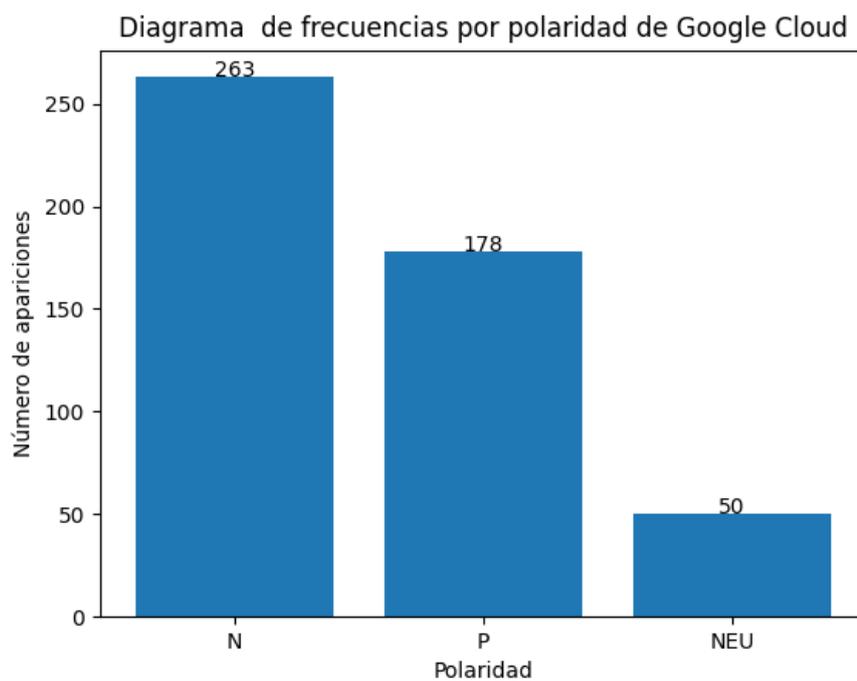


Figura 6.10: Frecuencia de aparición por polaridad en Google Cloud.

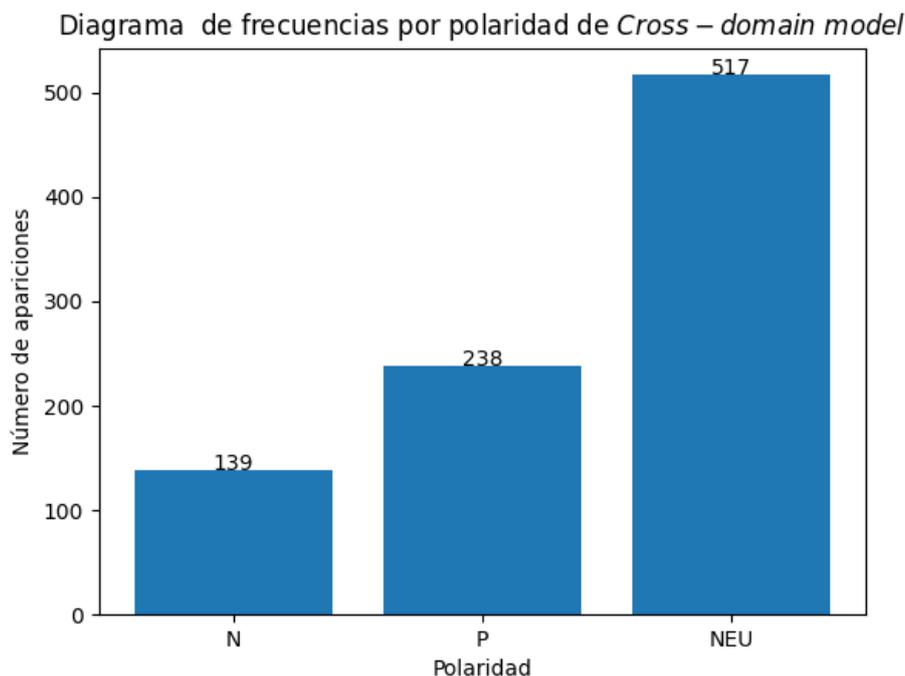


Figura 6.11: Frecuencia de aparición por polaridad en *cross-domain model*.

En las imágenes 6.11, 6.9 y 6.10 aparece el número de aspectos clasificados en cada clase por los tres modelos. Un punto a destacar es que el modelo de dominio cruzado detecta la totalidad aspectos, por lo que los valores absolutos no son comparables. Para facilitar la comparación se ha creado la tabla 6.6.

Herramienta	Negativo	Neutral	Positivo
Meaning Cloud	46,37 %	11,29 %	42,34 %
Google Cloud	53,56 %	10,19 %	36,25 %
<i>Cross-domain model</i>	15,55 %	57,83 %	26,62 %
Etiquetado correcto	39,37 %	9,06 %	51,57 %

Tabla 6.6: Comparativa frecuencia de aparición por polaridad.

En la tabla 6.6 se puede apreciar que el modelo de dominio cruzado tiende a clasificar muchos más aspectos como neutrales que los otros modelos generalistas de Meaning Cloud y Google Cloud. Meaning Cloud y Google Cloud clasifican aproximadamente el mismo porcentaje de aspectos como neutrales.

6.3.3. Distribución por categorías de herramientas comerciales

A continuación, se va a analizar la distribución por categorías de cada herramienta, en la figura 6.12 se analiza para Meaning Cloud y en 6.13 para Google Cloud.

En la figura 6.12 se puede observar como la categoría más frecuente es VLE, con 68 apariciones que la segunda categoría, IMAGEN, que ha aparecido un total de 47 veces y el resto de categorías tiene 25 o menos apariciones.

En la figura 6.13 se puede ver la distribución de aparición de entidades según su Categoría. Es destacable que la categoría con más apariciones es *OTHER* con 385 apariciones, seguida por *WORK_OF_ART* y *PERSON*, con 42 y 41 apariciones respectivamente. El

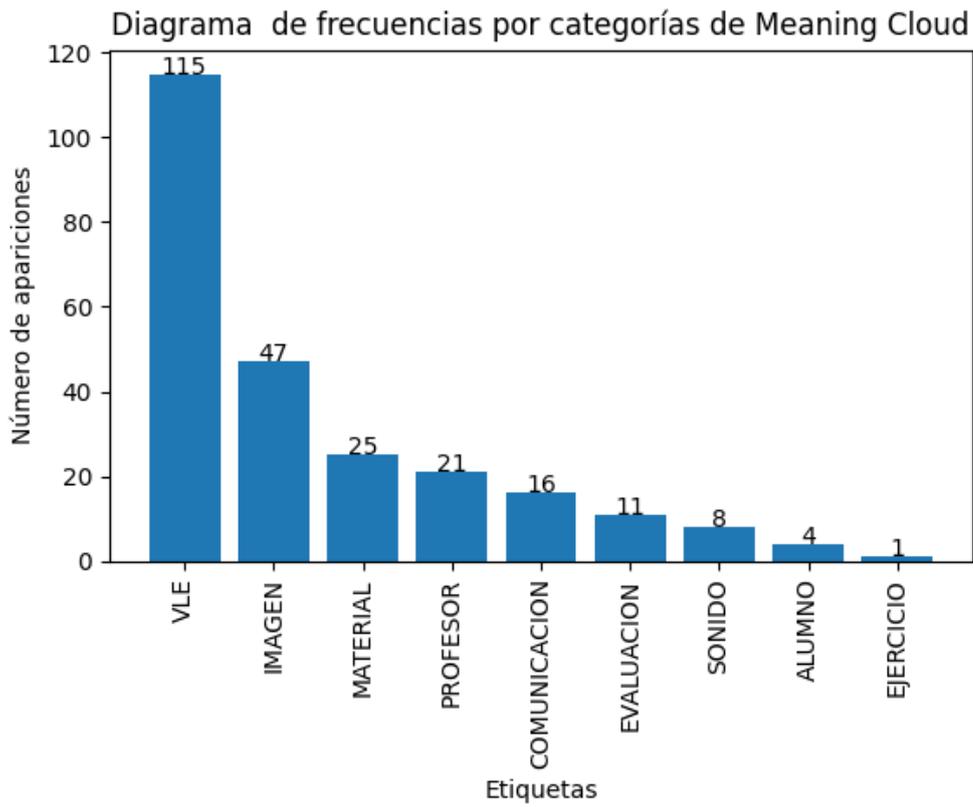


Figura 6.12: Frecuencia de aparición por categorías en Meaning Cloud.

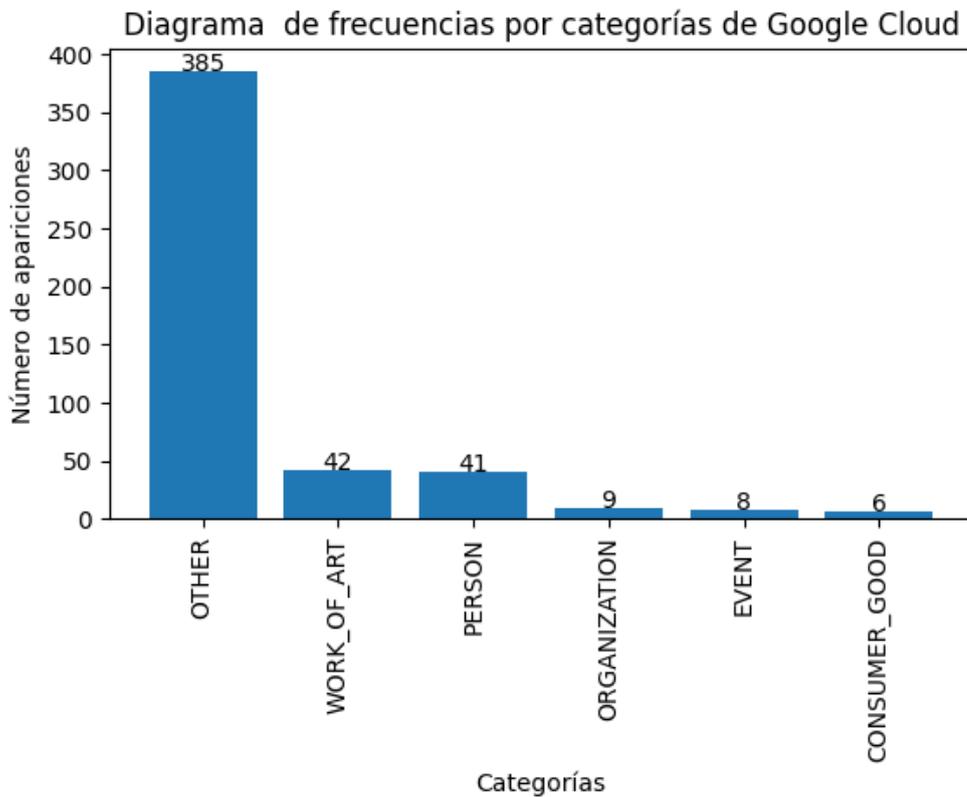


Figura 6.13: Frecuencia de aparición por categorías en Google Cloud.

resto de categorías no alcanza las diez apariciones. De esto se extrapola que la ontología utilizada por Google Cloud no es correcta para el dominio de las plataformas de aprendizaje virtual y los aspectos detectados no tienen hueco en sus diversas categorías, por lo que la mayor parte de los aspectos se derivan a la categoría *OTHER*.

6.3.4. Distribución por aspectos

A continuación, se presenta los aspectos con las diez frecuencias mayores en cada herramienta.

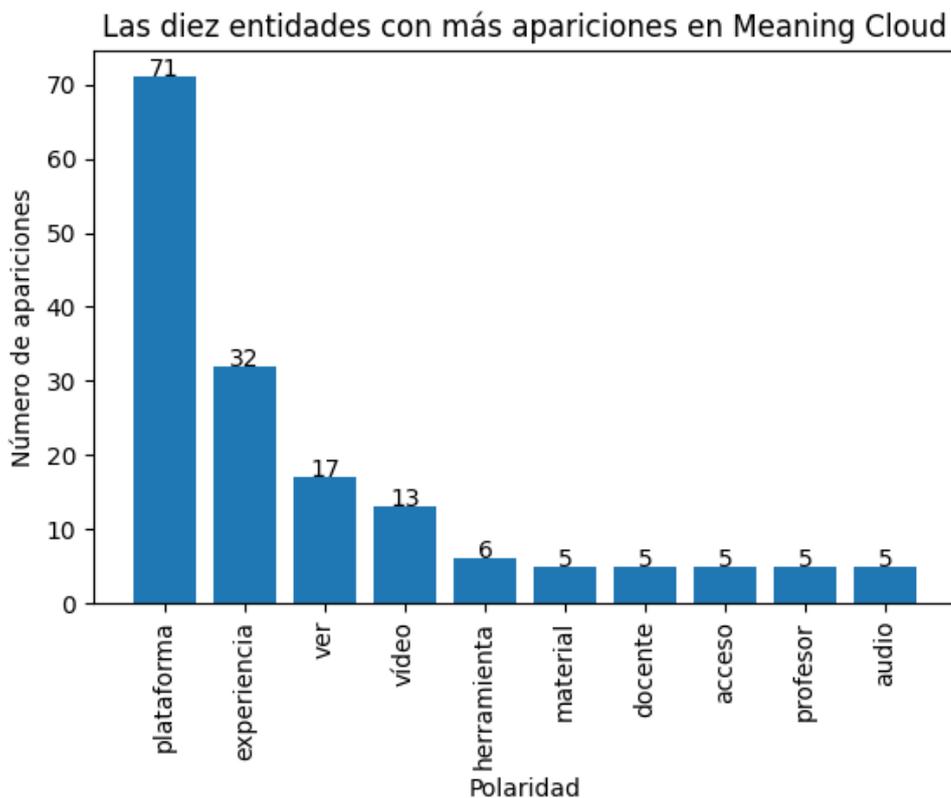


Figura 6.14: Frecuencia de aparición por entidad en Meaning Cloud.

En la imagen 6.14 se puede observar como plataforma es el aspecto más detectado seguido por experiencia, también destaca que a partir del cuarto aspecto la frecuencia de aparición comienza a ser inferior a 6.

En la imagen 6.15 el aspecto plataforma es el más frecuente, seguido por experiencia y curso, ambos con 32 apariciones. Es destacable el hecho de que vídeos y video aparezcan en la gráfica, puesto que son variaciones de un mismo lexema. Esto es debido a que la respuesta ofrecida por Google Cloud no devuelve los aspectos lematizados.

En la imagen 6.16 se puede apreciar la existencia de 288 aspectos omitidos, siendo este el tipo de aspectos más frecuente, seguido por el aspecto plataforma con 77 apariciones y curso con 71. Es destacable el hecho de que vídeos y video aparezcan en la gráfica, puesto que son variaciones de un mismo lexema. Esto es debido a que la respuesta ofrecida por el modelo de dominio cruzado no devuelve los aspectos lematizados.

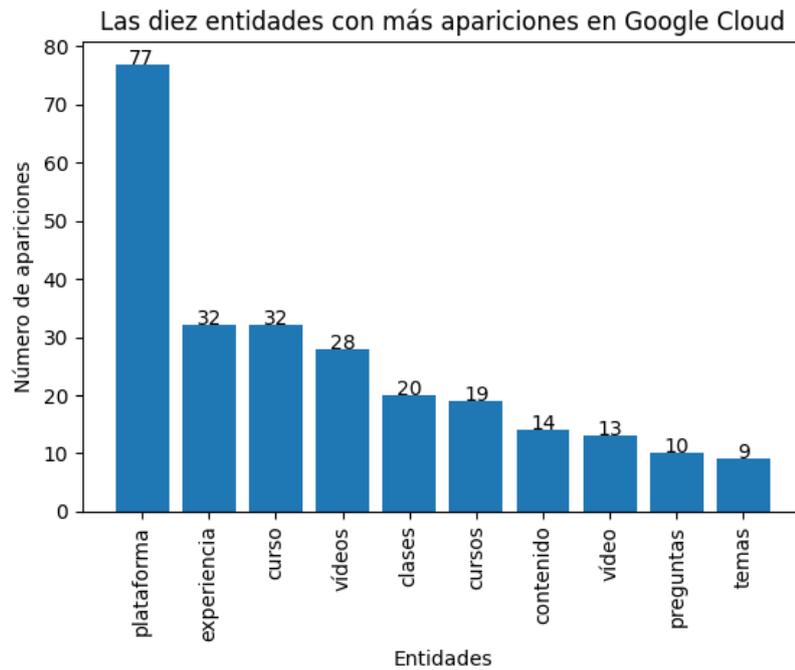


Figura 6.15: Frecuencia de aparición por entidad en Google Cloud.

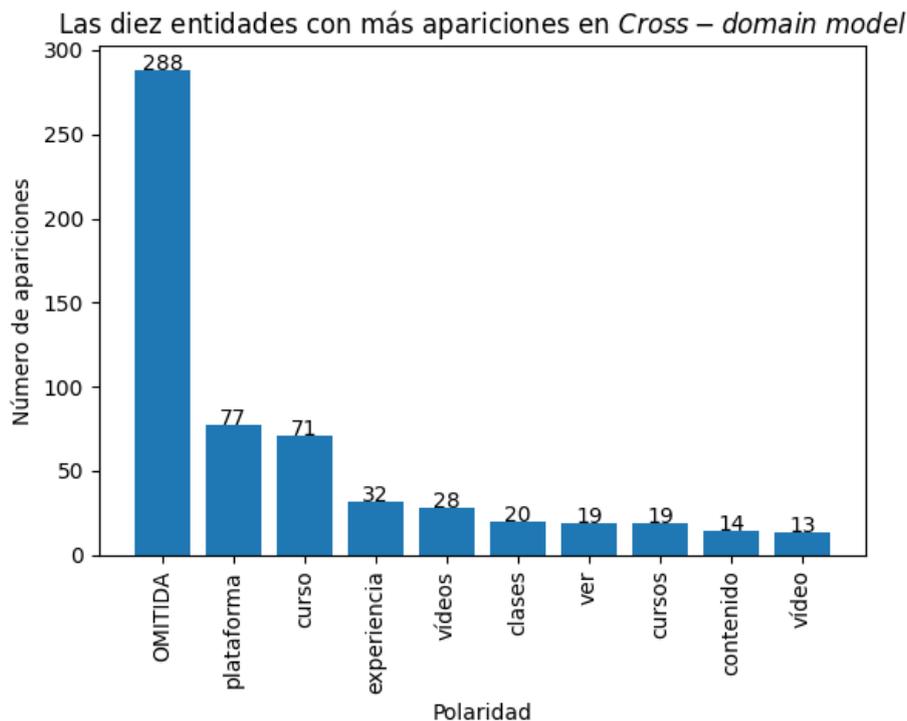


Figura 6.16: Frecuencia de aparición por entidad en *cross-domain model*.

6.3.5. Comparativa evaluación de herramientas

En este trabajo se han utilizado, evaluado y analizado dos herramientas comerciales de procesamiento de lenguaje natural y modelos propios entrenados por el equipo de investigación en otro dominio.

En general, los modelos Google Cloud han obtenido resultados más precisos, clasificando mejor para las clases positiva y neutral. Además, entre todas las muestras ne-

gativas del corpus detectadas marca más muestras como negativas que cualquier otra herramienta. Por otra parte, Meaning Cloud es más precisa detectando aspectos negativos. Los modelos de dominio cruzado tienen una sobretendencia a clasificar los aspectos como neutrales.

Herramienta	Accuracy	MF_1	Precision _P	Precision _{NEU}	Precision _N	Recall _P	Recall _{NEU}	Recall _N	F1 _P	F1 _{NEU}	F1 _N
Meaning Cloud	0,548	0,494	0,788	0,172	0,833	0,848	0,607	0,261	0,817	0,268	0,397
Google Cloud	0,815	0,616	0,901	0,235	0,799	0,820	0,080	0,951	0,861	0,119	0,868
cross-domain model	0,424	0,433	0,908	0,112	0,813	0,469	0,716	0,321	0,618	0,194	0,460

Tabla 6.7: Comparativa evaluación herramientas comerciales.

En primer lugar, vamos a comparar los resultados de la evaluación de las herramientas. Google Cloud ha obtenido resultados más precisos, clasificando mejor para las clases positiva y neutral. Además, entre todas las muestras negativas del corpus detecta marca más muestras como negativas que cualquier otra herramienta. Por otra parte, Meaning Cloud es más precisa detectando aspectos negativos. El modelo de dominio cruzado utilizado es la herramienta que recupera más muestras neutrales.

A nivel global Google Cloud es la herramienta más precisa y además obtiene un mayor valor de MF_1 , el cual representa un balance entre la precisión de la clasificación y la cantidad de etiquetas de cada clase que se etiquetan con esa misma clase.

Todas herramientas tienen dificultades para clasificar muestras de la clase neutral, esto se debe a que es la clase con un menor número de muestras.

CAPÍTULO 7

Experiencia de usuario

En este capítulo en primer lugar presentaremos diversas herramientas de medición para plataformas interactivas digitales, prestando especial atención en aquellas centradas en entornos de aprendizaje virtual. A continuación se realizará un estudio de la experiencia de usuario basado en los resultados obtenidos mediante técnicas de procesamiento del lenguaje natural y análisis de sentimiento a nivel aspectual en la sección 6.

7.1 Introducción

La experiencia de usuario es un concepto multidimensional y centrado en las necesidades humanas, por eso las interacciones persona-computador (conocidas generalmente por su acrónimo en inglés HCI *Human Computer Interaction*) deben ser consideradas cuando se diseña y evalúa cualquier sistema interactivo [28]. Pese a esto la evaluación de la experiencia de usuario en todo tipo de sistema sigue, especialmente en los sistemas de aprendizaje virtual, continúa siendo un tema que está en un estado incipiente de la investigación [33]. Este hecho junto con los prometedores resultados del procesamiento del lenguaje natural plantea la posibilidad de usar herramientas de procesamiento del lenguaje natural para evaluar dichos sistemas de manera automática y en concreto para medir la experiencia de usuario.

7.2 Herramientas de medición para plataformas interactivas

Actualmente la normativa ISO 9241-210:2019 [28] marca el estándar de UX en plataformas interactivas digitales, no obstante, esta normativa es generalista, de manera que se van a analizar diversas propuestas de herramientas de medición para plataformas de aprendizaje virtual. Las propuestas destacadas son: el cuestionario UEQ [42], las cuatro dimensiones propuestas por Panagiotis Zaharias y Christopher Pappas en [25] y los requisitos presentados por Hana Ovesleová en [53].

7.2.1. User Experience Questionnaire

El cuestionario UEQ (User Experience Questionnaire) es un cuestionario cuyo objetivo es medir la experiencia de usuario. Consta de 26 elementos a valorar agrupados en un total de seis categorías, en la figura 7.1 se puede ver el conjunto de elementos y en qué categoría se clasifican. Las categorías se pueden clasificar a su vez en tres dimensiones: pragmática, que incluye claridad, eficiencia y fiabilidad; hedonista, que incluye motiva-

ción y novedad y una última dimensión que engloba las impresiones generales, la cual está formada únicamente por la categoría atracción [54].

Si bien este cuestionario es de carácter generalista existen adaptaciones para aplicarse a entornos de aprendizaje virtual, como la presentada por R. Sanchis et al. en [27] y utilizada para conseguir el corpus de datos de este trabajo.

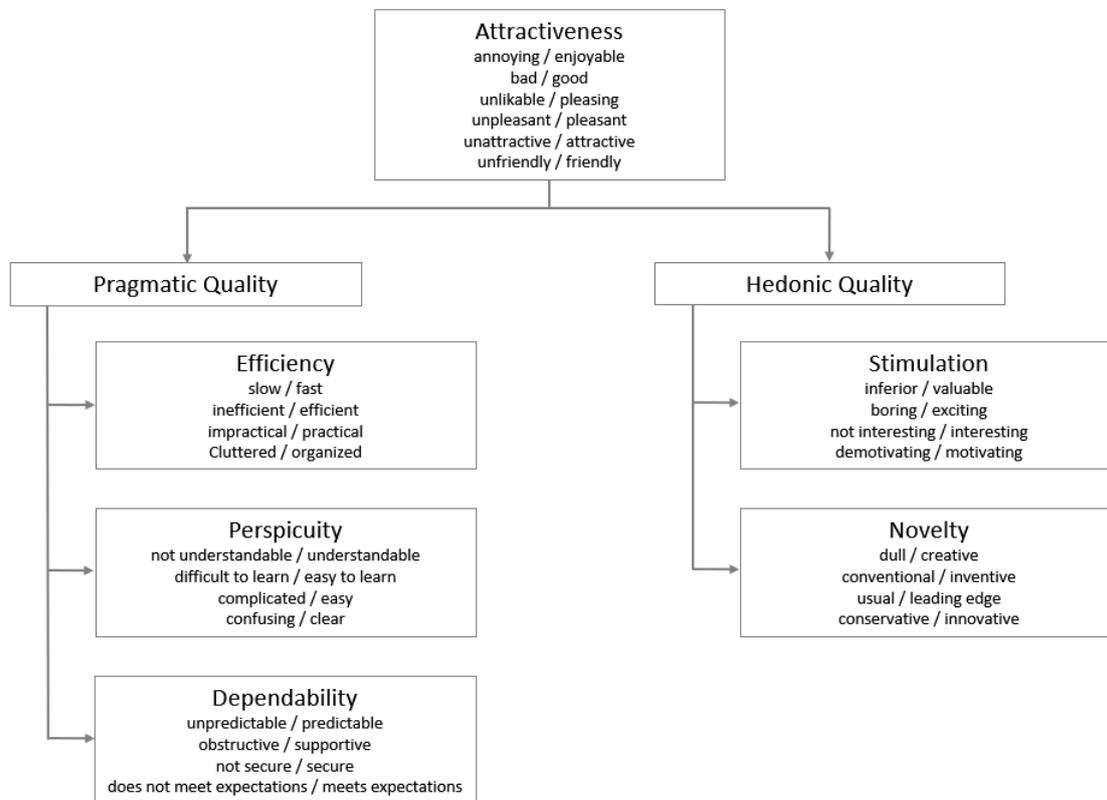


Figura 7.1: Estructura general del UEQ. [54]

Concretamente cada categoría mide:

- **Claridad:** la facilidad de aprender a usar el producto.
- **Eficiencia:** la capacidad usar el producto sin realizar esfuerzos innecesarios.
- **Fiabilidad:** la medida en la que el usuario se siente en control de la situación.
- **Motivación:** lo estimulante que es usar el producto.
- **Novedad:** lo innovativo e interesante que es el producto.
- **Atracción:** la impresión general del producto.

7.2.2. Modelo de Zaharias y Pappas

Panagiotis Zaharias y Christopher Pappas proponen un modelo de evaluación aplicable solamente a sistemas de aprendizaje virtual basado en las cuatro dimensiones que se pueden ver en la figura 7.2 [25].

Los conceptos comprendidos en cada categoría son los siguientes:

- **Pragmática:** agrupa los conceptos relacionados con las metas del aprendizaje.

- **Aprendizaje real:** agrupa todos los tipos de actividades que permiten un aprendizaje activo.
- **Autonomía y relaciones:** aglutina las relaciones sociales con miembros de la comunidad educativa y los retos que tiene que afrontar el estudiante para usar la plataforma de una manera eficiente.
- **Motivación y compromiso:** agrupa todos aquellos aspectos que ayudan a promover la finalización de los estudios.

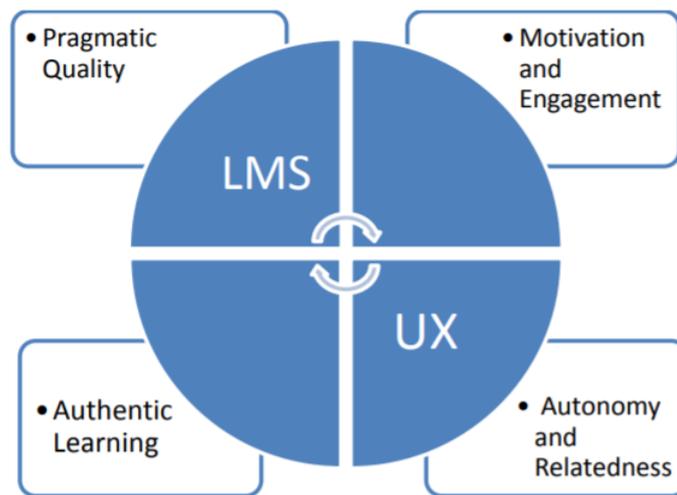


Figura 7.2: Dimensiones para evaluar un sistema de aprendizaje virtual. [25]

7.2.3. Requisitos de Ovesleová

Hana Ovesleová realiza una aproximación basada en las emociones, los aspectos emocionales y las herramientas metodológicas que pueden trabajar con sentimientos. Esto se debe a que la condición emocional del alumnado afecta en gran medida al proceso de aprendizaje [53]. Para reforzar esta idea defiende que el proceso de aprendizaje y la metodología empleada en este es un punto de partida para el diseño del sistema de aprendizaje virtual, el cual se debe ver influenciado tanto por la metodología de aprendizaje aplicada durante el curso como por los materiales disponibles.

Los nueve requisitos propuestos son:

- **Captar la atención:** la plataforma debe facilitar mantener la atención mediante patrones de diseño derivados de los modelos cognitivos del alumnado.
- **Informar de los objetivos:** los alumnos deben ser informados de los objetivos y todas las herramientas a su disposición deben ser únicas respecto a funcionalidad y uso.
- **Estimular el recuerdo de conocimientos previos:** se debe facilitar la relación entre el conocimiento previo y el conocimiento nuevo.
- **Presentación de estímulos:** la información se debe proporcionar de forma secuencial para que sea estimulante.
- **Enseñanza guiada:** cuando se presentan conocimientos nuevos al alumno se debe pedir retroalimentación a este para poder apoyarle.

- **Mejorar el rendimiento:** se debe aplicar todo el conocimiento adquirido para mejorar el rendimiento del aprendizaje.
- **Retroalimentar al alumnado:** se debe informar al alumnado de si su avance es correcto de forma periódica.
- **Evaluar el desempeño del alumnado:** se debe medir el desempeño del alumnado en función de los objetivos fijados previamente.
- **Mejorar la retención y transferencia de conocimiento:** se debe estimular la aplicación del conocimiento adquirido en entornos externos a la plataforma de aprendizaje.

7.3 Relación entre dimensiones de la UX y nuestra ontología

De los trabajos presentados en la sección anterior se puede concluir que todos los autores coinciden en la gran importancia que tienen las relaciones sociales y el material disponible en el aprendizaje online. Por este motivo para evaluar de manera automática los sistemas de aprendizaje virtual se propone la clasificación y relación de las categorías presentes en nuestra ontología 5.3 en dos dimensiones:

- **Conexiones sociales:** agrupa todo el conjunto de herramientas que permiten la comunicación y colaboración interpersonal en los sistemas de aprendizaje virtual. En esta dimensión entran las siguientes categorías de nuestra ontología:
 - Alumno.
 - Comunicación.
 - Profesor.
- **Material de aprendizaje:** aglutina todo el conjunto de materiales empleados en el proceso de aprendizaje, desde exámenes o ejercicios hasta vídeos y su calidad gráfica y de sonido. En esta dimensión entran las siguientes categorías de nuestra ontología:
 - Ejercicio.
 - Evaluación.
 - Imagen.
 - Material.
 - Sonido.

En la tabla 7.1 se puede ver como se relaciona nuestra ontología respecto a dos de las tres herramientas explicadas en el apartado 7.2 y las dos dimensiones propuestas en esta misma sección. Solamente se utilizan en la comparación el modelo de Zaharias y Pappas, 7.2.2, y los requisitos de Ovesleová, 7.2.3, porque son las dos únicas herramientas presentadas centradas en entornos de aprendizaje interactivo. Si bien el UEQ ha sido utilizado para evaluar al usuario de aprendizaje virtual no es un cuestionario especializado.

Categoría ontológica	Modelo Zaharias-Pappas	Requisitos Ovesleová	Dimensiones propias
VLE	Pragmática	Captar la atención	-
Alumno	Autonomía y relaciones	Enseñanza guiada	Conexiones sociales
Comunicación	Autonomía y relaciones	Enseñanza guiada	Conexiones sociales
Ejercicio	Aprendizaje real	Mejorar la retención y transferencia del conocimiento	Material de aprendizaje
Evaluación	Pragmática	Retroalimentar al alumnado	Material de aprendizaje
Imagen	Aprendizaje real	Presentación de estímulos	Material de aprendizaje
Material	Aprendizaje real	Presentación de estímulos	Material de aprendizaje
Profesor	Autonomía y relaciones	Retroalimentar al alumnado	Conexiones sociales
Sonido	Aprendizaje real	Presentación de estímulos	Material de aprendizaje

Tabla 7.1: Relación entre diversas herramientas de medición UX y las categorías de ontología para el análisis de sentimiento en plataformas de aprendizaje online.

7.4 Estadísticas y análisis de la experiencia de usuario

Para realizar el análisis de la experiencia de usuario a nivel dimensional se van a utilizar los datos obtenidos de la herramienta Meaning Cloud utilizados en la fase de experimentación. En la sección 6.1.4 se pueden encontrar más detalles sobre el conjunto de datos que se va a utilizar. Esta decisión se debe a que es el conjunto de datos obtenidos mediante una herramienta de análisis de sentimientos cuya clasificación en categorías se ajusta en mayor medida a nuestra ontología 6.1.1 .

El primer paso realizado ha sido reorganizar los aspectos que forman nuestro conjunto de datos en las dos dimensiones a analizar. Es importante destacar que al no utilizar la categoría VLE el conjunto de datos a tratar es menor que el conjunto de aspectos detectados por Meaning Cloud. La distribución de los aspectos a analizar y la cantidad de aspectos que forman cada dimensión se pueden observar en la imagen 7.3. En esa misma imagen se puede percibir que la dimensión “Conexiones sociales” presenta un menor número de aspectos que la dimensión “Material de aprendizaje”, esto se debe por a la mayor cantidad de categorías incluidas en esta última dimensión y la inclusión en esta dimensión de dos de las tres categorías con mayor frecuencia de aparición. Esto último se puede visualizar en la figura 6.12.

Distribución de dimensiones.

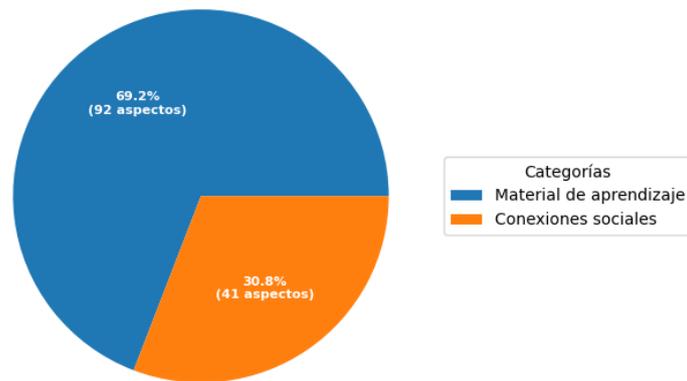


Figura 7.3: Distribución relativa de dimensiones.

En la figura 7.4 podemos observar en qué medida los comentarios de cada categoría de la dimensión y de la propia dimensión son positivos o negativos. La calificación más abundante tanto de la dimensión completa como sus distintas categorías es neutral.

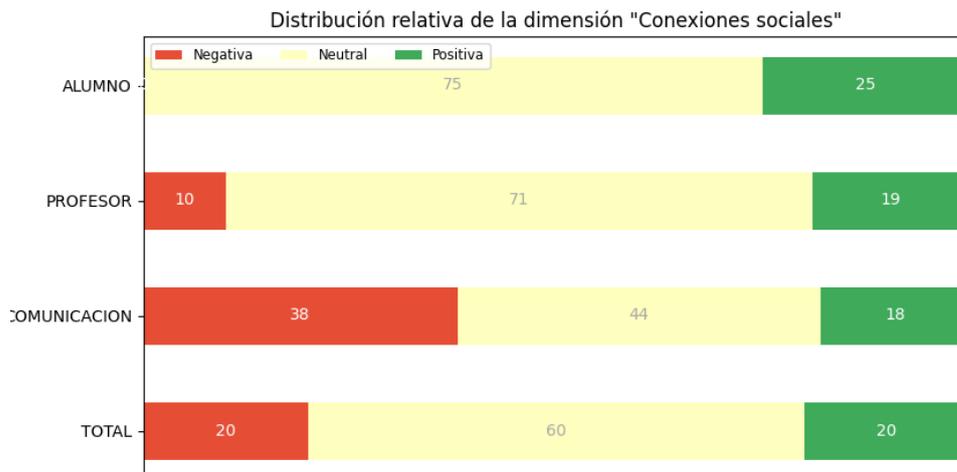


Figura 7.4: Distribución relativa de las categorías de la dimensión "Conexiones sociales".

En la figura 7.5 podemos apreciar que si bien la calificación más frecuente y la de la dimensión completa es neutral el aspecto SONIDO tiene una clasificación claramente negativa con el 76 % de los aspectos clasificados en la polaridad negativa.

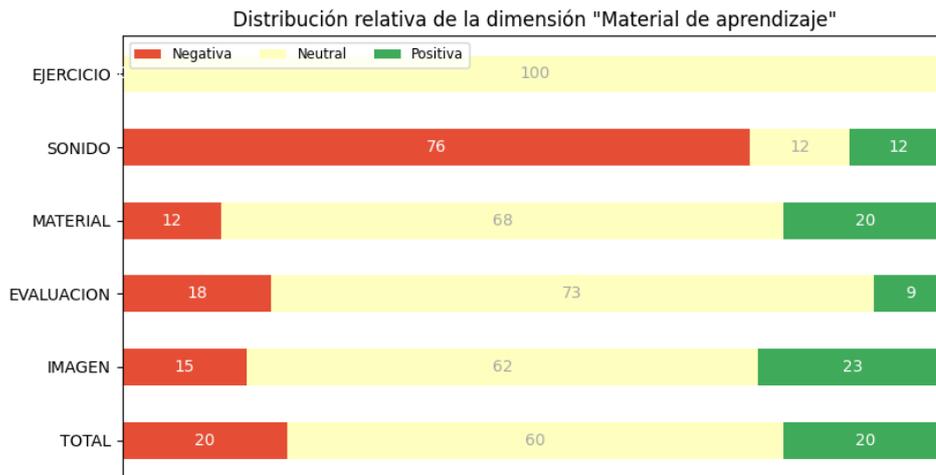


Figura 7.5: Distribución relativa de las categorías de la dimensión "Material de aprendizaje".

7.5 Conclusiones del análisis de la UX en entornos de aprendizaje virtual con herramientas de análisis de sentimiento para el NLP

Es importante recordar que los datos utilizados provienen de diversas plataformas de aprendizaje virtual, ver apartado 5.2, por lo que el análisis de las estadísticas calculadas en el apartado 7.4 no puede aspirar a dar un análisis que trate el origen de datos como si este proviniera de una única fuente.

En las figuras 7.4 y 7.5 se puede apreciar que la valoración general de ambas dimensiones es neutral, por lo que las plataformas de las que proveen los datos no ofrecen la posibilidad de relacionarse satisfactoriamente en el entorno académico virtual ni proveen de un material que cumpla con las expectativas del alumnado.

Prestando especial atención a la dimensión "Conexiones sociales" se puede apreciar que ALUMNO es la única categoría que no tiene ninguna valoración negativa y que COMUNICACIÓN recibe el mayor porcentaje de críticas, pese a esto existe un equilibrio entre la proporción de aspectos positivos y negativos.

Analizando la dimensión "Material de aprendizaje" volvemos a obtener un equilibrio en el porcentaje de aspectos con valoración positiva y con valoración negativa. Destaca la categoría ejercicio que únicamente recibe opiniones neutrales y la categoría sonido, que recibe un 76 % de críticas negativas, siendo la categoría que sufre la peor calificación de las dos dimensiones analizadas.

La predominancia de valores neutrales en la evaluación de las dos dimensiones presentadas es coherente con los resultados obtenidos en la figura 6.9. Sin embargo en los valores obtenidos de la experimentación se observan más valoraciones que positivas, al contrario que en las dimensiones. Este hecho es explicable porque en el análisis mediante las dos dimensiones no se ha utilizado la categoría VLE.

El análisis de la experiencia de usuario mediante un análisis de sentimientos a nivel aspectual nos ha permitido analizar dos importantes dimensiones de la experiencia de usuario y desglosar las dimensiones en categorías más específicas que dan información concreta de qué características de la experiencia de usuario hay que mejorar.

Conclusiones y trabajos futuros

Para concluir el trabajo se presentarán las conclusiones obtenidas de este, se evaluará la consecución de objetivos de acuerdo con la rúbrica 1.1 y se propondrá una serie de trabajos futuros que permitan seguir ahondando en el tema tratado.

8.1 Conclusiones

En este trabajo se han analizado y comparado las herramientas comerciales de procesamiento del lenguaje natural más importantes del mercado prestando especial atención a sus características aplicables al análisis de sentimiento a nivel de aspecto.

El siguiente paso ha sido investigar las diversas plataformas de etiquetado existentes con la finalidad de encontrar una que facilitara el proceso de anotación. Una vez elegidas las plataformas a comparar se han investigado en mayor medida, se han comparado y se ha elegido Doccano, ya que era la herramienta que cumplía mejor los requisitos. Una vez elegida la plataforma se ha creado una guía de utilización para facilitar el trabajo a los etiquetadores.

A continuación, se ha descrito y procesado el corpus de datos para la experimentación con la finalidad de extraer toda la información posible de este. Utilizando el conocimiento previo existente sobre el dominio y la información extraída se ha procedido a crear una ontología propia que permitiera clasificar los aspectos presentes en el texto y se ha procedido a realizar el etiquetado del corpus de datos entre tres etiquetadores y la evaluación del resultado de este proceso.

Después se ha realizado la experimentación con modelos generalistas de dos de las tres herramientas analizadas previamente, concretamente con Google Cloud y Meaning Cloud. Este proceso se ha descrito paso a paso comenzando por la creación de nuestra ontología en las herramientas, cuando era posible, y terminando con la evaluación de las respuestas de las plataformas. Una vez terminada la evaluación de las herramientas, se ha procedido a analizar los resultados obtenidos. La experimentación finaliza probando modelos propios, de dominio cruzado, entrenados por el grupo de investigación con Tweets.

El último paso ha consistido en llevar a cabo un breve análisis de la experiencia de usuario mediante los resultados obtenidos durante la experimentación. Las fases en las que se ha dividido este proceso son las siguientes: presentación de diversas herramientas de medición para plataformas digitales, presentación de dos dimensiones de la experiencia de usuario y su relación con nuestra ontología, análisis de la experiencia de usuario mediante estas dos dimensiones y finalmente las conclusiones obtenidas del análisis.

8.2 Evaluación de objetivos

De acuerdo con la rúbrica 1.1, todos los objetivos del trabajo se han alcanzado tal y como muestra la tabla 8.1

Objetivos	Objetivo no cumplido	Objetivo trabajado insuficientemente	Objetivo alcanzado suficientemente	Objetivo cumplido
Investigar y elegir de plataforma de etiquetado				✓
Investigar, analizar y utilizar diversas plataformas de NLP				✓
Etiquetar el conjunto de datos				✓
Realizar un análisis de la experiencia de usuario				✓

Tabla 8.1: Evaluación de objetivos.

8.3 Trabajos futuros

En este trabajo nos hemos centrado en explorar las posibilidades que ofrecen las herramientas comerciales que permiten realizar análisis de sentimiento a nivel aspectual para analizar la experiencia de usuario en el dominio del aprendizaje virtual. De cara a trabajos que profundicen en el tema abordado se proponen las siguientes líneas de investigación:

- **Creación de un modelo del dominio:** en el caso de que se consiguieran más datos para el corpus de experimentación se podría entrenar un modelo de clasificación para el dominio del aprendizaje virtual y se podría comparar con los resultados obtenidos en este trabajo.
- **Utilización de un modelo que recopile y devuelva el contexto:** en este trabajo se ha analizado la polaridad de los aspectos, pero no se ha recuperado el contexto que acompaña a los aspectos en cada frase, podría ser interesante utilizar un modelo que lo recuperara para poder extraer más información para evaluar la experiencia de usuario.
- **Analizar una plataforma de aprendizaje virtual:** una posible línea de investigación para continuar este trabajo es analizar una plataforma de aprendizaje virtual mediante el método presentado en el capítulo 7 y mediante otro método para comparar los análisis resultantes.

Bibliografía

- [1] Meltem Huri Baturay. An overview of the world of moocs. *Procedia-Social and Behavioral Sciences*, 174:427–433, 2015.
- [2] Anuradha Akuratiya and Nadeeshika Meddage. Students' perception of online learning during covid-19 pandemic: A survey study of it students. IV:755–758, 09 2020.
- [3] Tony Mullen and Nigel Collier. Sentiment analysis using support vector machines with diverse information sources. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pages 412–418, Barcelona, Spain, July 2004. Association for Computational Linguistics.
- [4] Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, HLT '05, page 347–354, USA, 2005. Association for Computational Linguistics.
- [5] Tetsuya Nasukawa and Jeonghee Yi. Sentiment analysis: Capturing favorability using natural language processing. In *Proceedings of the 2nd International Conference on Knowledge Capture*, K-CAP '03, page 70–77, New York, NY, USA, 2003. Association for Computing Machinery.
- [6] Rudy Prabowo and Mike Thelwall. Sentiment analysis: A combined approach. *Journal of Informetrics*, 3(2):143–157, 2009.
- [7] Chenghua Lin and Yulan He. Joint sentiment/topic model for sentiment analysis. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, CIKM '09, page 375–384, New York, NY, USA, 2009. Association for Computing Machinery.
- [8] Casey Whitelaw, Navendu Garg, and Shlomo Argamon. Using appraisal groups for sentiment analysis. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, CIKM '05, page 625–631, New York, NY, USA, 2005. Association for Computing Machinery.
- [9] Julia Heidemann, Mathias Klier, and Florian Probst. Online social networks: A survey of a global phenomenon. *Computer Networks*, 56(18):3866–3878, 2012. The WEB we live in.
- [10] Pollyanna Gonçalves, Matheus Araújo, Fabrício Benevenuto, and Meeyoung Cha. Comparing and combining sentiment analysis methods. In *Proceedings of the First ACM Conference on Online Social Networks*, COSN '13, page 27–38, New York, NY, USA, 2013. Association for Computing Machinery.
- [11] M.D. Devika, C. Sunitha, and Amal Ganesh. Sentiment analysis: A comparative study on different approaches. *Procedia Computer Science*, 87:44–49, 2016. Fourth

- International Conference on Recent Trends in Computer Science and Engineering (ICRTCSE 2016).
- [12] Wei Li, Luyao Zhu, Yong Shi, Kun Guo, and Erik Cambria. User reviews: Sentiment analysis using lexicon integrated two-channel cnn-lstm family models. *Applied Soft Computing*, 94:106435, 2020.
- [13] Manuel Carlos Díaz-Galiano, M. Vega, E. Casasola, Luis Chiruzzo, Miguel Ángel García Cumbreiras, Eugenio Martínez-Cámara, D. Moctezuma, Arturo Montejó Ráez, Marco Antonio Sobrevilla Cabezudo, Eric Sadit Tellez, Mario Graff, and Sabino Miranda-Jiménez. Overview of tass 2019: One more further for the global spanish sentiment analysis corpus. In *IberLEF@SEPLN*, 2019.
- [14] Google Cloud. Google cloud. <https://cloud.google.com/>, 2021.
- [15] IBM. Natural Language Understanding. <https://www.ibm.com/watson/services/natural-language-understanding/>, 2019.
- [16] Microsoft Azure. Microsoft azure. <https://azure.microsoft.com/es-es/services/cognitive-services/text-analytics/>, 2021.
- [17] Meaning Cloud LLC. Meaning cloud. <https://www.meaningcloud.com/es>, 2021.
- [18] Carlos Henriquez, Ferran Pla, Lluís-F Hurtado, and Jaime Guzman. Aspect-based sentiment analysis using ontologies and machine learning. *PROCESAMIENTO DEL LENGUAJE NATURAL*, (59):49–56, 2017.
- [19] Tareq Al-Moslmi, Marc Gallofré Ocaña, Andreas L. Opdahl, and Csaba Veres. Named entity extraction for knowledge graphs: A literature overview. *IEEE Access*, 8:32862–32881, 2020.
- [20] Mohammad Erfan Mowlaei, Mohammad Saniee Abadeh, and Hamidreza Keshtarvarz. Aspect-based sentiment analysis using adaptive aspect-based lexicons. *Expert Systems with Applications*, 148:113234, 2020.
- [21] Ganpat Singh Chauhan, Yogesh Kumar Meena, Dinesh Gopalani, and Ravi Nahta. A two-step hybrid unsupervised model with attention mechanism for aspect extraction. *Expert Systems with Applications*, 161:113673, 2020.
- [22] Weixin Zeng, Jiuyang Tang, Xiang Zhao, Bin Ge, and Weidong Xiao. Named entity disambiguation via probabilistic graphical model with embedding features. In Long Cheng, Andrew Chi Sing Leung, and Seiichi Ozawa, editors, *Neural Information Processing*, pages 16–27, Cham, 2018. Springer International Publishing.
- [23] Efrén Rama-Maneiro, Juan C. Vidal, and Manuel Lama. Collective disambiguation in entity linking based on topic coherence in semantic graphs. *Knowledge-Based Systems*, 199:105967, 2020.
- [24] Yotam Eshel, Noam Cohen, Kira Radinsky, Shaul Markovitch, Ikuya Yamada, and Omer Levy. Named entity disambiguation for noisy text. pages 58–68, 01 2017.
- [25] Panagiotis Zaharias and Christopher Pappas. Quality management of learning management systems: A user experience perspective. *Current Issues in Emerging eLearning*, 3(1):5, 2016.
- [26] Ergonomics of human-system interaction — part 210: Human-centred design for interactive systems. Standard, 2019.

- [27] Rosario Sanchis Font, María José Castro Bleda, Carlos Turró Ribalta, and Ignacio Despujol Zabala. Integración del “user experience questionnaire short” en moocs upv. In *IN-RED 2018. IV Congreso Nacional de Innovación Educativa y Docencia en Red*, pages 393–404. Editorial Universitat Politècnica de València, 2018.
- [28] Ergonomics of human-system interaction — part 210: Human-centred design for interactive systems. Standard, 2019.
- [29] Nigel Bevana, Jurek Kirakowskib, and Jonathan Maissela. What is usability. In *Proceedings of the 4th International Conference on HCI*. Citeseer, 1991.
- [30] Iván Balmaceda Castro, Cristian Rusu, and Silvana Aciar. Customer experience in e-learning: A systematic mapping study. In Gabriele Meiselwitz, editor, *Social Computing and Social Media. Participation, User Experience, Consumer Experience, and Applications of Social Computing*, pages 158–170, Cham, 2020. Springer International Publishing.
- [31] Estelle de Kock, Judy van Biljon, and Adèle Botha. User experience of academic staff in the use of a learning management system tool. In *Proceedings of the Annual Conference of the South African Institute of Computer Scientists and Information Technologists*, volume 26-28-, page 1–10. ACM, 2016.
- [32] Maria Rauschenberger, Martin Schrepp, Manuel Perez Cota, Siegfried Olschner, and Jörg Thomaschewski. Efficient Measurement of the User Experience of Interactive Products. How to use the User Experience Questionnaire (UEQ). Example: Spanish Language Version. *International Journal of Interactive Multimedia and Artificial Intelligence*, 2(1):39–45, 03/2013 2013.
- [33] Rosario Sanchis-Font, Maria Jose Castro-Bleda, and José-Ángel González. Applying sentiment analysis with cross-domain models to evaluate user experience in virtual learning environments. In Ignacio Rojas, Gonzalo Joya, and Andreu Catala, editors, *Advances in Computational Intelligence*, pages 609–620, Cham, 2019. Springer International Publishing.
- [34] Nicolás Díaz Roussel. *Estudio comparativo de herramientas para tareas de Procesamiento de Lenguaje Natural*. PhD thesis, 2020.
- [35] Google Cloud Natural Language. Google cloud. <https://cloud.google.com/natural-language>, 2021.
- [36] Mariana Neves and Jurica Ševa. An extensive review of tools for manual annotation of documents. *Briefings in Bioinformatics*, 22(1):146–163, 12 2019.
- [37] Fabian Gringel. The best free labeling tools for text annotation in nlp. <https://dida.do/blog/the-best-free-labeling-tools-for-text-annotation-in-nlp>, 2020.
- [38] Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun’ichi Tsujii. brat: a web-based tool for NLP-assisted text annotation. In *Proceedings of the Demonstrations Session at EACL 2012*, Avignon, France, April 2012. Association for Computational Linguistics.
- [39] Hiroki Nakayama, Takahiro Kubo, Junya Kamura, Yasufumi Taniguchi, and Xu Liang. doccano: Text annotation tool for human, 2018. Software available from <https://github.com/doccano/doccano>.

- [40] Jan-Christoph Klie, Michael Bugert, Beto Boullosa, Richard Eckart de Castilho, and Iryna Gurevych. The INCEpTION platform: Machine-assisted and knowledge-oriented interactive annotation. In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, pages 5–9, Santa Fe, New Mexico, 2018.
- [41] Rodrigo Moraes, João Francisco Valiati, and Wilson P Gavião Neto. Document-level sentiment classification: An empirical comparison between svm and ann. *Expert Systems with Applications*, 40(2):621–633, 2013.
- [42] Maria Rauschenberger, Martin Schrepp, Manuel Perez Cota, Siegfried Olschner, and Jörg Thomaschewski. Efficient Measurement of the User Experience of Interactive Products. How to use the User Experience Questionnaire (UEQ). Example: Spanish Language Version. *International Journal of Interactive Multimedia and Artificial Intelligence*, 2(1):39–45, 03/2013 2013.
- [43] J. Cohen. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1):37, 1960.
- [44] Klaus Krippendorff. Reliability in content analysis. *Human communication research*, 30(3):411–433, 2004.
- [45] William A. Scott. Reliability of Content Analysis: The Case of Nominal Scale Coding. *Public Opinion Quarterly*, 19(3):321–325, 01 1955.
- [46] Mary McHugh. Interrater reliability: The kappa statistic. *Biochemia medica : časopis Hrvatskoga društva medicinskih biokemičara / HDMB*, 22:276–82, 10 2012.
- [47] Klaus Krippendorff. Reliability in content analysis: Some common misconceptions and recommendations. *Human communication research*, 30(3):411–433, 2004.
- [48] Nathalie Japkowicz. Why question machine learning evaluation methods. In *AAAI workshop on evaluation methods for machine learning*, pages 6–11, 2006.
- [49] Meaning Cloud LLC. Meaning cloud. <https://www.meaningcloud.com/developer/documentation/ontology>, 2021.
- [50] Google Cloud. Google cloud ontology. <https://cloud.google.com/natural-language/docs/reference/rest/v1/Entity?hl=es-419>, 2021.
- [51] Rosario Sanchis-Font, Maria Jose Castro-Bleda, Jose-Angel Gonzalez, Ferran Pla, and Lluís-F Hurtado. Cross-domain polarity models to evaluate user experience in e-learning. *Neural processing letters*, pages 1–17, 2020.
- [52] Eugenio Martinez-Camara, Manuel C Diaz-Galiano, M Angel Garcia-Cumbreras, Manuel Garcia-Vega, and Julio Villena-Román. Overview of tass 2017. *Proceedings of TASS*, pages 13–21, 2017.
- [53] Hana Ovesleová. User-interface supporting learners’ motivation and emotion: A case for innovation in learning management systems. In Aaron Marcus, editor, *Design, User Experience, and Usability: Novel User Experiences*, pages 67–75, Cham, 2016. Springer International Publishing.
- [54] Martin Schrepp. User experience questionnaire handbook. *All you need to know to apply the UEQ successfully in your project*, 2015.

APÉNDICE A

Guía etiquetado Doccano

Doccano es una herramienta de código abierto utilizada para la anotación manual. Sus dos características principales son que permite el etiquetado colaborativo y que es multilingüe. El servidor de Doccano que se va a utilizar está alojado en la plataforma online Heroku, concretamente en la url: <https://pruebatfgaspectos.herokuapp.com/>.

A.1 Primeros pasos

Para comenzar a etiquetar el corpus se deben seguir los siguientes pasos:

1. Acceder a <https://pruebatfgaspectos.herokuapp.com/>.
2. Iniciar sesión en el sistema.
 - a) Clicar en *login*. Ver imagen A.1.

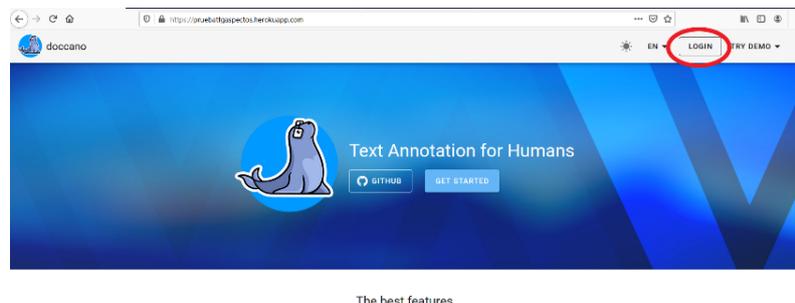


Figura A.1: Login.

- b) Introducir credenciales.
3. Escoger proyecto a etiquetar. Para este trabajo el proyecto es: TFG-Aspectos. Ver imagen A.2
 4. Clicar en comenzar anotación. Ver A.3

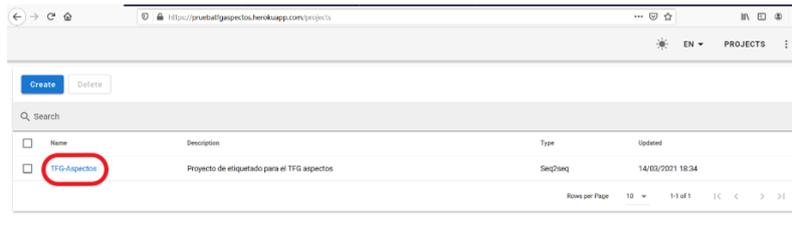


Figura A.2: Elección del proyecto.

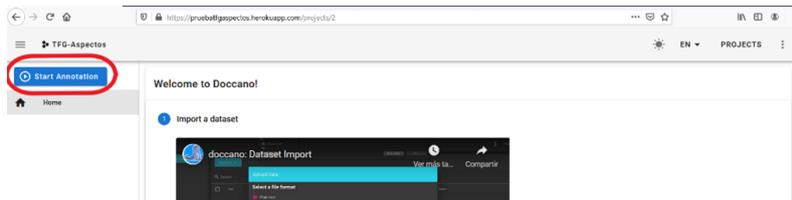


Figura A.3: Comenzar anotación.

A.2 Anotación

Para etiquetar una frase seguir el siguiente procedimiento:

1. Copiar la frase a etiquetar.
2. Sustituir la etiqueta de polaridad por la polaridad detectada.
3. Guardar etiquetado (apretar tecla intro).

Copiar la frase a etiquetar. Sustituir la etiqueta de polaridad por la polaridad detectada. Guardar etiquetado (apretar tecla intro).

Ejemplo: El contenido de los |apuntes#AMATERIAL#Polaridad| también es bueno así como las |clases#ACLASE#Polaridad|. El contenido de los |apuntes#AMATERIAL#P| también es bueno así como las |clases#ACLASE#P|.

Conjunto de polaridades:

- P: positiva
- N: negativa
- NEU: neutra

P: positiva **N:** negativa **NEU:** neutra

En caso de duda o caso especial usar los comentarios para dejar constancia.

A.3 Comentarios

En caso de que se tengan dudas del etiquetado, exista un caso especial en el etiquetado o se quiera destacar algo de este se debe escribir un comentario que explique el problema. Para escribir un comentario se puede utilizar el botón destacado en la imagen [A.4](#).

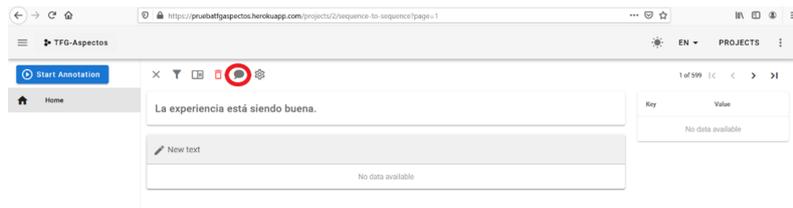


Figura A.4: Crear comentarios.

A.4 Guía auxiliar

Además de esta guía, dentro del proyecto de etiquetado se puede consultar una guía breve de etiquetado. Para verla se debe apretar el botón destacado en la imagen [A.5](#).

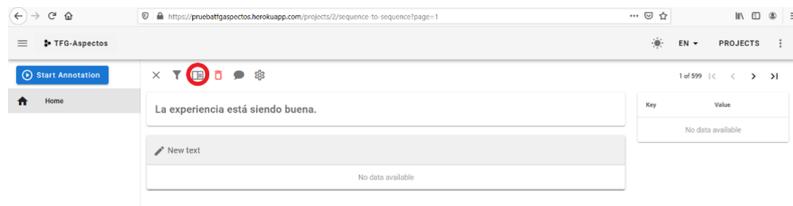


Figura A.5: Ver guía auxiliar.

Y se verá la imagen [A.6](#).

Annotation Guideline

Para etiquetar una frase seguir el siguiente procedimiento:

1. Copiar la frase a etiquetar.
2. Sustituir la etiqueta de polaridad por la polaridad detectada.
3. Guardar etiquetado (apretar intro).

Ejemplo:

El contenido de los |apuntes#AMATERIAL#Polaridad| también es bueno así como las |clases#ACLASE#Polaridad|.

El contenido de los |apuntes#AMATERIAL#P| también es bueno así como las |clases#ACLASE#P|.

Conjunto de polaridades:

P: positiva

N: negativa

NEU: neutra

En caso de duda o caso especial usar los comentarios para dejar constancia.

[Close](#)

Figura A.6: Guía auxiliar.

APÉNDICE B

Aspectos por categoría

- **VLE**: curso, entorno, experiencia, master, máster, plataforma, herramienta, aplicación, interfaz, postgrado, posgrado, navegador.
- **Alumno**: alumno, estudiante, integrante, participante.
- **Comunicación**: anuncio, comunicación, correo, correo-electrónico, foro, interacción, mensaje, acceso, notificación, aviso, chat, email, mail, post.
- **Ejercicio**: actividad, ejercicio, tarea, cuestionario, encuesta, preguntas, preguntar, respuesta, entregable, entrega.
- **Evaluación**: evaluación, examen, pruebas, revisión, test, certificado, certificación, corrección, corregir, nota, calificación.
- **Imagen**: apuntador, audiovisual, material audiovisual, dispositivo señalador, señalador, enfocar, enfoque, filmación, filmado, imagen, láser, mostrar, powerpoint, ppt, proyectar, puntero, señalar, ver, vídeo, reproducir, pantalla pantallazo, slide (ENG), resolución, HD, cámara.
- **Material**: apuntes, clase, conferencia, contenido, diapositiva, lección, material, módulo, presentación, recursos, tema, transparencia, archivo, charla, disertación, explicación, exposición.
- **Profesor**: docente, docencia, enseñanza, maestro, ponente, profesor, tutor, instructor, presentador.
- **Sonido**: audio, audiovisual, entender, escuchar, oír, sonido, volumen.