

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

ESCUELA DE POSGRADO



REVISIÓN SISTEMÁTICA SOBRE LA  
APLICACIÓN DE ONTOLOGÍAS DE DOMINIO  
EN EL ANÁLISIS DE SENTIMIENTO

Tesis para optar el grado de Magíster en Informática con mención en  
Ciencias de la Computación que presenta

CÉSAR AUGUSTO OLIVARES POGGI

Dirigido por

DR. HÉCTOR ANDRÉS MELGAR SASIETA

Jurado

MG. CLAUDIA MARÍA DEL PILAR ZAPATA DEL RÍO

DR. HÉCTOR ANDRÉS MELGAR SASIETA

MG. LUIS ALBERTO FLORES GARCÍA

San Miguel, 2016

## Resumen

El análisis de sentimiento es un área de creciente investigación en los campos del procesamiento de lenguaje natural y la recuperación de información. En los últimos años ha habido un aumento en la aplicación de técnicas semánticas en el análisis de sentimiento, en particular con el apoyo de la aplicación de ontologías de dominio. Sin embargo, en la literatura actual no se cuenta con un estudio que reporte de manera sistemática los beneficios alcanzados con la aplicación de ontologías de dominio al análisis de sentimiento. Esta revisión sistemática tiene por objetivos realizar dicha síntesis, reportar el grado de generalización de las investigaciones realizadas, verificar el aprovechamiento de la riqueza expresiva de las ontologías de dominio y señalar el estado del arte actual en la representación de las emociones humanas por medio de ontologías de dominio en su aplicación al análisis de sentimiento.

Se identificó 9 distintos problemas del análisis del sentimiento a los que se aplicó ontologías de dominio y un total de 22 beneficios de dicha aplicación. Los beneficios más reportados son: (1) el soporte para una representación estructurada de las opiniones y la vinculación de datos; (2) mayor precisión y exhaustividad en la clasificación de la polaridad; y (3) soporte para la representación de modelos emocionales.

Como investigación futura se sugiere profundizar en el empleo de ontologías de dominios para analizar el sentimiento a nivel de conceptos, modelar el proceso de análisis de sentimiento, estandarizar la elaboración de ontologías de productos e integrar diversos modelos emocionales, así como aprovechar mejor la expresividad semántica y capacidad de razonamiento de las ontologías de dominio.

**Palabras clave.** Sentiment Analysis, Opinion Mining, Domain Ontologies, Systematic Literature Review

# Índice general

<b>1. Introducción</b>	<b>7</b>
1.1. Análisis de sentimiento . . . . .	8
1.2. Ontologías . . . . .	10
<b>2. Método</b>	<b>13</b>
2.1. Revisión sistemática . . . . .	13
<b>3. Protocolo de Revisión</b>	<b>15</b>
3.1. Preguntas de Investigación . . . . .	15
3.2. Estrategia de búsqueda y selección de fuentes . . . . .	16
3.2.1. Definición de los términos de búsqueda . . . . .	16
3.2.2. Selección de fuentes . . . . .	19
3.2.3. Documentación del proceso de búsqueda . . . . .	19
3.2.4. Criterios de inclusión y exclusión . . . . .	20
3.2.5. Criterios de estimación de calidad . . . . .	21
3.2.6. Proceso de selección . . . . .	21
3.3. Estrategia de extracción de datos . . . . .	22
3.3.1. Formulario de extracción de datos . . . . .	22
3.3.2. Procedimiento para la extracción de datos . . . . .	22
3.4. Estrategia de síntesis de datos . . . . .	22
<b>4. Ejecución de la Revisión Sistemática de la Literatura</b>	<b>25</b>
4.1. Búsqueda y selección de estudios primarios . . . . .	25
4.2. Extracción y síntesis de datos . . . . .	27
4.3. Visión general de los estudios incluidos . . . . .	27
4.3.1. Años de publicación . . . . .	27
4.3.2. Fuentes de publicación . . . . .	29
4.3.3. Centros de investigación . . . . .	29
4.3.4. Número de citas . . . . .	29

<b>5. Resultados de la Revisión Sistemática de la Literatura</b>	<b>33</b>
5.1. Contribución al análisis de sentimiento (RQ1)	33
5.1.1. Niveles de análisis de sentimiento (RQ1.1)	33
5.1.2. Fases del análisis de sentimiento (RQ1.2)	35
5.1.3. Problemas del análisis de sentimiento y beneficios reportados (RQ1.3)	37
5.2. Aplicabilidad a otros dominios (RQ2)	51
5.2.1. Resultados	51
5.2.2. Análisis y discusión	53
5.3. Expresividad y razonamiento ontológico (RQ3)	54
5.3.1. Componentes de representación del conocimiento (RQ3.1)	54
5.3.2. Inferencias y razonamiento ontológico (RQ3.2)	56
5.4. Representación de emociones (RQ4)	57
5.4.1. Resultados	57
5.4.2. Análisis y discusión	58
<b>6. Trabajos futuros y Conclusiones</b>	<b>61</b>
6.1. Trabajos futuros	61
6.2. Conclusiones	63
<b>Bibliografía</b>	<b>65</b>

# Capítulo 1

## Introducción

El análisis de sentimiento es un área de creciente investigación en los campos del procesamiento de lenguaje natural y la recuperación de información. Hay un notable interés y necesidad por parte de compañías, políticos, así como consumidores, de contar con herramientas que les permitan identificar y sintetizar las opiniones y sentimientos que las personas expresan en blogs, reseñas de productos, redes sociales y en general cualquier tipo de repositorio digital.

Las primeras investigaciones en análisis de sentimiento se concentraron en identificar la polaridad (positiva o negativa) de las opiniones por medio de métodos estadísticos a partir de una representación «bag of words» de los documentos en los que ellas se expresan, es decir, sin tomar en consideración el significado de las palabras ni las relaciones sintácticas o semánticas entre sí. Posteriores investigaciones han realizado caracterizaciones más detalladas de los sentimientos, distinguiendo niveles de polaridad, identificando si un documento expresa o no una opinión (análisis de la subjetividad), diferenciando los sentimientos de las distintas oraciones o frases dentro de un mismo documento, diferenciando los sentimientos atribuidos a distintos atributos de un mismo producto, o más recientemente identificando las emociones humanas manifestadas en las opiniones.

Sin embargo, los sentimientos humanos son usualmente expresados con mucha sutileza y unas mismas palabras pueden indicar un sentimiento u otro dependiente del contexto en el que se encuentran. Se hace necesario un análisis semántico que reconozca los conceptos a los que aluden las palabras y cómo éstos indican diversos sentimientos en un dominio u otro.

Con esta finalidad en los últimos años las investigaciones en análisis de sentimiento han venido introduciendo el uso de ontologías de dominio y otras tecnologías de web semántica. No obstante, en la literatura actual no se cuenta con un estudio que reporte de manera sistemática los beneficios alcanzados

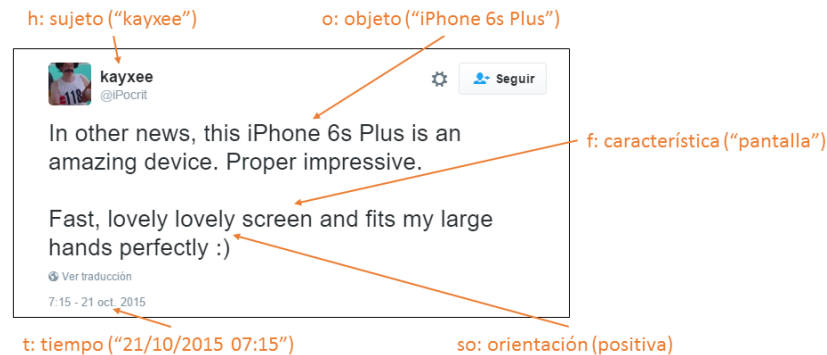


Figura 1.1: Elementos de una opinión

con la aplicación de tecnologías semánticas al análisis de sentimiento.

## 1.1. Análisis de sentimiento

El análisis de sentimiento o minería de opinión ha sido definido como el estudio computacional de las opiniones, sentimientos y emociones expresadas en texto[68]. Recientemente se considera también como parte de esta área el análisis de otros tipos de fuentes digitales, tales como contenido visual[13] o multimodal[44].

Liu [68] ha representado una opinión como un quintuplo (o, f, so, h, t), tal como se muestra en la figura 1.1:

- o: objeto (p.ej. «iPhone»)
- f: característica o aspecto (p.ej. «pantalla»)
- so: orientación o polaridad de la opinión sobre la característica f del objeto o (p.ej. polaridad positiva, o un valor numérico correspondiente a un grado dentro de una escala de polaridad)
- h: el sujeto de la opinión (p.ej. «Juan Pérez»)
- t: el momento en el que la opinión ha sido expresada

Kumar y Sebastian [60] distinguen las siguientes tareas en el proceso del análisis de sentimiento:

- Clasificación de la subjetividad
- Clasificación del sentimiento
- Extracción del sujeto de la opinión
- Extracción del objeto y la característica/aspecto

## 1.1. ANÁLISIS DE SENTIMIENTO

9

A estas tareas se podría añadir la sumarización de los sentimientos encontrados.

El análisis de sentimiento ha venido evolucionando y pasando por diversos niveles de granularidad [60, 115]:

- *A nivel de documento.* Consiste en determinar la polaridad o sentimiento general de un documento completo. Se considera el documento entero como una única opinión emitida por un sujeto sobre un único objeto. Usualmente depende de la determinación del sentimiento a nivel de palabra y a nivel de frase.
- *A nivel de oración.* Consiste en calcular la polaridad individual de cada oración que contenga una opinión [54]. Se clasifica primero la oración en objetiva o subjetiva. Cada oración subjetiva es considerada una única opinión emitida por un sujeto sobre un único objeto. También depende de la determinación del sentimiento a nivel de palabra y a nivel de frase.
- *A nivel de palabra.* Consiste en utilizar la polaridad asociada a cada palabra para inferir la polaridad del texto que las contiene. [39]
- *A nivel de frase.* Se diferencia de la granularidad a nivel de palabra en que se considera grupos de evaluación, en vez de palabras, como unidades mínimas de expresión. [119] Se considera el contexto de cada palabra para corregir su polaridad general.
- *A nivel de aspecto/característica.* Consiste en identificar el sentimiento de manera diferenciada para cada atributo de un producto. [90]
- *A nivel de concepto.* Consiste en utilizar el recurso a grandes bases de conocimiento para tomar como unidad mínima de expresión los conceptos expresados en lenguaje natural de manera implícita o explícita. [20]

En los últimos años el análisis de sentimiento ha ido más allá de la mera detección de la polaridad, y busca identificar estados emocionales tales como la alegría, ira, la tristeza o incluso la ironía [5, 18, 14]. En este punto, el análisis de sentimiento se conecta con las investigaciones realizadas en el campo de la computación afectiva, esto es, la computación que está relacionada con, que surge de, o que influencia las emociones [85] [2]. La computación afectiva nació con la motivación de introducir la empatía en la interacción persona-computador, mediante sistemas capaces de reconocer, procesar y expresar emociones. Sus campos de aplicación han estado tradicionalmente ligados al reconocimiento de emociones en la voz y el rostro. También se ha realizado estudios para el reconocimiento de emociones en el texto escrito, usualmente con el objetivo de generar la conversión de texto a voz con una entonación

acorde a la emoción subyacente en el discurso analizado [86, 30, 4, 42]. En esto se diferencia del análisis de sentimiento, cuya motivación está más en la automatización de las opiniones, sentimientos y emociones expresadas por muchos individuos sobre diversos temas.

Mientras que la polaridad y grados de subjetividad son fácilmente representables con escalas numéricas, las emociones humanas constituyen un fenómeno mucho más complejo, ambiguo e impreciso [37]. Tanto la computación afectiva como el análisis de sentimiento requieren por ello apoyarse en modelos desarrollados por la psicología para representar el espacio de las emociones humanas, tales como los de Plutchik [88], Ekman [38], Parrot [82], o Mehrabian [74].

Se tiene tres tipos fundamentales de representación de las emociones:

- *Representación categorial.* Cada emoción corresponde a un valor nominal discreto. Por ejemplo, en el modelo de Ekman se enuncia seis emociones básicas: ira, disgusto, miedo, alegría, tristeza y sorpresa.
- *Representación dimensional.* Las distintas emociones son representadas como la combinación de varios factores o dimensiones continuas. Por ejemplo, en el modelo de Mehrabian, el entusiasmo corresponde a la combinación de un grado alto de placer, un grado alto de activación y un grado moderado de dominancia.
- *Representación componencial.* Las emociones son entendidas más bien como un proceso que define un perfil de actividad en diversos componentes donde cada uno gatilla el siguiente: evento, evaluación, respuesta emocional, preparación para la acción, conducta y regulación. A partir de la evaluación de un evento, se produce una respuesta emocional que a su vez orienta la conducta. [43, 73]

## 1.2. Ontologías

En los últimos años, las investigaciones en análisis de sentimiento han venido introduciendo el uso de ontologías para pasar de una clasificación de sentimientos a nivel de términos a una clasificación semántica, esto es, a nivel de conceptos [20, 50, 79, 19]. Lenguajes como RDF y OWL permiten la representación e integración de datos para una web semántica [10] con la riqueza conceptual de las ontologías [53].

Una ontología es una definición formal de los términos y relaciones básicas que comprenden el vocabulario de un área temática así como las reglas para combinar términos y relaciones para definir extensiones del vocabulario [77].



## 1.2. ONTOLOGÍAS

11

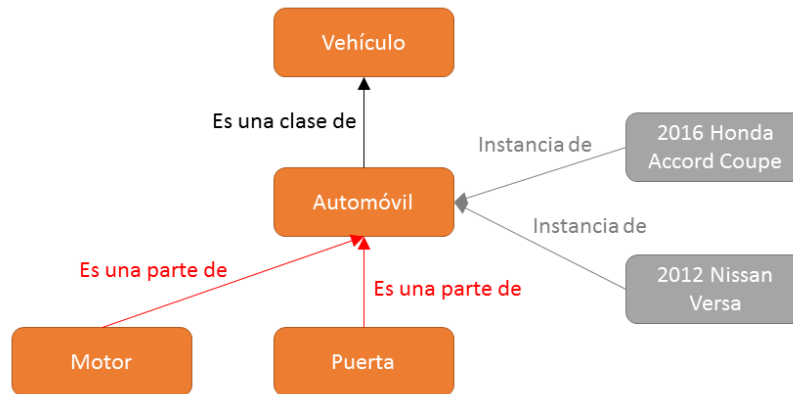


Figura 1.2: Algunos elementos de una ontología: clases (anaranjado), instancias (gris), relaciones de clase (negro), relaciones ad-hoc (rojo).

Independientemente del lenguaje de formalización, a nivel de conceptualización una ontología consta de los siguientes componentes principales [28] (ver figura 1.2):

- *Conceptos* o clases de entidades, p.ej.: automóvil, motor.
- *Relaciones*, que pueden ser (1) *de clase (taxonómicas)*, p.ej.: automóvil *es una clase de* vehículo; o (2) relaciones ad hoc, p.ej.: motor *es una parte de* automóvil.
- *Instancias*, que representan elementos que pertenecen a una clase, p.ej.: 2016 Honda Accord Coupe.
- *Constantes*, que describen valores inmutables o que no cambian mucho en el tiempo, p.ej.: velocidad máxima para conducir en carretera en Perú.
- *Atributos*, que describen propiedades (1) de instancias, p.ej.: *velocidad máxima* de 2016 Honda Accord Coupe; o (2) de clases, p.ej.: *número de puertas* de la clase «automóvil de dos puertas».
- *Axiomas formales*, o expresiones lógicas que definen restricciones en la ontología, p.ej.: automóvil sólo puede tener un tipo de tracción (simple o doble).
- *Reglas*, que permiten inferir conocimiento en la ontología, tales como valores de atributos o instancias de relaciones, p.ej.: si la ubicación del timón es a la derecha, el automóvil no es legal en Perú.

Según el criterio de clasificación se puede distinguir diversos tipos de ontologías[48]. Entre ellos destacan:

- *Ontologías de nivel superior.* Son aquellas que representan conceptos y nociones generales a los que pertenecen los elementos raíz de toda otra ontología. No están asociadas a ningún dominio en particular. Como ejemplo se puede citar la ontología Cyc [64] y la ontología de Sowa [107].
- *Ontologías de dominio.* Son aquellas que proveen vocabularios acerca de conceptos propios de un dominio, sus relaciones, las actividades que tienen lugar en ese dominio, las teorías y principios elementales que lo gobiernan. A este tipo pertenecen las ontologías utilizadas en las investigación que abarca esta revisión sistemática, tales como Human Emotion Ontology (HEO) [50] (ver figura 5.2), OntoEmotion [40] (ver figura 5.3) y Onyx [96] (ver figura 5.4).
- *Ontologías lingüísticas.* Son aquellas que describen constructos semánticos en vez de modelar un dominio específico. Son muy empleadas como bases de datos lexicográficas para el procesamiento de lenguaje natural. Como ejemplo se puede citar BabelNet [?], WordNet [?, ?] y SentiWordNet [?].

# Capítulo 2

## Método

### 2.1. Revisión sistemática

Este estudio se apoya en la metodología señalada por B. Kitchenham [58] para la realización de revisiones sistemáticas. Una revisión sistemática es un proceso formal y repetible para identificar, evaluar e interpretar toda la investigación disponible relacionada con una pregunta de investigación.

El proceso de realización de revisión sistemática comprende las siguientes fases y etapas:

1. Planificar la revisión
  - Identificación de la necesidad de una revisión
  - Encargo de una revisión (opcional)
  - Especificación de la(s) pregunta(s) de investigación
  - Desarrollo de un protocolo de revisión
  - Evaluación del protocolo de revisión (opcional)
2. Realizar la revisión
  - Identificación de la investigación
  - Selección de estudios primarios
  - Evaluación de la calidad de los estudios
  - Extracción de datos y monitoreo
  - Síntesis de los datos
3. Reportar la revisión
  - Especificación de los mecanismos de difusión

- Formateo del reporte principal
- Evaluación del reporte (opcional)

Es importante tomar en cuenta que estas etapas no son completamente secuenciales y que muchas de ellas implican un proceso iterativo de progresiva revisión y refinamiento.

El protocolo de revisión debe especificar el contexto de la revisión, las preguntas de investigación, la estrategia de búsqueda, los criterios de selección de los estudios, los procedimientos para la selección de los estudios, los procedimientos para la evaluación de la calidad de los estudios, la estrategia de extracción de datos y de la síntesis de los datos extraídos, la estrategia de difusión y el cronograma del proyecto de revisión.

Para la estructuración de las preguntas de investigación se recomienda usar los criterios PICOC (*Population, Intervention, Comparison, Outcome, Context*) propuestos por [8].

La estrategia de búsqueda comprende la selección de fuentes y la definición de los términos de búsqueda. En esta etapa se debe elegir las fuentes o bases de datos más relevantes y que contengan los estudios más importantes y de mayor impacto sobre el tema de investigación.

Luego de una búsqueda inicial en las fuentes seleccionadas se puede realizar una verificación y análisis de las referencias contenidas en los estudios seleccionados para validar que no se haya dejado de lado estudios relevantes.

Los criterios de selección tienen como objetivo identificar todos los estudios primarios que contienen evidencia directa acerca de la pregunta de investigación, así como reducir el riesgo de parcialidad en la revisión. Deben estar basados en la pregunta de investigación y pueden ser refinados durante el proceso de búsqueda.

La evaluación de la calidad busca asegurar la credibilidad de los estudios a incluir desde el punto de vista de su imparcialidad y validez interna y externa.

El proceso de extracción de datos tiene como objetivo identificar en cada uno de los estudios los elementos de información que se necesita para responder a las preguntas de investigación.

El proceso de análisis de datos tiene el fin de sintetizar los datos de un modo tal que se responda a las preguntas de investigación.

## Capítulo 3

# Protocolo de Revisión

### 3.1. Preguntas de Investigación

Esta revisión sistemática tuvo como objetivo identificar de qué manera las ontologías de dominio vienen siendo aplicadas en el proceso de análisis de sentimiento. Según ello, se buscó responder a la siguiente pregunta de investigación principal: *¿De qué manera el uso de ontologías de dominio está apoyando el análisis de sentimiento?*

Para estructurar los elementos de la pregunta de investigación se usó los criterios PICOC propuestos en [8], según se muestra en la tabla 3.1. Dado que como parte del estudio no se incluye la comparación de la aplicación de ontologías de dominio con otras aproximaciones, métodos o modelos, el elemento “Comparison” de PICOC no aplica.

Tabla 3.1: Criterios PICOC

Population	Análisis de sentimiento de cualquier tipo de contenido digital (textual, visual, multimodal, etc.)
Intervention	Aplicación de ontologías de dominio
Comparison	<i>No aplica</i>
Outcome	Cualquier tipo de mejora o beneficio reportado
Context	Estudios de investigación primaria en el área de las ciencias de la computación orientadas a mejorar cualquier aspecto del proceso de análisis de sentimiento.

A partir de la pregunta general de investigación se formuló un conjunto de

preguntas específicas para los diversos aspectos de la aplicación de ontologías de dominio en el análisis de sentimiento que se desea investigar:

RQ1. ¿Cómo está contribuyendo el uso de ontologías de dominio a la solución de problemas específicos en el análisis de sentimiento?

RQ1.1 ¿Qué nivel de análisis de sentimiento está siendo apoyado con el uso de ontologías de dominio?

RQ1.2 ¿Qué fases del análisis de sentimiento están siendo apoyadas con el uso de ontologías de dominio?

RQ1.3 ¿Qué problemas en el análisis de sentimiento están siendo abordados con el apoyo de ontologías de dominio y qué beneficios han sido reportados con su aplicación?

RQ2. ¿Las investigaciones reportadas son generalizables a dominios diferentes al de las ontologías aplicadas durante la experimentación?

RQ3. ¿En qué medida se está aprovechando la expresividad y mecanismos de razonamiento propios de las ontologías de dominio en el análisis de sentimiento?

RQ3.1 ¿Qué componentes de representación del conocimiento propios de las ontologías están siendo aplicados en el análisis de sentimiento?

RQ3.2 ¿Qué razonadores están siendo usados para realizar inferencias en la aplicación de ontologías de dominio en el análisis de sentimiento?

RQ4. ¿Cómo se está representando las emociones en las ontologías de dominio aplicadas en el análisis de sentimiento?

RQ4.1 ¿En qué modelo psicológico está fundada la clasificación de emociones usada en las ontologías de emociones aplicadas en el análisis de sentimiento?

RQ4.2 ¿Qué tipo de representación es usada para las emociones en las ontologías de emociones aplicadas en el análisis de sentimiento?

## 3.2. Estrategia de búsqueda y selección de fuentes

### 3.2.1. Definición de los términos de búsqueda

Para la definición de los términos de búsqueda se tomó en cuenta las sugerencias presentadas en [58], en particular las siguientes:

- Descomposición de la pregunta de investigación en sus componentes PICOC.
- Búsquedas preliminares usando varias combinaciones de términos de búsqueda derivadas de la pregunta de investigación.

### 3.2. ESTRATEGIA DE BÚSQUEDA Y SELECCIÓN DE FUENTES 17

- Adición de sinónimos, abreviaturas y formas ortográficas alternativas si aplicasen.
- Revisión de títulos, sumarios y palabras clave.
- Uso de operadores lógicos AND y OR.
- Verificación de que las cadenas de búsqueda de prueba recuperen la lista preliminar de estudios primarios ya identificados.

#### Elementos de la cadena de búsqueda

Para identificar los elementos de la cadena de búsqueda se partió por tomar como referencia los principales términos discriminantes del cuadro 3.1. Los términos “sentiment analysis” fueron desde un principio acompañados por la alternativa “opinion mining”. El resultado se ve en la tabla 3.2. Con dichos términos se procedió a una primera fase de búsqueda preliminar.

Tabla 3.2: Términos derivados de PICOC

Population	sentiment analysis, opinion mining
Intervention	ontolog*
Comparison	
Outcome	
Context	computer science.

#### Búsqueda preliminar

Se hizo búsquedas preliminares en Google, Google Scholar y Scopus con el fin de verificar la existencia de revisiones sistemáticas existentes, hacer una primera estimación del volumen de estudios relevantes e identificar un primer conjunto de publicaciones relevantes. Para las búsquedas se usó la cadena “ontology AND (“sentiment analysis” OR “opinion mining”)”. No se encontró revisiones sistemáticas previas sobre el tema.

Para estimar el volumen de estudios se tomó como referencia la búsqueda en Scopus, en la que se restringió la búsqueda a publicaciones clasificadas en el área de las Ciencias de la Computación. La cadena de búsqueda empleada fue: “ALL ( ontolog\* ) AND TITLE-ABS-KEY ( “sentiment analysis” OR “opinion mining” ) AND ( LIMIT-TO ( SUBJAREA , “COMP” ) ) ”.

El 23 de abril de 2015 se obtuvo 357 resultados, ordenados por fecha de publicación, de más recientes a más antiguos. Se revisó los sumarios de las

primeras 100 publicaciones y se estimó de modo preliminar que aproximadamente un 30% de ellas podría ser de interés para el tema de investigación.

Se registró un primer conjunto de publicaciones relevantes al tema de investigación. En este proceso se consideró también estudios secundarios. Algunas publicaciones fueron parte de los resultados de búsqueda, otras fueron identificadas por ser citadas en uno o más de los estudios revisados. Éstas son reportadas en la tabla 3.3.

Tabla 3.3: Publicaciones relevantes identificadas en la búsqueda preliminar

Estudios primarios	[89, 101, 102, 17, 115, 21, 6, 3, 18, 59]
Estudios secundarios	[99, 80, 57, 60, 20, 5]

### Adición de términos y concatenación de la cadena de búsqueda

A partir de los títulos, sumarios, palabras clave y contenido las publicaciones relevantes identificadas en la búsqueda preliminar, se identificó sinónimos y términos relacionados alternativos para completar la cadena de búsqueda. Éstos se reportan en la tabla 3.4.

Tabla 3.4: Sinónimos y términos alternativos de búsqueda

sentiment analysis	opinion mining, opinion analysis, subjectivity analysis, polarity mining, sentiment classification
ontology	ontologies, ontological
computer science	

Dado que la investigación tuvo como objetivo identificar artículos del área de conocimiento del análisis de sentimiento, se añadió como restricción, donde se pudiera, que el término “sentiment analysis” o sus sinónimos fueran parte de los metadatos de la publicación (título, sumario o palabras clave).

Teniendo ello en consideración, se unió los términos con los operadores OR y AND, obteniéndose la siguiente cadena de búsqueda, enunciada en la sintaxis de la base de datos Scopus:

```
ALL ( ontolog* )
AND
TITLE-ABS-KEY ( "sentiment analysis"
```



### 3.2. ESTRATEGIA DE BÚSQUEDA Y SELECCIÓN DE FUENTES 19

```

OR "opinion mining"
OR "opinion analysis"
OR "subjectivity analysis"
OR "polarity mining"
OR "sentiment classification" )
AND
  LIMIT-TO ( SUBJAREA , "COMP" )

```

#### 3.2.2. Selección de fuentes

La selección de bases de datos fue realizada a partir de otras revisiones sistemáticas del área de ciencias de la computación y de las sugerencias de la web de la Pontificia Universidad Católica del Perú para esta área. Las bases de datos consultadas fueron:

- ACM Digital Library<sup>1</sup>
- IEEE Xplore<sup>2</sup>
- ProQuest<sup>3</sup>
- ScienceDirect<sup>4</sup>
- Scopus<sup>5</sup>
- Springer Link<sup>6</sup>
- Web of Science<sup>7</sup>

#### 3.2.3. Documentación del proceso de búsqueda

El proceso de búsqueda fue documentado con la finalidad de hacerlo lo más transparente y replicable posible. Para las búsquedas en bases de datos electrónicas, siguiendo lo indicado por Kitchenham [58], se documentó la siguiente información:

- Nombre de la base de datos
- Estrategia de búsqueda para la base de datos

<sup>1</sup><http://dl.acm.org/>

<sup>2</sup><http://ieeexplore.ieee.org/>

<sup>3</sup><http://search.proquest.com/>

<sup>4</sup><http://www.sciencedirect.com/>

<sup>5</sup><https://www.scopus.com/>

<sup>6</sup><http://link.springer.com/>

<sup>7</sup><http://isiknowledge.com/>

- Fecha de búsqueda
- Años cubiertos por la búsqueda
- Número de resultados
- Número de duplicados
- Número de nuevos artículos relevantes
- Campos examinados durante la búsqueda

### 3.2.4. Criterios de inclusión y exclusión

Se incluyó los estudios que cumplieron con todos los siguientes criterios de inclusión:

1. El estudio realiza un procedimiento de experimentación en análisis de sentimiento.
2. El estudio usa ontologías de dominio como parte del procedimiento de análisis de sentimiento.
3. El estudio realiza una validación empírica de sus hallazgos.
4. El estudio está escrito en inglés, español, francés, italiano o alemán.

No se hizo distinción en el tipo de experimentación realizada ni se puso restricciones en el tipo de validaciones empíricas. Sin embargo, para la inclusión no se consideró que se haya usado ontologías de dominio en la experimentación; se juzgó necesario que éstas hayan sido usadas como parte del análisis de sentimiento.

No se puso límite en cuanto a la fecha de publicación de los estudios. El tema de estudio corresponde a un área nueva de investigación y en las etapas preliminares no se encontró estudios anteriores a 2008.

Se omitió en la revisión sistemática los estudios que cumplieron con el siguiente criterio de exclusión:

1. El estudio es redundante y menos detallado con respecto otro estudio de la misma autoría. Este criterio se aplica cuando dos o más estudios de los mismos autores reportan una misma investigación sin que se detalle mayor información sobre el uso y beneficios de la aplicación de ontologías de dominio. Se selecciona el estudio con mayor detalle, o si no hubiera diferencia, el primero en haber sido publicado.

### 3.2. ESTRATEGIA DE BÚSQUEDA Y SELECCIÓN DE FUENTES 21

#### 3.2.5. Criterios de estimación de calidad

Se elaboró una lista de preguntas para evaluar la calidad y credibilidad de los estudios desde el punto de vista de su imparcialidad y de su validez interna y externa. Se tomó como referencia las preguntas enumeradas en [58]. Dado que la pregunta de investigación de la revisión sistemática es amplia y abierta a diversos tipos de trabajo, las preguntas escogidas fueron generales y no abarcaron procesos específicos del análisis de sentimiento.

1. ¿Se indica claramente el objetivo del estudio?
2. ¿El estudio indica su contribución en relación con el estado del arte actual?
3. ¿Se describe con claridad el marco conceptual del estudio?
4. ¿Se describe con claridad la técnica propuesta?
5. ¿Se presenta un caso de estudio replicable?
6. ¿Se discute los resultados del estudio?
7. ¿Se discute las limitaciones del estudio?
8. ¿La discusión y conclusiones son claras y coherentes?

Cada pregunta tuvo tres opciones de respuesta que siguen la siguiente escala: sí (1), no (0), parcialmente (0.5). Por lo tanto, los resultados variaron entre 0 (pésima calidad) y 8 (óptima calidad).

#### 3.2.6. Proceso de selección

El proceso de selección se realizó en dos fases, que se describen a continuación.

##### Selección inicial

En la primera etapa del proceso de selección se revisó los títulos y sumarios de los estudios recuperados y se eliminó aquellos que claramente no cumplían con los criterios de inclusión y exclusión.

### Selección final

En la segunda etapa del proceso de selección se recuperó y revisó el texto completo de los estudios restantes para verificar el cumplimiento de los criterios de inclusión y exclusión. Los estudios fueron evaluados según los criterios de estimación de calidad. Se determinó excluir los estudios con un puntaje menor a 4. Dicho puntaje, que corresponde al 50 % del puntaje máximo, fue establecido en el protocolo de revisión como umbral aceptable de calidad para las publicaciones.

## 3.3. Estrategia de extracción de datos

La información de los estudios seleccionados fue extraída y registrada en vistas a responder a las preguntas de investigación. A continuación se describe el formulario y el procedimiento empleados en esta etapa.

### 3.3.1. Formulario de extracción de datos

El formulario de extracción de datos contuvo tanto la información general sobre cada estudio como los detalles completos que permitiesen describir cómo cada uno de ellos respondía a las preguntas de investigación de la revisión sistemática. El formulario constó de los campos indicados en la tabla 3.5. Cada uno de los campos tuvo anexo un campo de notas adicionales.

### 3.3.2. Procedimiento para la extracción de datos

Para cada uno de los estudios seleccionados, se completó los formularios de extracción indicados en la sección 3.3.1.

Se leyó completamente la introducción y conclusión de los estudios. Se leyó con detenimiento el texto restante en la medida que fue necesario para extraer los datos requeridos.

## 3.4. Estrategia de síntesis de datos

Una vez que se extrajo los datos de cada estudio primario, se consolidó y tabuló la información correspondiente a cada una de las cuatro preguntas de investigación. Las tablas obtenidas fueron utilizadas para facilitar el análisis y síntesis de la información en vistas a responder cada una de las preguntas de investigación planteadas.

3.4. ESTRATEGIA DE SÍNTESIS DE DATOS

Tabla 3.5: Campos del formulario de extracción de datos

Campo	Descripción	RQ
Id	E[número]. P.ej: E001	General
Fecha de extracción		General
Autores		General
Título		General
Tipo de fuente	Revista, congreso o capítulo de libro.	General
Fuente	Nombre de la revista, congreso o libro.	General
Año de publicación		General
Afiliación	Instituciones de afiliación de los investigadores.	General
País	País(es) de afiliación de los investigadores.	General
Nivel	Qué nivel de análisis de sentimiento está siendo apoyado con el uso de ontologías de dominio.	RQ1.1
Fase(s)	Qué fases del análisis de sentimiento están siendo apoyadas con el uso de ontologías de dominio.	RQ1.2
Problemas abordados	Qué problemas en el análisis de sentimiento están siendo abordados con el apoyo de ontologías de dominio.	RQ1.3
Comparaciones reportadas	Con qué otras aproximaciones se compara la propuesta del estudio.	RQ1.3
Beneficios reportados	Qué beneficios han sido reportados al aplicar las ontologías de dominio para la solución de problemas en el análisis de sentimiento.	RQ1.3
Generalizable a otros dominios	Si las investigaciones reportadas son generalizables a dominios diferentes al de las ontologías aplicadas durante la experimentación. (Sí/No)	RQ2
Dominio de las ontologías	A qué dominios pertenecen las ontologías de dominio aplicadas.	RQ2
Se reutiliza ontologías existentes	Sí/No.	RQ2
Ontologías reutilizadas	Nombre de las ontologías existentes reutilizadas.	RQ2
Se construye ontologías reutilizables	Sí/No.	RQ2
Ontologías construidas	Nombre de las ontologías reutilizables construidas.	RQ2
Modela atributos	Si los conceptos e instancias de la ontología tienen asociados atributos. (Sí/No)	RQ3.1
Modela relaciones taxonómicas	Si los conceptos e instancias de la ontología están asociados con relaciones taxonómicas. (Sí/No)	RQ3.1
Modela relaciones ad-hoc	Si los conceptos e instancias de la ontología están asociados con otros tipos de relaciones ad-hoc. (Sí/No)	RQ3.1
Modela axiomas formales	Si la ontología define axiomas formales. (Sí/No)	RQ3.1
Modela reglas	Si la ontología define reglas que puedan ser usadas en inferencias. (Sí/No)	RQ3.1
Realiza inferencias	Si se realiza inferencias con la ontología. (Sí/No)	RQ3.2
Razonadores utilizados	Qué razonadores son usados para realizar las inferencias.	RQ3.2
Modela emociones	Si se representa emociones en la ontología. (Sí/No)	RQ4
Modelo psicológico asumido	Qué modelo psicológico de clasificación de las emociones es empleado.	RQ4.1
Tipo de representación emocional	Categorial, dimensional o componencial.	RQ4.2

## Capítulo 4

# Ejecución de la Revisión Sistemática de la Literatura

### 4.1. Búsqueda y selección de estudios primarios

Se ejecutó el protocolo de revisión definido. Se hizo uso de la aplicación StArt (*State of the Art through Systematic Reviews*)<sup>1</sup> para dar soporte a la ejecución del protocolo de revisión. StArt ha sido evaluada empíricamente, verificándose que es una ayuda efectiva en la realización de Revisiones Sistemáticas de la Literatura, facilitando la aplicación de sus procesos [52].

La búsqueda en las bases de datos seleccionadas se realizó entre los días 14 y 27 de julio de 2015. Una búsqueda complementaria en la base de datos Scopus se realizó el día 30 de setiembre de 2015, durante la fase de selección final, luego de verificarse que haber restringido las búsquedas al área de Ciencias de la Computación había excluido de los resultados varias publicaciones identificadas en la etapa de elaboración del protocolo de revisión y que aparecían citadas en las publicaciones que venían siendo aceptadas. Se verificó que ello se debía a que tales publicaciones estaban clasificadas erróneamente en la base de datos Scopus. La búsqueda complementaria repitió la cadena inicial pero esta vez poniendo como restricción que las publicaciones no hubieran sido clasificadas en el área de Ciencias de la Computación.

En la figura 4.1 se muestra el proceso de selección y el número de publicaciones identificadas en cada fase. En total, las búsquedas dieron como resultado 1445 publicaciones. Se excluyó 300 publicaciones duplicadas, la mayoría de las cuales fueron identificadas por la aplicación StArt. Después de

---

<sup>1</sup>[http://lapes.dc.ufscar.br/tools/start\\_tool](http://lapes.dc.ufscar.br/tools/start_tool)

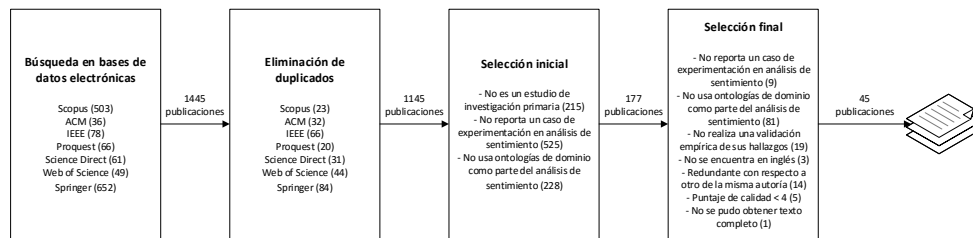


Figura 4.1: Fases del proceso de selección

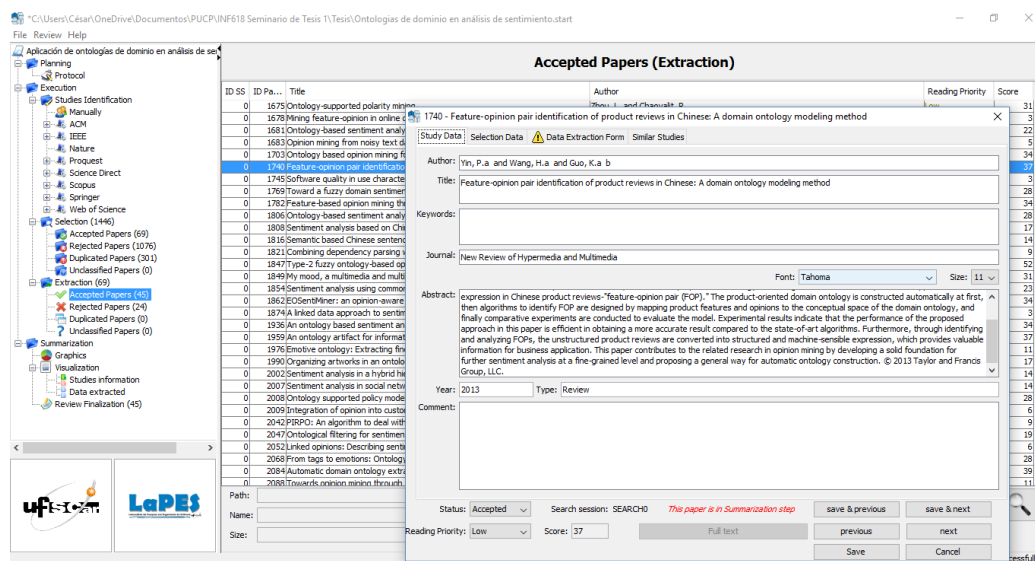


Figura 4.2: Detalle de formulario de extracción en la herramienta StArt, utilizada como soporte para la ejecución del protocolo de revisión

la eliminación de duplicados se quedó con 1145 publicaciones.

A continuación se procedió a la etapa de selección inicial, revisándose los títulos y sumarios de las publicaciones y eliminándose aquellas que claramente no cumplían con los criterios de inclusión y exclusión. Como resultado de la selección inicial se excluyó 968 publicaciones, quedando 177 publicaciones que pasaron a la fase de selección final.

En la fase de selección final se recuperó el texto completo de las 177 publicaciones y se aplicó los criterios de inclusión y exclusión, excluyéndose 126 publicaciones. Asimismo, las 51 publicaciones restantes fueron evaluadas según los criterios de calidad y se excluyó 5 publicaciones que recibieron un puntaje de calidad menor a 4. Una publicación ([112]) fue excluida por no estar disponible en fuentes en línea y por no haber recibido respuesta de los autores. Finalmente, quedaron sólo 45 estudios primarios, los cuales son listados en la tabla 4.1 con los respectivos años de publicación y puntajes de

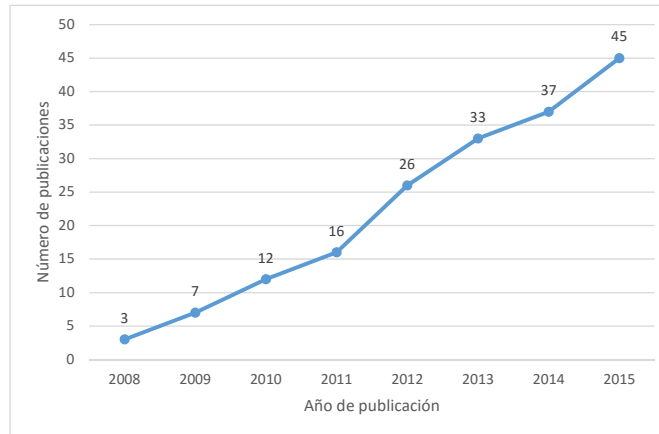


Figura 4.3: Número acumulado de publicaciones por año

calidad.

En la tabla 4.2 se sintetiza el resultado final de las búsquedas realizadas en las bases de datos electrónicas.

## 4.2. Extracción y síntesis de datos

Se leyó detenidamente las 45 publicaciones y se extrajo los datos relevantes con la ayuda de la herramienta StArt según el formulario establecido en el protocolo de revisión (ver tabla 3.5 y figura 4.2). En la fase de síntesis de datos se analizó los datos para identificar criterios de agrupamiento. Los resultados de la síntesis se describen a continuación.

## 4.3. Visión general de los estudios incluidos

### 4.3.1. Años de publicación

En la figura 4.3 se presenta el número acumulado de publicaciones seleccionadas. No se encontró publicaciones previas a 2008. Desde entonces, el número acumulado de publicaciones se ha incrementado todos los años. El año de mayor crecimiento fue 2012 (10 publicaciones). Los siguientes años (2013 y 2014) tuvieron un crecimiento de 7 y 4 publicaciones, respectivamente. Al momento de las búsquedas se registró 8 publicaciones relevantes para el año 2015.



Tabla 4.1: Publicaciones incluidas

ID	Año	Autor	Puntaje de calidad	País(es) de afiliación
S01	2008	Cheng et al. [26]	6.5	Alemania
S02	2008	Softic et al. [105]	7.0	Austria
S03	2008	Zhou et al. [126]	8.0	EE.UU.
S04	2009	Brooke et al. [16]	6.5	Canadá, EE.UU.
S05	2009	Dey et al. [32]	7.5	India
S06	2009	Lau et al. [63]	7.5	Australia, China
S07	2009	Zhao et al. [125]	7.0	China
S08	2010	Chen et al. [25]	7.0	China
S09	2010	Ding et al. [33]	6.0	China
S10	2010	García-Crespo et al. [47]	8.0	España
S11	2010	Somprasertsri et al. [106]	7.0	Tailandia
S12	2010	Wei et al. [115]	8.0	Noruega
S13	2011	Baldoni et al. [6]	5.5	Italia
S14	2011	Grassi et al. [50]	7.5	Italia, Reino Unido
S15	2011	Westerski et al. [118]	7.5	España
S16	2011	Yuan et al. [124]	5.5	China
S17	2012	Baracho et al. [7]	6.5	Brasil
S18	2012	Chaves et al. [23]	6.0	Brasil, Portugal
S19	2012	Colace et al. [27]	6.5	EE.UU., Italia
S20	2012	Husaini et al. [55]	7.5	Hungría, Turquía
S21	2012	Leopairote et al. [65]	5.5	Tailandia
S22	2012	Li et al. [67]	5.0	China
S23	2012	Liu et al. [71]	6.5	China
S24	2012	Wei et al. [116]	8.0	Noruega
S25	2012	Yaakub et al. [120]	6.0	Australia
S26	2012	Yang et al. [121]	6.5	China
S27	2013	Bertola et al. [11]	4.5	Italia
S28	2013	Kontopoulos et al. [59]	8.0	Grecia
S29	2013	Park et al. [81]	6.0	EE.UU.
S30	2013	Ramteke et al. [92]	8.0	India
S31	2013	Sam et al. [95]	4.5	China, Reino Unido
S32	2013	Sykora et al. [108]	6.0	Reino Unido
S33	2013	Yin et al. [122]	7.5	China
S34	2014	Nithish et al. [78]	5.5	India
S35	2014	Peñalver-Martínez et al. [83]	8.0	España
S36	2014	Sánchez-Rada et al. [97]	7.5	España
S37	2014	Wang et al. [114]	6.0	China
S38	2015	Agarwal et al. [1]	6.0	India
S39	2015	Ali et al. [3]	7.0	India
S40	2015	Cotfas et al. [29]	5.5	Francia, Rumania
S41	2015	Leotta et al. [66]	7.0	Italia
S42	2015	Liu et al. [70]	5.5	China
S43	2015	Reforgiato et al. [93]	7.5	Italia
S44	2015	Shi et al. [103]	7.5	China, EE.UU.
S45	2015	Teja Santosh et al. [110]	5.0	India

#### 4.3. VISIÓN GENERAL DE LOS ESTUDIOS INCLUIDOS

29

Tabla 4.2: Resultados de búsqueda por bases de datos

Fuente	Resultados	Duplicados	Artículos relevantes
Scopus	503	23	35
ACM	36	32	1
IEEE	78	66	1
Proquest	66	20	2
Science Direct	61	31	0
Web of Science	49	44	1
Springer	652	84	5

#### 4.3.2. Fuentes de publicación

En la tabla 4.3 se presenta las fuentes de publicación de los 46 estudios seleccionados, junto con el número de estudios correspondientes a cada fuente. De un total de 39 fuentes, 14 son revistas científicas y 25 son actas de conferencias. La distribución es marcadamente dispersa, siendo 2 el número máximo de publicaciones por fuente.

#### 4.3.3. Centros de investigación

Para identificar los centros de investigación más activos, se examinó los datos de afiliación de los autores de las publicaciones. En las 45 publicaciones se identificó un total de 58 instituciones de afiliación diferentes, con una distribución marcadamente dispersa, en la que el número máximo de publicaciones por institución es de 2. En la tabla 4.4 se muestra las cinco instituciones de investigación con más de una publicación sobre aplicación de ontologías de dominio en análisis de sentimiento.

De un total de 65 instituciones de afiliación, 61 son académicas y sólo 4 son comerciales (S01, S04, S05, S36).

En la figura 4.4 se muestra la distribución de estudios por el país de afiliación. De un total de 22 países, China es el país con más publicaciones (17), seguido de la India (7), Italia y España (6).

#### 4.3.4. Número de citas

Para estimar el impacto de los estudios en la literatura académica se tomó como referencia el número de citas recibidas por los estudios según Google Scholar<sup>2</sup>. Se debe tener que el número real de citas podría ser distinto y es un dato en continua variación, por lo que es meramente referencial. Los estudios más recientes han recibido comprensiblemente un número menor de citas. El número de citas recibidas por los 45 estudios varía desde 0 hasta 90, con un

<sup>2</sup><http://scholar.google.com/>, fecha de acceso 28/11/2015.

Tabla 4.3: Distribución de estudios por fuentes de publicación

Tipo	Fuente	Total
Revista	Cognitive Computation	2
Revista	Expert Systems with Applications	2
Conferencia	Annual Meeting of the Association for Computational Linguistics (ACL)	2
Conferencia	International Conference on Information Systems (ICIS)	2
Conferencia	International Workshop on Emotion and Sentiment in Social and Expressive Media (ESSEM)	2
Conferencia	International Workshop on Social Data on the Web (SDoW)	2
Revista	Applied Intelligence	1
Revista	Computational Intelligence and Neuroscience	1
Revista	International Journal of Digital Content Technology and its Applications	1
Revista	International Journal of Technology	1
Revista	International Journal of e-Education, e-Business, e-Management and e-Learning	1
Revista	International Journal on Document Analysis and Recognition	1
Revista	Journal of Experimental and Theoretical Artificial Intelligence	1
Revista	Journal of Information Technology	1
Revista	Journal of Shanghai Jiaotong University (Science)	1
Revista	Journal of Universal Computer Science	1
Revista	Journal of the American Society for Information Science and Technology	1
Revista	New Review of Hypermedia and Multimedia	1
Conferencia	Asian Conference on Intelligent Information and Database Systems (ACIIDS)	1
Conferencia	CCF Conference on Natural Language Processing and Chinese Computing (NLPCC)	1
Conferencia	Communications in Computer and Information Science	1
Conferencia	IADIS International Conference on Intelligent Systems and Agents	1
Conferencia	International CIKM Workshop on Topic-sentiment Analysis for Mass Opinion	1
Conferencia	International Conference on Advanced Computing (ICoAC)	1
Conferencia	International Conference on Applications of Natural Language to Information Systems (NLDB)	1
Conferencia	International Conference on Asian Language Processing (IALP)	1
Conferencia	International Conference on Digital Information Management (ICDIM)	1
Conferencia	International Conference on Digital Information and Communication Technology and its Applications (DICTAP)	1
Conferencia	International Conference on Distributed Multimedia Systems (DMS)	1
Conferencia	International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)	1
Conferencia	International Conference on Knowledge Science, Engineering and Management (KSEM)	1
Conferencia	International Conference on Language Resources and Evaluation (LREC)	1
Conferencia	International Congress on Image and Signal Processing (CISP)	1
Conferencia	International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)	1
Conferencia	International Universal Communication Symposium (IUCS)	1
Conferencia	International Workshop on Finance and Economics on the Semantic Web (FEOSW)	1
Conferencia	International Workshop on Metamodels, Ontologies and Semantic Technologies (MOST)	1
Conferencia	International Workshop on New Challenges in Distributed Information Filtering and Retrieval (DART)	1
Conferencia	International Workshops on On the Move to Meaningful Internet Systems (OTM)	1
	<b>Total</b>	<b>45</b>

## 4.3. VISION GENERAL DE LOS ESTUDIOS INCLUIDOS

31

Tabla 4.4: Distribución de estudios por instituto de afiliación de los autores

Institución de Afiliación	Conteo de estudios
Capital Normal University, China	2
Norwegian University of Science and Technology, Noruega	2
Queensland University of Technology, Australia	2
Tongji University, China	2
Universidad Carlos III de Madrid, España	2
Universidad Politécnica de Madrid, España	2
Università degli Studi di Torino, Italia	2

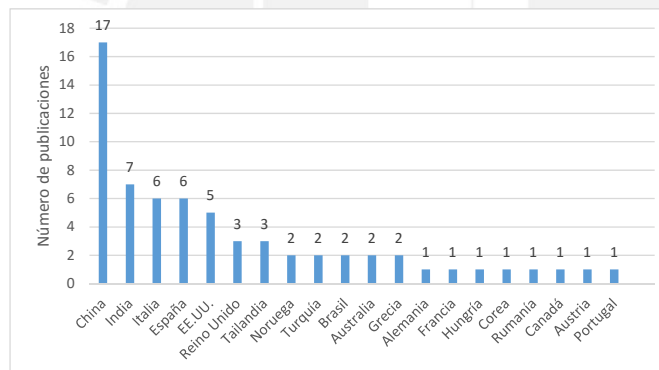


Figura 4.4: Estudios por país de afiliación

Tabla 4.5: Número de citas recibidas por los estudios seleccionados

Citado por	0-10	11-30	31-50	51-70	71-90
Número de estudios	30	5	4	2	4

Tabla 4.6: Estudios más citados

Estudio	Autor y título	Total de citas (Google Scholar)	Citas por otros estudios de la revisión
S05	Dey et al. Opinion mining from noisy text data	90	0
S11	Somprasertsri et al. Mining feature-opinion in online customer reviews for opinion summarization	81	1
S28	Kontopoulos et al. Ontology-based sentiment analysis of twitter posts	73	4
S03	Zhou et al. Ontology-supported polarity mining	71	8
S12	Wei et al. Sentiment learning on product reviews via Sentiment Ontology Tree	65	5
S10	García-Crespo et al. SEMO: a framework for customer social networks analysis based on semantics	63	1
S07	Zhao et al. Ontology based opinion mining for movie reviews	45	4
S14	Grassi et al. Sentic Web: A New Paradigm for Managing Social Media Affective Information	43	0
S13	Baldoni et al. From tags to emotions: Ontology-driven sentiment analysis in the social Semantic Web	33	6
S06	Lau et al. Automatic domain ontology extraction for context-sensitive opinion mining	31	3

promedio de 17, primer cuartil 2, mediana 4 y tercer cuartil 16. En la tabla 4.5 se presenta el detalle de la tasa de cita recibidas. Sólo 10 de los 45 estudios han recibido un número de citas mayor a 30. De ellos, sólo 4 han recibido un número muy alto de citas (más de 70 citas). En la tabla 4.6 se presenta los diez estudios con más de 20 citas recibidas. Como referencia para observar su impacto en el conjunto de estudios seleccionados, se indica asimismo el número de citas recibidas por otros de los 45 estudios seleccionados.

## Capítulo 5

# Resultados de la Revisión Sistemática de la Literatura

### 5.1. Contribución al análisis de sentimiento (RQ1)

*¿Cómo está contribuyendo el uso de ontologías de dominio a la solución de problemas específicos en el análisis de sentimiento?*

La primera pregunta de investigación tiene por objetivo identificar cómo está contribuyendo el uso de ontologías de dominio a la solución de problemas específicos en el análisis de sentimiento? Ha sido dividida en cuatro subpreguntas en las que se identifica respectivamente los niveles, las fases y los problemas del análisis del sentimiento en los que se ha aplicado ontologías de dominio, así como los beneficios reportados.

#### 5.1.1. Niveles de análisis de sentimiento (RQ1.1)

*¿Qué nivel de análisis de sentimiento está siendo apoyado con el uso de ontologías de dominio?*

#### Resultados

Esta pregunta de investigación tiene por fin identificar los niveles de granularidad de análisis de sentimiento que están siendo apoyados con la aplicación de ontologías de dominio. La categorización de los niveles fue tomada de [60, 115, 20] y está descrita en 1.1: a nivel de documento, oración, palabra, frase, aspecto/característica y concepto.

Tabla 5.1: Aplicación de ontologías de dominio en los diferentes niveles de granularidad de análisis de sentimiento

Nivel	Estudios	Conteo	%
Documento	S02, S03, S04, S06, S07, S10, S13, S14, S19, S22, S27, S29, S30, S31, S35, S36, S38, S40, S41	19	42.2 %
Oración	S26, S32, S37, S43, S44	5	11.1 %
Palabra	S06, S16, S26, S32, S39, S40, S41, S42, S44	9	20.0 %
Frase	S43	1	2.2 %
Aspecto/Característica	S01, S05, S07, S08, S09, S11, S12, S15, S17, S18, S20, S21, S23, S24, S25, S28, S33, S34, S39, S40, S42, S43, S45	23	51.1 %
Concepto	S14	1	2.2 %

En la tabla 5.1 se presenta el detalle de los estudios por nivel de granularidad de análisis de sentimiento en la aplicación de ontologías de dominio. Dado que un mismo estudio puede aplicar a más de un nivel de granularidad, la suma de conteos o porcentajes es mayor a la totalidad de estudios.

Los niveles más observados son: *aspecto/característica* (23 estudios, 51.1 %) y *documento* (19 estudios, 42.2 %). Los demás niveles tiene una aplicación marcadamente menor: *palabra* (9 estudios, 20 %), *oración* (5 estudios, 11.1 %), *frase* (1 estudio, 2.2 %) y *concepto* (1 estudio, 2.2 %).

### Análisis y discusión

El nivel *aspecto/característica* es el más representado en los estudios, con una presencia de más del 50 %. Este resultado es consistente con la investigación preliminar y en cierta medida esperado, pues las ontologías de dominio son muy apropiadas para representar el conocimiento sobre las diversas características o aspectos de los productos o servicios sobre los cuales se realiza el análisis de sentimiento.

En segundo lugar de presencia, también con un número significativo de estudios, está el nivel *documento*. Este es el nivel de menor granularidad, en el que se calcula el sentimiento general de todo el documento considerado como una única opinión sobre un único objeto. Sólo dos estudios (S07 y S40) aplican las ontologías de dominio simultáneamente a nivel de *aspecto/característica* y a nivel de *documento*. En S07 [125] se usa la ontología para calcular la polaridad con respecto a cada característica y luego para derivar la polaridad general de un documento como una transferencia desde los nodos de nivel más bajo al de nivel más alto en la ontología. En S40 [29] se emplea ontologías de dominio tanto para recuperar tweets sobre cada uno de los atributos de un producto como para identificar las emociones presentes en cada tweet.

Una mayor granularidad a nivel de *oración* se da en sólo cinco estudios.

## 5.1. CONTRIBUCIÓN AL ANÁLISIS DE SENTIMIENTO (RQ1) 35

Cuatro de ellos (S26 [121], S32 [108], S37 [114] y S44 [103]) aplican una ontología de dominio para identificar las emociones expresadas en cada oración de los documentos analizados. En S43 [93] se aplica una ontología de dominio para representar las opiniones y determinar la polaridad a la vez a nivel de aspecto/característica y de oración tomando en cuenta las dependencias entre los diferentes elementos de la opinión.

Los anteriores niveles de granularidad suelen depender de una anotación del sentimiento a nivel de una unidad mínima de análisis que puede ser una palabra, una frase o un concepto. El más sencillo de estos es el nivel de *palabra*, que está representado por nueve estudios. Cinco de ellos (S40 [29], S26 [121], S32 [108], S39 [3] y S44 [103]) usan una ontología de dominio para identificar la emoción asociada a cada palabra. Otros tres (S06 [63], S16 [124], S42 [70]) se apoyan en una ontología de dominio para determinar la polaridad de las palabras cuyo sentido varía dependiendo del dominio o del objeto o aspecto/característica al cual se aplican. Finalmente, en S41 [66] se usa dos ontologías de dominio, una que representa documentos negativos y una que representa documentos positivos, para identificar la polaridad de los documentos a partir de las palabras que los conforman.

Un estudio reciente (S43 [93]) realiza un análisis a nivel de *frase* con el apoyo de ontologías de dominio tomando en cuenta las dependencias semánticas entre los diferentes elementos de la opinión.

La granularidad a nivel de *concepto* sólo está presente en un estudio (S14 [50]) en el que se toma como unidad de análisis los nodos de ConceptNet identificados<sup>1</sup> y se usa una ontología de dominio para inferir la categoría emocional de cada uno a partir de vectores de cuatro dimensiones emocionales.

Los resultados permiten verificar que todos los niveles de granularidad de análisis de sentimiento han sido apoyados con la aplicación de ontologías de dominio. Si bien el número de estudios correspondientes es muy diferente entre unos niveles y otros, se puede concluir que la aplicación de ontologías de dominio no depende del nivel de granularidad aplicada.

### 5.1.2. Fases del análisis de sentimiento (RQ1.2)

*¿Qué fases del análisis de sentimiento están siendo apoyadas con el uso de ontologías de dominio?*

---

<sup>1</sup>ConceptNet [51] es una representación a manera de grafo dirigido del corpus de información recolectada por el proyecto Open Mind Common Sense, establecido desde el año 2000 para recopilar la información que denominamos sentido común, esto es, cosas obvias que la gente normalmente sabe y es por ello usualmente implícita, como una red de relaciones entre conceptos.



Tabla 5.2: Aplicación de ontologías de dominio en las diferentes fases del análisis de sentimiento

Fase	Estudios	Conteo	%
Clasificación de la subjetividad	S15, S43	2	4.4 %
Clasificación del sentimiento	S03, S04, S06, S07, S10, S12, S13, S14, S15, S16, S19, S22, S23, S24, S26, S27, S29, S30, S32, S37, S38, S39, S40, S41, S42, S43, S44, S45	28	62.2 %
Extracción del sujeto de la opinión	S15, S43, S45	3	6.7 %
Extracción del objeto y la característica/aspecto	S01, S02, S03, S04, S06, S07, S08, S09, S11, S12, S15, S17, S18, S20, S21, S23, S24, S25, S33, S35, S36, S39, S42, S43, S45	25	55.6 %
Sumarización	S05, S17, S20, S21, S27, S33, S34	7	15.6 %

Tabla 5.3: Número de fases del análisis de sentimiento en las que los estudios revisados aplican ontologías de dominio

Número de fases	Conteo de estudios
0	2
1	26
2	14
3	1
4	2

## Resultados

Esta pregunta de investigación tiene por fin identificar las fases del análisis de sentimiento que fueron apoyadas por ontologías de dominio en los estudios revisados. La categorización de las fases fue tomada de [60] y está descrita en 1.1.

En la tabla 5.2 se presenta el detalle de los estudios según la fase del análisis de sentimiento en la que se aplicó ontologías de dominio. La suma de conteos o porcentajes es mayor a la totalidad de estudios debido a que en un estudio se puede hacer la aplicación en una, más de una o ninguna de las fases.

Las fases en las que más se observa la aplicación de ontologías de dominio son: *Clasificación del sentimiento* (28 estudios, 62.2 %) y *Extracción del objeto y la característica/aspecto* (25 estudios, 55.6 %). La aplicación durante las demás fases es comparativamente menor: *Sumarización* (7 estudios, 15.6 %), *Extracción del sujeto de la opinión* (3 estudios, 6.7 %) y *Clasificación de la subjetividad* (2 estudio, 4.4 %).

## 5.1. CONTRIBUCIÓN AL ANÁLISIS DE SENTIMIENTO (RQ1) 37

### Análisis y discusión

En la tabla 5.3 se presenta el número de fases del análisis de sentimiento en las que los estudios revisados aplican ontologías de dominio. La mayoría de estudios (88.9 %) realizan la aplicación durante una o dos fases. Ninguno lo hace durante las cinco fases.

Dos estudios (S28 [59] y S31 [95]) no reportan ninguna fase. En ambos casos se siguió un proceso diverso al tomado como referencia.

Otros dos estudios (S15 [118] y S43 [93]) realizan la aplicación en cuatro de las cinco fases. En ambos casos ésta consiste en representar estructuralmente los diversos elementos de las opiniones e irlos vinculando con otros conjuntos de datos (*linked data*).

La aplicación de ontologías de dominio durante la fase de *Clasificación del sentimiento* es la más frecuente (62.2 %, 28 estudios), aunque el tipo de aplicación es variado en los distintos estudios. En algunos casos las ontologías de dominio apoyan la clasificación a nivel de emociones, en otros apoyan la clasificación de las palabras cuya polaridad varía según el contexto, en otros apoyan el cálculo de polaridades generales a partir de polaridades calculadas para aspectos/características particulares, en otros apoyan una representación estructurada de las opiniones. El detalle de las diversas aplicaciones se verá en la siguiente sección (5.1.3).

La aplicación durante la fase de *Extracción del objeto y la característica/aspecto* es también muy frecuente (55.6 %, 25 estudios). Ello es esperado dado que las ontologías de dominio son muy apropiadas para la representación de los diversos tipos de objetos y sus características. Por este mismo motivo son también idóneas para los casos en que se realiza una sumarización de las opiniones (15.6 %, 7 estudios), usualmente identificando el sentimiento en relación a cada uno de los aspectos o características del objeto de opinión.

El apoyo en las fases de *Clasificación de la subjetividad* (S15 [118] y S43 [93]) y *Extracción del sujeto de la opinión* (S15 [118], S43 [93] y S45 [110]) tiene sólo el alcance de facilitar la representación estructurada de las opiniones y su vinculación con otros conjuntos de datos (*linked data*).

### 5.1.3. Problemas del análisis de sentimiento y beneficios reportados (RQ1.3)

*¿Qué problemas en el análisis de sentimiento están siendo abordados con el apoyo de ontologías de dominio y qué beneficios han sido reportados con su aplicación?*

Tabla 5.4: Aplicación de ontologías de dominio a problemas del análisis de sentimiento

Id	Problema	Estudios	Conteo	%
P1	Representación de las opiniones	S02, S10, S15, S29, S36, S40, S43, S45	8	17.8%
P2	Análisis de discurso	S04	1	2.2%
P3	Clasificación de la polaridad	S03, S06, S07, S12, S16, S19, S22, S23, S24, S38, S41, S42, S44	13	28.9%
P4	Identificación de expectativas frustradas (thwarting)	S30	1	2.2%
P5	Clasificación de emociones	S10, S13, S14, S26, S27, S32, S37, S40, S44	9	20.0%
P6	Identificación de tópicos/atributos	S01, S03, S07, S09, S12, S17, S18, S20, S21, S23, S24, S25, S35, S38, S39, S42	16	35.6%
P7	Identificación de pares tópico-sentimiento	S08, S11, S33, S39, S42	5	11.1%
P8	Recuperación de información	S02, S07, S15, S28, S31, S34, S36, S39, S40	9	20.0%
P9	Sumarización	S05, S17, S20, S21, S27, S33, S34, S37	8	17.8%

## Resultados

Esta pregunta de investigación tiene por fin identificar los problemas del análisis de sentimiento que están siendo abordados con el apoyo de ontologías de dominio, así como los beneficios reportados con su aplicación.

La categorización de los problemas fue realizada luego de la extracción, como parte de la fase de síntesis. En la tabla 5.4 se presenta el detalle de los 9 problemas del análisis de sentimiento en los que se aplicó ontologías de dominio. La suma de conteos o porcentajes es mayor a la totalidad de estudios debido a que en un estudio se puede aplicar ontologías de dominio para más de un problema del análisis de sentimiento.

Los problemas en los que más se observa la aplicación de ontologías de dominio son: *P6: Identificación de tópicos/atributos* (16 estudios, 35.6%) y *P3: Clasificación de la polaridad* (13 estudios, 28.9%). Dos problemas fueron identificados cada uno en un solo estudio (2.2%): *P2: Análisis de discurso* y *P4: Identificación de expectativas frustradas (thwarting)*.

En cada uno de los casos se tabuló la información reportada acerca de los beneficios obtenidos con la aplicación de ontologías de dominio. En los estudios con resultados comparables a los reportados en investigaciones previas, se registró además las correspondientes referencias bibliográficas. Sólo 14 estudios (31.1%) reportan resultados comparables con estudios anteriores: S01, S06, S08, S11, S12, S23, S24, S25, S26, S27, S32, S33, S38 y S40.

En la tabla 5.5 se presenta una síntesis de los beneficios reportados por los distintos estudios, así como las referencias de los estudios con cuyos resultados se realiza una comparación directa.

## 5.1. CONTRIBUCIÓN AL ANÁLISIS DE SENTIMIENTO (RQ1)

39

Tabla 5.5: Beneficios de la aplicación de ontologías de dominio a problemas del análisis de sentimiento

Id	Problema	Beneficio	Estudios	Estudios con los que se compara
P1	Representación de las opiniones	Representación estructurada de las opiniones	S02, S10, S15, S29, S36, S40, S43, S45	[105]
		Vinculación con otros conjuntos de datos ( <i>linked data</i> )	S02, S15, S36, S43, S45	[105]
P2	Análisis de discurso	Soporte para análisis de patrones de cambio de polaridad y profundidad de tópicos	S04	
P3	Clasificación de la polaridad	Mayor precisión y exhaustividad	S03, S06, S07, S16, S23, S38	[54, 90, 35, 115, 76, 117, 109]
		Clasificación de textos sobre más de un producto a la vez	S12, S24	[22, 115]
		Clasificación de términos de sentimiento variable según contexto	S16, S42	
		Ponderación del grado de relevancia de opiniones individuales	S22, S38	
		Soporte para análisis multilingüe	S41	
P4	Identificación de expectativas frustradas ( <i>thwarting</i> )	Soporte para reconocer documentos <i>thwarted</i>	S30	
P5	Clasificación de emociones	Representación de modelo emocional	S10, S13, S14, S26, S27, S37, S40, S44	[113]
		Calcular emociones predominantes por medio de inferencias semánticas	S13	
		Interoperabilidad de modelos emocionales	S14	
		Comparación y combinación de emociones	S27	
		Soporte para visualización de emociones	S14, S27	
		Mayor precisión	S32	[31]
P6	Identificación de tópicos/atributos	Mayor precisión y exhaustividad	S01, S09, S25	[90, 98, 54]
P7	Identificación de pares tópico-sentimiento	Mayor precisión y exhaustividad	S08, S11, S33	[69, 72, 54, 90, 91, 125, 84, 62, 63]
		Normalización del lenguaje	S11	
		Identificación en caso de tópicos implícitos	S33	[54, 91, 125, 84, 62, 63]
P8	Recuperación de información	Mayor precisión y exhaustividad	S02, S15, S28, S39	
P9	Sumarización	Sumarización con diversos niveles de especificidad	S05, S17, S21	
		Sumarización de emociones	S27	

### Análisis y discusión

**P1: Representación de las opiniones** Un 17.8 % de los estudios utiliza una ontología de dominio para modelar la representación de las opiniones o emociones identificadas en los documentos objeto del análisis de sentimiento.

En S02 [105] se emplea una ontología sobre minería de opinión, la cual actúa como nexo entre el dominio de aplicación y los datos a analizar. Con ella se representa formalmente las opiniones en formato RDF y se las vincula con entidades definidas en otros conjuntos de datos (*linked data*). El modelo propuesto es así independiente del origen de los datos y del dominio de aplicación.

En S10 [47] se usa la *Customer Emotion Ontology (CEO)* [46] tanto como para representar productos y atributos como para representar el modelo emocional y las emociones asociadas a cada producto.

En S15 [118] se propone la ontología MARL para modelar el conocimiento sobre Minería de Opinión. En relación a [105] añade nuevas propiedades que permiten asignar valores numéricos a la polaridad. Asimismo, la vinculación a otros conjuntos de datos tales como PROV y DBpedia hace posible realizar búsquedas *Internet-wide* y comparar o agregar opiniones de diversas fuentes.

En S29 [81], con una aproximación basada en la ciencia del diseño, se propone la *Information Systems Sentiment Ontology*, una ontología que modela el análisis del sentimiento a partir del modelo componencial de las emociones. En particular, se modela la percepción de los usuarios y las emociones asociadas, y se predice la intención conductual (compra del producto o cambio por otro).

En S36 [97], gracias al uso de Marl [118] como ontología de minería de opinión, de Onyx [96] como ontología de emociones y de *Financial Industry Business Ontology (FIBO)* [9] como ontología del dominio financiero, se hace posible el análisis conjunto de opiniones expresadas en diferentes medios sociales.

En S40 [29] se aplica tres ontologías de dominio: una, llamada Twitter, que modela tweets; otra que basada en OntoEmotion [41] representa categorías emocionales; y otra, denominada TweetOntoSense, que modela el análisis de sentimiento y vincula tweets con emociones específicas sobre entidades mencionadas en ellos.

En S43 [93] se presenta OntoSentilo, una aproximación que usa la ontología del mismo nombre para soportar una representación formal en RDF de todos los elementos la opinión: sujeto de la opinión, contexto, tópico, subtópico, disparador, característica, y las relaciones entre ellos, así como resolver dichos elementos como *linked data* con alto grado de precisión.

En S45 [110] el uso de la ontología Marl [118] hace posible obtener efi-

## 5.1. CONTRIBUCIÓN AL ANÁLISIS DE SENTIMIENTO (RQ1)

41

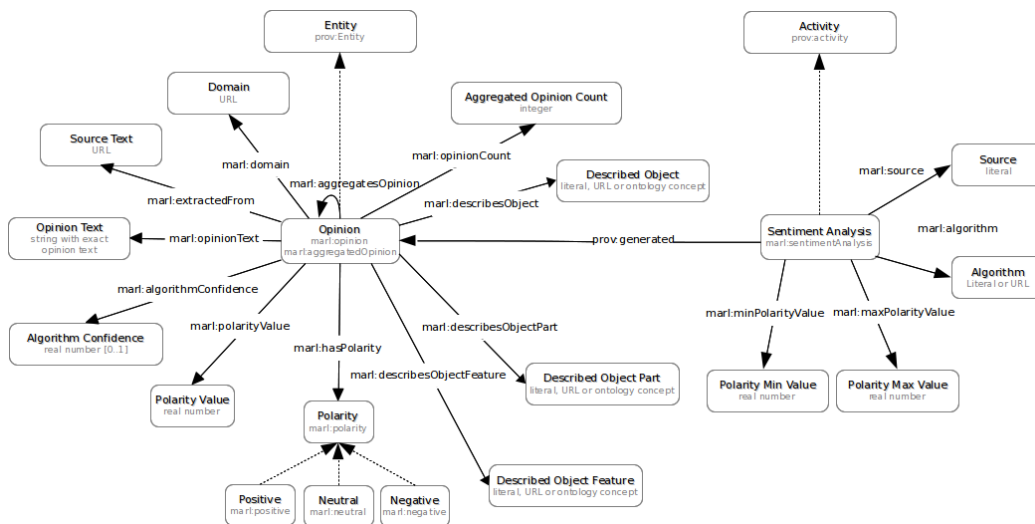


Figura 5.1: Ontología Marl versión 1.0 [118]

cientemente datos acerca de opiniones en forma de linked data.

**Beneficios** *Representación estructurada de las opiniones.* Como se ve en la tabla 5.5, un beneficio inmediato de la aplicación de ontologías de dominio en relación a P1 es la capacidad de las ontologías para modelar de manera estructurada y semánticamente rica los diversos elementos de la opinión, sus atributos y las relaciones entre ellos, e incluso el procedimiento mismo de análisis de sentimiento. Con ello, el modelo de análisis es independiente del origen de los datos y del dominio de aplicación. Como ilustración de ello, en la figura 5.1 se presenta la versión 1.0 de la ontología Marl, introducida en S15 [118] y reutilizada en S36 [97] y S45 [110]. Marl tiene soporte para representar el análisis de sentimiento a nivel de aspecto/característica, considera tanto polaridades categóricas (positivo/neutral/negativo) como polaridades continuas con valores reales y apoya también la agregación de opiniones individuales.

*Vinculación con otros conjuntos de datos (linked data).* La representación de las opiniones con el apoyo de ontologías de dominio hace también posible la integración con otros conjuntos de datos a manera de *linked data*. Por ejemplo, en S43 [93], siempre que es posible, se resuelve como *linked data* tanto los sujetos de opinión como la identidad de los tópicos de opinión en recursos tales como DBpedia.

**P2: Análisis de discurso** Se reportó un caso (2.2%) en el que con el apoyo de ontologías de dominio se realizó un análisis de patrones de discurso en documentos de opinión.

En S04 [16] se utiliza una ontología del dominio de los videojuegos para analizar los patrones de cambio de polaridad y profundidad de tópicos en un conjunto de reseñas de usuarios.

**Beneficios** *Soporte para analizar los patrones de cambio de polaridad y profundidad de tópicos.* En el caso citado, con el apoyo de la ontología se identificó patrones que relacionan tópicos y polaridad:

- Ciertos tópicos tienen mayor sesgo hacia una polaridad u otra.
- Ciertos tópicos son mejores predictores de la polaridad general.
- Hay una correlación negativa entre la profundidad taxonómica del tópico y su grado de relevancia para la polaridad general.

Los resultados de análisis de este tipo son valiosos para la optimización de los procesos de clasificación de la polaridad.

**P3: Clasificación de la polaridad** Un 31.1% de los estudios aprovecha la aplicación de ontologías de dominio para mejorar el proceso de clasificación de la polaridad.

En S03 [126] se propone segmentar en oraciones los documentos, según las diversas propiedades de un concepto, y luego calcular la polaridad general como un promedio ponderado de las polaridades individuales de cada segmento según la importancia de cada propiedad.

En S06 [63] se plantea la construcción automática de un lexicon de sentimiento específico de un dominio con el apoyo de una ontología de dominio difusa. La ontología modela relaciones taxonómicas y no taxonómicas entre productos y sus atributos. Para cada característica de los productos se calcula una polaridad particular para las palabras que la modifican. Ello permite atender al hecho de que ciertas palabras cambian su polaridad dependiendo del objeto al cual se aplican. P.ej.: «pequeña» se considera positiva en el contexto de «laptop», mientras que usualmente es considerada negativa.

En S07 [125] se hace una propuesta para la minería de opinión en reseñas de películas en la que se emplea una ontología de dominio para calcular la polaridad con respecto a cada característica de las películas y realizar una transferencia de polaridad de los nodos (características) a su raíz (cada película de manera general).

En S12 [115], S23 [71] y S24 [116] el uso de ontologías de dominio con estructura arbórea (*Sentiment Ontology Trees - SOT*) permite abordar la identificación de atributos y la clasificación de la polaridad como problemas

## 5.1. CONTRIBUCIÓN AL ANÁLISIS DE SENTIMIENTO (RQ1) 43

de clasificación jerárquica. Las propuestas se diferencian en rendimiento, precisión y exhaustividad.

En S16 [124] se aplica una ontología sobre productos digitales para diferenciar la polaridad de una palabra de sentimiento variable según la característica a la que se aplica.

En S19 [27] se construye de manera automática tres ontologías ad hoc para un dominio específico: una de documentos positivos, una de negativos, una de neutrales. Luego se clasifica la polaridad de un nuevo documento dependiendo de una función de similaridad con las ontologías construidas.

En S22 [67] se usa una ontología de dominio para representar el conocimiento sobre el control de precios inmobiliarios, en la que se identifica las entidades de opinión (palabras que expresan un sentimiento) y las palabras cuya coocurrencia las hace relevantes (relación de conexión), irrelevantes (relación mutex) o modifica su intensidad (modificadores).

En S38 [1] se propone utilizar ontologías para identificar las características importantes de un dominio y luego considerar en la clasificación de la polaridad solamente las opiniones asociadas a las características relevantes presentes en la ontología.

En S41 [66] se construye dos ontologías por dominio, una con términos positivos y otra con términos negativos. Luego se determina la polaridad de los documentos individuales usando un clasificador de texto multilingüe que los compara con las ontologías construidas.

En S42 [70] se genera de manera automática una ontología afectiva específica para un dominio. La ontología afectiva modela la relación semántica entre los productos, componentes, atributos y palabras afectivas. Con ella se trata de superar la dependencia de dominio de ciertas palabras afectivas.

En S44 [103] se emplea una ontología difusa de emociones para anotar la polaridad y nivel de intensidad de las palabras evaluativas y emocionales presentes en cada oración.

**Beneficios** *Mayor precisión y exhaustividad.* Seis de los estudios dan evidencia de que con el apoyo del conocimiento modelado en ontologías de dominio se puede mejorar la precisión y exhaustividad de la clasificación de la polaridad.

En S03 [126] se constata una mejora en la precisión general usando tanto un modelado de lenguaje de n-gramas como la técnica General Inquirer [104]. El mismo estudio reporta también que se obtuvieron resultados más balanceados (menos sesgados) entre la precisión positiva y negativa.

En S06 [63] el uso de una ontología se emplea para tomar en cuenta el contexto (dominio) de la opinión y se reporta una mejor precisión en la



clasificación en relación a FBS [54], OPINE [90] y Opinion Observer [35].

En S23 se mejora la precisión y exhaustividad de la clasificación de la polaridad (en relación a Wei [115]) al enriquecer la ontología con conjuntos difusos de sinónimos y con la identificación de relaciones «is-a» entre sus conceptos.

*Clasificación de textos sobre más de un producto a la vez.* En S12 [115] y S24 [116] Wei et al., el uso de ontologías de dominio con estructura arbórea (*Sentiment Ontology Trees - SOT*) hace posible analizar el sentimiento de documentos que versan sobre más de un producto, dado que en esta aproximación el análisis de sentimiento es abordado como un problema de clasificación. El algoritmo utilizado, denominado HL-SOT se basa en una generalización del algoritmo de aprendizaje en línea H-RLS [22].

*Clasificación de términos de sentimiento variable según contexto.* El soporte de ontologías de dominio en el proceso de clasificación de la polaridad hace posible considerar el contexto de la opinión para identificar la correcta polaridad de términos de sentimiento variable, esto es, dependientes del contexto en el cual se empleen. Por ejemplo la palabra «alto» tiene una polaridad positiva en «alto salario», pero negativa en «alto consumo de combustible».

*Ponderación del grado de relevancia de opiniones individuales.* Las ontologías de dominio dan también soporte al proceso de ponderar la relevancia de diversas opiniones al momento de agregarlas. En S22 [67] se hace la ponderación a partir de la identificación de indicadores de relevancia, mientras que en S38 [1] se considera como relevantes solamente las opiniones sobre características modeladas en la ontología, cuyo grado de relevancia se determina en proporción inversa al nivel de profundidad del nodo correspondiente en la ontología.

*Soporte para análisis multilingüe.* Dado que las ontologías modelan conceptos que pueden tener asociados uno o más términos en diversos idiomas, la aplicación de ontologías de dominio permite dar también soporte al análisis de sentimiento de texto en diversos idiomas. En S41 [66] se realiza así un análisis de polaridad multilingüe en documentos de cinco idiomas.

**P4: Identificación de expectativas frustradas (*thwarting*)** Uno de los estudios (S30 [92]) realiza un particular análisis sobre la manera de identificar los documentos en los cuales se presentan giros abruptos en la orientación del discurso. El *thwarting* o expectativas frustradas se define como el fenómeno por el cual el autor levanta en primer lugar un conjunto de expectativas sobre el tema, para luego presentar deliberadamente un contraste con ellas [80]. Se da por ejemplo cuando se enumera una a una las buenas prestaciones de un producto para luego finalmente hacer un giro y señalar

## 5.1. CONTRIBUCIÓN AL ANÁLISIS DE SENTIMIENTO (RQ1) 45

que ellas no son de ninguna utilidad por faltarle una característica imprescindible. En este ejemplo el análisis de sentimiento encuentra la dificultad de que la polaridad del documento sería mayoritariamente positiva mientras que la polaridad general atribuida por el autor no lo es. Con la ayuda de una ontología de dominio se puede redefinir el fenómeno de (*thwarting*) como aquel en el que la polaridad general del documento está en contraste con la polaridad de la mayoría de su contenido.

**Beneficios** *Soporte para reconocer documentos «thwarted».* La aplicación de una ontología de dominio permite redefinir el fenómeno de (*thwarting*) aquel en el que la polaridad general del documento está en contraste con la polaridad de la mayoría de su contenido. Sin embargo, en S30 [92] se parte del dato previo de la polaridad general y las polaridades de las partes, con lo cual no se ha desarrollado aún un método para calcular la polaridad general en los casos donde se presente este fenómeno.

**P5: Clasificación de emociones** Un 20% de los estudios se apoya en ontologías de dominio para clasificar el sentimiento con más detalle que la polaridad e identificar las emociones expresadas.

En S10 [47] se presenta un marco de trabajo para un análisis de las emociones de los clientes en un sistema de Gestión de Relacionamiento con Clientes (CRM). En la ontología denominada CEO se representa los productos y sus atributos, así como las emociones asociadas por los cliente a ellos.

En S13 [6] y S27 [11] se emplea una ontología emocional para analizar las etiquetas que los visitantes asignan a las obras de arte de un museo virtual y a partir de ellas identificar las emociones asociadas a cada una de ellas.

En S14 [50] se usa una ontología de emociones humanas para inferir la categoría emocional asociada a un vector emocional de cuatro dimensiones obtenido como salida de un proceso de *sentic computing*.

En S26 [121], S32 [108], S37 [114] y S40 [29] se realiza un análisis de emociones categoriales definidas en una ontología emocional de apoyo.

En S44 [103] se emplea una ontología difusa de emociones para anotar la polaridad y nivel de intensidad de las palabras evaluativas y emocionales presentes en cada oración.

**Beneficios** *Representación de modelo emocional.* Un beneficio inmediato de la aplicación de una ontología de dominio para la clasificación de emociones es el soporte para la representación de modelos emocionales con diversos grados de complejidad. Este beneficio se verificó en todos los estudios que realizaron una clasificación de emociones.

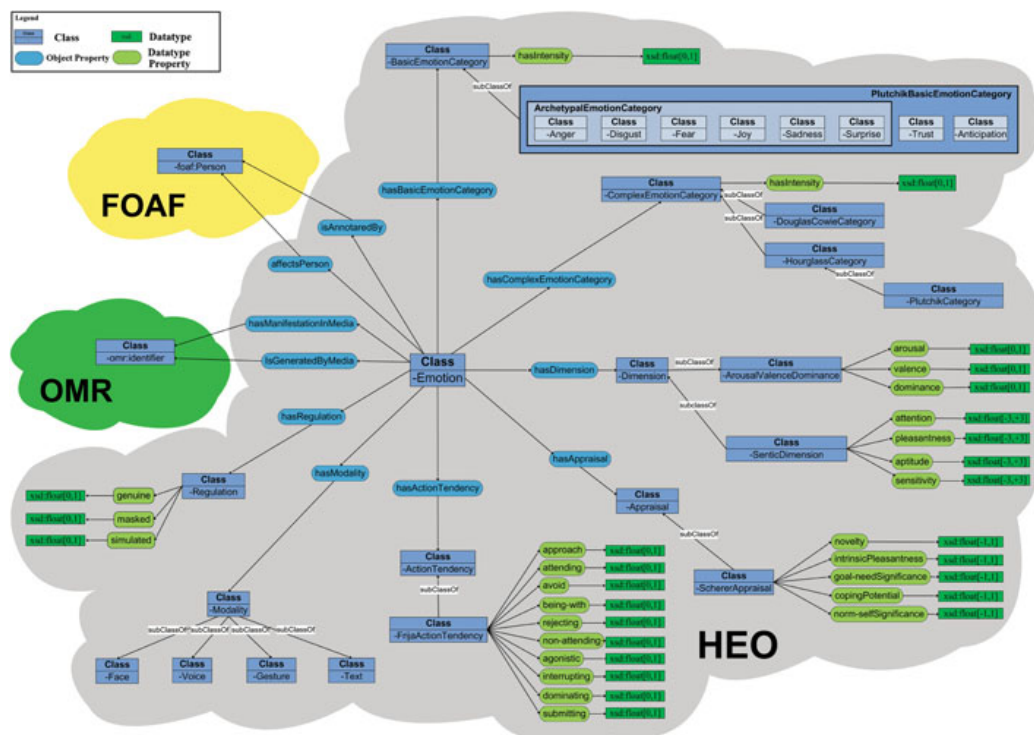


Figura 5.2: Human Emotion Ontology (HEO) [49]

*Calcular emociones predominantes por medio de inferencias semánticas.* Los mecanismos de inferencia semántica son aprovechados en S13 [6] para identificar las emociones predominantes asociadas a cada una de las obras de arte etiquetadas por los visitantes de un museo virtual. El uso de una ontología permite estructurar las emociones y las palabras relacionadas a ellas en diversos idiomas. Con ello se puede extraer el contenido emocional de las emociones y calcular las emociones predominantes por medio de inferencias.

*Interoperabilidad de modelos emocionales.* Dada la variedad de modelos emocionales propuestos, se hace necesario definir las equivalencias y mecanismos de traducción de la representación emocional entre diversos modelos. En S14 [50] se emplea para ello la Ontología de Emociones Humanas (HEO) (ver figura 5.2), la cual provee flexibilidad en la representación de las emociones, permitiendo el uso de un conjunto amplio y extensible de descriptores para definir las principales características de una emoción. Ello posibilita una efectiva interoperabilidad entre diferentes aplicaciones que usen diferentes vocabularios o modelos emocionales. Permite mapear conceptos y propiedades que corresponden a diferentes representaciones emocionales.

*Comparación y combinación de emociones.* En S27 [11], sobre la base de S13 [6], no sólo se calcula las emociones predominantes sino que se realiza

## 5.1. CONTRIBUCIÓN AL ANÁLISIS DE SENTIMIENTO (RQ1) 47

inferencias semánticas para comparar y combinar las emociones extraídas usando un modelo dimensional de representación de emociones (Plutchik).

*Soporte para visualización de emociones.* La organización y estructuración del conocimiento que hacen posibles las ontologías es aprovechada también en vistas a la visualización de las emociones detectadas como parte del análisis de sentimiento. En S14 [50] la capacidad de inferencias de OWL se aplica a la Ontología de Emociones Humanas (HEO) y se explota para realizar una administración y visualización eficiente de la información codificada en la base de conocimiento. En S27 [11] se reporta que el uso de la ontología y el modelo emocional de Plutchik inspiraron una interface gráfica interactiva para visualizar y sumarizar los resultados del algoritmo de detección de emociones.

*Mayor precisión.* En S32 [108] se realiza un análisis de sentimiento en el que cada pieza de texto recibe 8 puntajes, uno para cada emoción básica del modelo aplicado. Se obtuvo un incremento en la precisión para la detección de emociones en relación a [31]. Asimismo, se logró una representación más específica de las emociones en relación a SentiStrength-2 [111] (análisis basado en lexicon pero sin estructuración ontológica).

**P6: Identificación de tópicos/atributos** Un 35.6 % de los estudios emplearon el conocimiento modelado en ontologías de dominio para la identificación de los tópicos y atributos presentes en los documentos de opinión. Este es el caso más común de aplicación de ontologías de dominio al análisis de sentimiento. Este resultado era esperado de acuerdo a la investigación preliminar. Una ontología de dominio es idónea para la representación de tópicos o productos y sus atributos, así como la relaciones entre ellos y otras entidades.

**Beneficios** *Mayor precisión y exhaustividad.* En relación a OPINE [90] y TermExtractor [98], en S01 [26] se obtuvo una mejora notable en la precisión y exhaustividad de la identificación de tópicos luego de aplicar una ontología de dominio y vincular los tópicos con los conceptos de dominio.

En S09 [33] también se reporta mejor precisión y exhaustividad cuando se considera como características para el modelo de clasificación de Conditional Random Fields los datos provenientes de una ontología de dominio (concepto, propiedad y/o relación correspondiente a cada palabra).

En S25 [120], en relación a [54], el uso de una ontología de dominio incrementa la precisión y exhaustividad en la identificación de los atributos de producto y en la identificación de las oraciones que representan una opinión.

**P7: Identificación de pares tópico-sentimiento** Un 11.1% de los estudios empleó ontologías de dominio como soporte para realizar una identificación de pares tópico-sentimiento, es decir, para emparejar correctamente las palabras o conceptos que representan un tópico con aquellos que representan una opinión sobre ese tópico particular. Este apoyo es especialmente provechoso cuando la relación entre el tópico y el sentimiento es indirecta o implícita. La relación es indirecta o implícita cuando un sentimiento referido a un atributo es predicado de un atributo más general o del mismo producto [34]. Por ejemplo, en la oración «El auto tiene un precio alto y se ve hermoso», se obtiene dos pares («precio-alto» y «auto-hermoso»), el primero de los cuales es una relación directa y el segundo es indirecta pues «hermoso» se refiere de manera directa al atributo «apariciencia» y sólo de manera indirecta al producto «auto».

En S08 [25] se emplea una ontología para identificar pares tópico-sentimiento en los que la relación es indirecta, es decir, para transferir el sentimiento de un tópico implícito a uno de más alto nivel para formar una relación indirecta.

En S11 [106] se presenta una metodología en la que se usa una ontología para identificar a qué atributo de un producto están dirigidas la opiniones teniendo en cuenta información sintáctica y semántica.

S33 [122] emplea una ontología para identificar a qué atributo implícito o explícito de un producto están dirigidas la opiniones.

En S39 [3] se puebla una ontología de dominio difusa con información de hoteles disponibles, clientes y proveedores, así como con un lexicon de opinión generado automáticamente, la cual es luego usada para extraer las características relevantes y asociar la polaridad a cada una de ellas.

En S42 [70] se emplea una ontología de dominio en la que se modela la relación semántica entre los productos, componentes, atributos y palabras afectivas.

**Beneficios** *Normalización del lenguaje.* En S11 [106] una ontología de dominio permite normalizar el lenguaje para distinguir entre las diferentes características del producto.

*Identificación en caso de tópicos implícitos.* En S33 [122] se consigue identificar las características explícitas e implícitas asociadas a las opiniones vertidas. Asimismo se sumariza los pares opinión-característica asociados a cada producto.

*Mayor precisión y exhaustividad.* En S08 [25] se reporta una mejora tanto de la precisión como de la exhaustividad en la asociación de los sentimientos a su tópico implícito, en relación a las técnicas de Par más cercano [69] y Análisis de dependencias sintácticas [72].

## 5.1. CONTRIBUCIÓN AL ANÁLISIS DE SENTIMIENTO (RQ1) 49

En S11 [106] se obtiene mejores resultados en la aplicación de aprendizaje de máquina para la extracción automática de reglas de identificación de los pares tópico-sentimiento. Se compara la técnica propuesta con un método basado en adyacencia [54], un método basado en patrones [90] y el mismo método propuesto pero sin el uso de la ontología de dominio.

En S33 [122] la metodología propuesta obtiene una mejor precisión y exhaustividad para la identificación de características explícitas, en comparación con un método de reglas de asociación y adyacencia explícita [54], un método de relaciones sintácticas [91], un método semi-supervisado con ontología de dominio [125, 84] y un método no supervisado con ontología de dominio [62, 63]. A diferencia de S33, las cuatro metodologías con las que se compara no identifican características implícitas.

**P8: Recuperación de información** Un 20% de estudios aplican ontologías de dominio para apoyar la recuperación de información a partir de los documentos de opinión. A diferencia del problema tradicional de análisis de sentimiento, en el que se recibe un conjunto de documentos que deben ser analizados, identificándose primero el tópico sobre el cual versan, la recuperación de información permite iniciar el proceso definiendo primero el tópico de análisis y luego recuperando para el análisis solamente un conjunto de documentos que versen sobre ese tópico.

En S02 [105] se aplica recuperación de información con apoyo de una ontología de dominio para recuperar opiniones sobre un tópico particular en un foro de discusión web.

En S15 [118] se usa la ontología para representar las opiniones con todos sus elementos y luego recuperar la información de la base de datos así generada.

En S28 [59], S34 [78] y S40 [29] se analiza el sentimiento sobre un conjunto de productos buscando y procesando *tweets* sobre cada una de las características de cada producto.

En S31 [95] se utiliza recuperación de información para responder preguntas como «¿Qué modelo de tablet es excelente?». Para ello se realiza una expansión de búsqueda mediante una ontología de productos y de polaridades.

En S36 [97] se usa las ontologías para identificar las entidades de dominio, para la representación y vinculación de datos como linked data. Una vez representados semánticamente los datos, se recupera la información mediante consultas semánticas.

En S39 [3] una ontología de dominio es poblada con información de hoteles disponibles, clientes y proveedores, así como con un lexicon de opinión gene-

rado automáticamente. Es usada para extraer las características relevantes, para asociar la polaridad a cada una de ellas y para recuperar información sobre hoteles con características y polaridades específicas por medio de DL y SPARQL.

**Beneficios** *Mayor precisión y exhaustividad.* En S02 [105] se obtiene una notable mejora en la exhaustividad de opiniones recuperadas en comparación a una indización de texto completo con Lucene<sup>2</sup>. La precisión de la recuperación es equivalente o superior.

En S15 [118] se reporta mayor precisión y flexibilidad de búsquedas con metadata. También se compara con una indización de texto completo.

En S28 [59] se reporta un alto grado de exhaustividad para recuperar tweets relevantes que permitan un análisis del sentimiento asociado a cada una de las distintas características de un producto.

**P9: Sumarización** Un 17.8 % de los estudios aplica ontologías de dominio como soporte para una sumarización del análisis del sentimiento. Gracias a la estructuración de la información modelada en las ontologías de dominio, éstas facilitan la sumarización de los resultados según múltiples criterios: sea por producto, por características, por niveles de polaridad, por emociones presentes, por fuente de origen, etc.

En S05 [32] se usa una ontología de dominio para sumarizar las opiniones sobre cada una de las características individuales de un producto.

En S17 [7] se emplea una ontología para identificar el objeto y atributo al que corresponden las opiniones, agregarlas y sumarizarlas.

En S20 [55] se aplica una ontología para identificar a qué aspecto de un servicio están dirigidas la opiniones.

En S21 [65] una ontología de dominio permite calcular y sumarizar puntajes independientes de polaridad para cada uno de los atributos de calidad en uso de un producto software.

En S27 [11] se sumariza y visualiza las emociones asignadas colectivamente a una obra de arte.

En S34 [78] se se aplica una ontología para recuperar tweets sobre cada una de las características de diversos productos y luego sumarizar los resultados de su análisis.

**Beneficios** *Sumarización con diversos niveles de especificidad.* En S05 [32] se agrega opiniones sobre productos y sus atributos con diversos niveles de especificidad: atributo, producto, usuario, sitio fuente.

---

<sup>2</sup><http://lucene.apache.org/java/docs/>

En S17 [7] dos ontologías (una de modelos de automóviles Fiat y una de modelos de otras marcas) permiten identificar sólo los documentos de la marca estudiada. Una ontología de componentes de automóviles permite calcular el sentimiento general sobre la marca estudiada en el nivel deseado y sumarizar el sentimiento asignado a los nodos descendientes.

*Sumarización de emociones.* En S27 [11] el uso de una ontología de dominio basada en el modelo emocional de Plutchik inspiró una interface gráfica interactiva para visualizar y sumarizar los resultados del algoritmo de detección de emociones.

## 5.2. Aplicabilidad a otros dominios (RQ2)

*¿Las investigaciones reportadas son generalizables a dominios diferentes al de las ontologías aplicadas durante la experimentación?*

### 5.2.1. Resultados

La segunda pregunta de investigación tiene por objetivo analizar si los resultados obtenidos en los estudios son generalizables a otros dominios de aplicación. Para poder analizar los resultados obtenidos se tabuló asimismo el dominio de las ontologías aplicadas en cada estudio y las ontologías existentes reutilizadas en ellos.

En la tabla 5.6 se presenta el dominio de las ontologías aplicadas en los estudios. La suma de conteos o porcentajes es mayor a la totalidad de estudios debido a que varios estudios experimentan con más de una ontología de dominio.

Durante la etapa de síntesis se agrupó los dominios en tres tipos, que son también presentados en la primera columna de la tabla 5.6:

- *Opiniones y emociones.* Comprende las ontologías que modelan las opiniones, las emociones, o su análisis.
- *Redes sociales.* Comprende las ontologías que modelan la información contenida en los documentos obtenidos de redes sociales.
- *Productos y servicios.* Comprende las ontologías que modelan los dominios sobre los que versan los documentos de opinión.

Un total de 15 estudios (33.3%) reutilizan ontologías preexistentes. En la tabla 5.7 se presenta las ontologías reutilizadas en los estudios.



Tabla 5.6: Dominio de las ontologías aplicadas

Tipo de dominio	Dominio	Estudios	Conteo	%
Opiniones y emociones	Emociones	S13, S14, S26, S27, S31, S32, S37, S40, S44	9	20.0%
Productos y servicios	Teléfonos móviles	S06, S08, S09, S19, S25, S28, S33, S34	8	17.8%
Opiniones y emociones	Minería de opinión	S02, S15, S29, S36, S43, S45	6	13.3%
Productos y servicios	Cámaras fotográficas	S06, S08, S11, S12, S24, S30	6	13.3%
Productos y servicios	Automóviles	S01, S02, S05, S08, S17	5	11.1%
Productos y servicios	Computadoras portátiles	S06, S08, S09, S23, S42	5	11.1%
Productos y servicios	Películas	S03, S07, S35, S38	4	8.9%
Productos y servicios	Hoteles	S18, S39, S41	3	6.7%
Opiniones y emociones	Análisis de emociones	S36, S40	2	4.4%
Productos y servicios	Productos electrónicos	S16, S31	2	4.4%
Productos y servicios	Relojes	S06	1	2.2%
Productos y servicios	Equipos deportivos	S06	1	2.2%
Productos y servicios	Finanzas	S36	1	2.2%
Productos y servicios	Videojuegos	S04	1	2.2%
Redes sociales	Redes sociales	S10	1	2.2%
Opiniones y emociones	Emociones de cliente	S10	1	2.2%
Productos y servicios	Políticas públicas	S20	1	2.2%
Productos y servicios	Calidad en uso de productos software	S21	1	2.2%
Productos y servicios	Control de precios inmobiliarios	S22	1	2.2%
Productos y servicios	Restaurantes	S38	1	2.2%
Productos y servicios	Software	S38	1	2.2%
Redes sociales	Twitter	S40	1	2.2%

Tabla 5.7: Ontologías reutilizadas

Ontologías reutilizadas	Dominio	Estudios	Conteo	%
Friend of a Friend (FOAF)	Redes sociales	S14, S36, S40	3	6.7%
Marl [118]	Minería de opinión	S15, S36, S45	3	6.7%
OntoEmotion [40]	Emociones	S13, S27, S40	3	6.7%
Semantically-Interlinked Online Communities (SIOC)	Redes sociales	S02, S36, S40	3	6.7%
Dbpedia	General	S02, S15	2	4.4%
DUTIR Emotion Ontology [123]	Emociones	S26, S37	2	4.4%
Customer Emotion Ontology (CEO) [47]	Emociones	S10	1	2.2%
DAML Event Ontology (DAML.org 2004)	Eventos	S29	1	2.2%
DOLCE+DnS Ultralite ontology [45]	Ontología de alto nivel	S43	1	2.2%
Dublin Core ontology	Recursos bibliográficos	S40	1	2.2%
Financial Industry Business Ontology (FIBO) [9]	Finanzas	S36	1	2.2%
GI2MO	Gestión de ideas	S36	1	2.2%
Hontology [24]	Hotelería	S18	1	2.2%
Movie Ontology (movieontology.org)	Películas	S35	1	2.2%
Ontology for Media Resources (OMR)	Recursos mediáticos	S14	1	2.2%
Open Social Network Dataset (OSND) [94]	Redes sociales	S10	1	2.2%
Simple Knowledge Organization System (SKOS)	Gestión del conocimiento	S02	1	2.2%

Tabla 5.8: Ontologías construidas

Dominio	Ontologías construidas	Estudio
Minería de opinión	Opinion Mining Core Ontology	S02
Minería de opinión	MARL - An Ontology for Opinion Mining	S15
Minería de opinión	Information Systems Sentiment Ontology	S29
Minería de opinión	TweetOntoSense	S40
Minería de opinión	OntoSentilo	S43
Emociones	OntoEmotion4arsemotica	S13
Emociones	Human Emotion Ontology (HEO)	S14
Emociones	Plutchik4arsemotica	S27
Emociones	EMOTIVE	S32
Emociones	Onyx	S36
Automóviles	CarOnto	S01
Hoteles	SentiHotel	S41
Redes sociales	Twitter Ontology	S40

Un total de 12 estudios (26.7%) construyen y proponen ontologías reutilizables. En la tabla 5.8 se presenta el dominio de las ontologías construidas en los estudios.

### 5.2.2. Análisis y discusión

En relación a la aplicabilidad de las investigaciones realizadas a dominios diferentes de las ontologías utilizadas se necesita hacer una distinción a partir de los tipos de ontologías aplicadas.

Las ontologías sobre el dominio de opiniones, emociones o redes sociales no son claramente generalizables a otros dominios.

La categoría de productos más utilizados para ejemplificar la aplicación de ontologías de dominio al análisis de sentimiento son los teléfonos móviles (17.8%), seguida de las cámaras fotográficas (13.3%), los automóviles (11.1%) y las computadoras portátiles (11.1%). Sin embargo, los estudios en que se los presenta son siempre generalizables a otros productos o servicios.

Se observa que el grado de reutilización de ontologías de productos y servicios es mínimo. Por el contrario, si bien no hay una estandarización en la representación semántica de las emociones o de la minería de opinión se encuentra un mayor grado de reutilización de estas ontologías de dominio.

Entre las ontologías del dominio de la minería de opinión destaca Marl<sup>3</sup> (ver figura 5.1), propuesta en S15 [118], S36 [97] y S45 [110].

<sup>3</sup><http://www.gsi.dit.upm.es/ontologies/marl/>



Figura 5.3: Fragmento de OntoEmotion [40]

Entre las ontologías de representación emocional destacan por su adopción OntoEmotion [40] (ver figura 5.3) y por su riqueza expresiva Human Emotion Ontology (HEO) [50] (ver figura 5.2) y Onyx<sup>4</sup> [96] (ver figura 5.4).

### 5.3. Expresividad y razonamiento ontológico (RQ3)

*¿En qué medida se está aprovechando la expresividad y mecanismos de razonamiento propios de las ontologías de dominio en el análisis de sentimiento?*

La tercera pregunta de investigación tiene por objetivo identificar de qué manera se viene aprovechando las posibilidades de expresividad y razonamiento que proveen las ontologías. Ha sido dividida en dos subpreguntas en las que se identifica respectivamente los componentes de representación del conocimiento presente en las ontologías de dominio aplicadas y los razonadores empleados para realizar inferencias semánticas.

#### 5.3.1. Componentes de representación del conocimiento (RQ3.1)

*¿Qué componentes de representación del conocimiento propios de las ontologías están siendo aplicados en el análisis de sentimiento?*

<sup>4</sup><http://www.gsi.dit.upm.es/ontologies/onyx/>

5.3. EXPRESIVIDAD Y RAZONAMIENTO ONTOLÓGICO (RQ3) 55

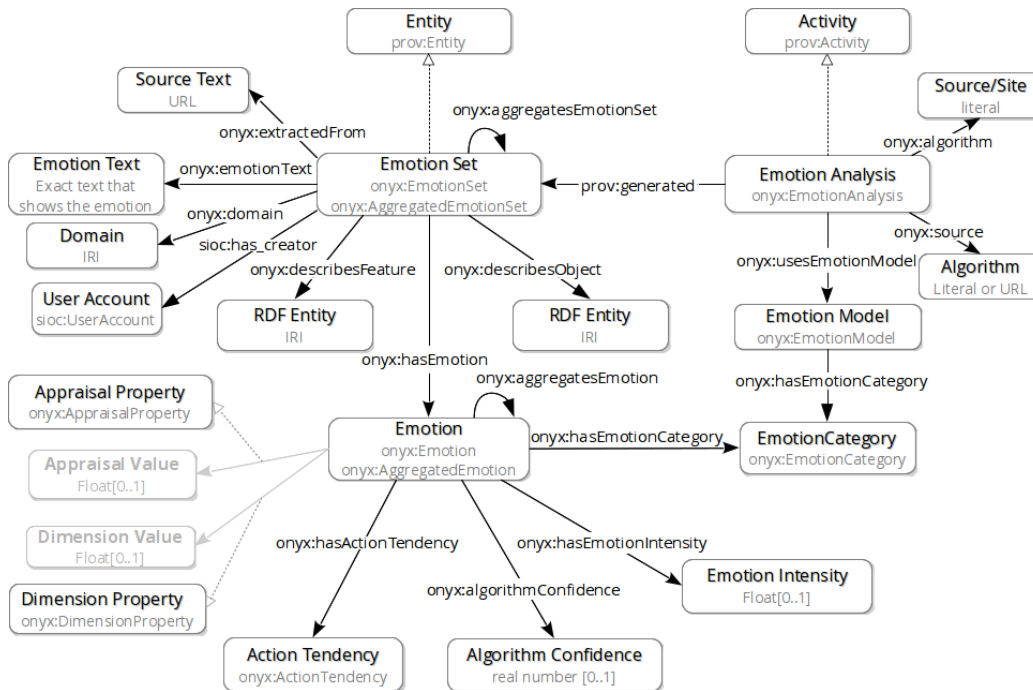


Figura 5.4: Ontología Onyx [96]

Resultados

Se tabuló los diversos componentes semánticos presentes en las ontologías de dominio aplicadas en los dominios. Los resultados se presentan en la tabla 5.9.

Análisis y discusión

Los dos componentes semánticos más representados son las relaciones taxonómicas (91.1%) y los atributos (88.9%). El resultado es esperado, puesto que estos son los componentes más elementales para la clasificación de productos y sus propiedades, presentes en el tipo de aplicación más común de las ontologías de dominio para el análisis de sentimiento.

Un 62.2% de estudios representaron relaciones no taxonómicas. Estas relaciones añaden riqueza expresiva al modelamiento de la vinculación entre conceptos.

El uso de reglas fue identificado en 6 estudios (13.3%), 3 de ellos con ontologías de emociones (S14 [50], S27 [11] y S36 [97]) 2 con ontologías de minería de opinión (S15 [118] y S43 [93]), y uno de ellos con una ontología de reservas hoteleras (S39 [3]).

Sólo 2 estudios (4.4%) reportaron haber incluido axiomas formales (S39

Tabla 5.9: Componentes de representación del conocimiento presentes en las ontologías de dominio empleadas

Componentes semánticos	Estudios	Conteo	%
Atributos	S01, S02, S03, S04, S05, S06, S07, S08, S09, S11, S12, S13, S14, S15, S16, S17, S18, S19, S21, S22, S23, S24, S25, S27, S28, S29, S30, S31, S32, S33, S34, S35, S36, S38, S39, S40, S41, S42, S43, S44	40	88.9 %
Relaciones taxonómicas	S01, S02, S04, S05, S06, S07, S08, S09, S10, S11, S12, S13, S14, S15, S16, S17, S18, S19, S20, S21, S22, S23, S24, S25, S27, S28, S29, S30, S31, S32, S33, S34, S35, S36, S38, S39, S40, S41, S42, S43, S44	41	91.1 %
Relaciones ad-hoc	S01, S02, S05, S06, S07, S08, S09, S11, S13, S14, S15, S16, S17, S18, S19, S22, S27, S28, S29, S33, S34, S35, S36, S38, S39, S40, S42, S43	28	62.2 %
Axiomas formales	S39, S43	2	4.4 %
Reglas	S14, S15, S27, S36, S39, S43	6	13.3 %

Tabla 5.10: Herramientas de razonamiento semántico utilizadas

Estudios	Herramientas de razonamiento
S10	KAON2 [75]
S13	Apache Jena
S14	Sesame [15]
S27	Hermit OWL Reasoner [100]
S39	DeLorean [12]
S40	Apache Jena

[3] y S43 [93]).

Se puede concluir que en las ontologías examinadas hay un bajo nivel de adopción de los elementos de mayor refinamiento semántico.

### 5.3.2. Inferencias y razonamiento ontológico (RQ3.2)

*¿Qué razonadores están siendo usados para realizar inferencias en la aplicación de ontologías de dominio en el análisis de sentimiento?*

#### Resultados

Se registró asimismo los estudios que reportaron haber realizado inferencias semánticas sobre los datos modelados según las ontologías de dominio. Sólo 6 estudios (13.3%) indican haber realizado inferencias. La lista de los estudios, con el detalle de la herramienta empleada para el razonamiento semántico se presenta en la tabla 5.10.

## 5.4. REPRESENTACIÓN DE EMOCIONES (RQ4)

57

Tabla 5.11: Modelos y tipos de representaciones emocionales

Estudio	Modelo emocional	Tipo de representación		
		Categorial	Dimensional	Componencial
S10	Laros y Steenkamp [61]	X		
S13	Parrot [82]	X		
S14	Permite usar cualquier modelo emocional.	X	X	
S26	DUTIR Emotion Ontology [123]	X		
S27	Plutchik [88]	X	X	
S29	Componencial [43]			X
S31	No se indica.	X		
S32	Modelo propio a partir de Ekman [38], Plutchik [87], Drummond [36] e Izard [56]	X		
S36	Permite usar cualquier modelo emocional.	X	X	
S37	No se indica.	X		
S40	Francisco [41]	X		
S44	Modelo propio. 8 categorías: Expect, Joy, Love, Surprise, Anxiety, Sorrow, Anger & Hate.	X		

**Análisis y discusión**

De los 6 estudios que reportan haber realizado inferencias, sólo dos de ellos (S27 [11] y S39 [3]) reportaron el detalle específico del razonador utilizado (respectivamente, Hermit y DeLorean). En el resto de estudios, sólo se reportó la herramienta o marco de trabajo empleado, tales como KAON2, Sesame o Apache Jena.

S39 [3] es el único estudio que utilizó, según los parámetros tabulados, el máximo de expresividad semántica y que además realizó inferencias semánticas.

Los resultados no dan evidencia de preferencias ni estandarización en cuanto a los razonadores empleados.

**5.4. Representación de emociones (RQ4)**

*¿Cómo se está representando las emociones en las ontologías de dominio aplicadas en el análisis de sentimiento?*

La última pregunta de investigación tiene por objetivo identificar cómo se está representando las emociones en las ontologías de dominio. La motivación viene del hecho de que en el estudio preliminar se constató que el análisis de sentimiento viene evolucionando desde una análisis de la polaridad hacia un análisis más detallado en el que se detecta las emociones expresadas.

Se presenta a continuación los resultados de esta pregunta de investigación y sus dos subpreguntas.

**5.4.1. Resultados**

Se encontró que un total de 12 estudios (26.7%), listados en la tabla 5.11, realizan un análisis a nivel de emociones apoyado con ontologías de dominio.

**Modelo emocional (RQ4.1)** *¿En qué modelo psicológico está fundada la clasificación de emociones usada en las ontologías de emociones aplicadas en el análisis de sentimiento?*

Esta subpregunta buscó observar los modelos emocionales en los que se basan los análisis de sentimiento a nivel de emociones. En la tabla 5.11 se lista el modelo emocional empleado por cada uno de los 12 estudios que realizaron este tipo de análisis.

Se puede observar que 2 de los estudios (S14 [50] y S36 [97]) son independientes del modelo emocional subyacente. Otros 3 estudios (S13 [6], S27 [11] y S29 [81]) asumen directamente un modelo psicológico general. Otros 3 asumen una clasificación establecida en otro estudio: S10 [47] asume la clasificación de emociones de cliente establecida en [61]; S26 [121] asume la clasificación implícita en una ontología existente [123]; y S40 [29] asume la clasificación empleada por [41]. Otros 2 (S32 [108] y S44 [103]) optan por un modelo propio. Finalmente, 2 estudios (S31 [95] y S37 [114]) no indican el origen o fuente de su clasificación.

**Tipo de representación (RQ4.2)** *¿Qué tipo de representación es usada para las emociones en las ontologías de emociones aplicadas en el análisis de sentimiento?*

En el estudio preliminar se había encontrado tres tipos o grupos de representaciones: categoriales, dimensionales y componenciales. En esta subpregunta se tabuló qué tipo de representación emocional es soportado por los estudios en los que se basan los análisis de sentimiento a nivel de emociones. En la tabla 5.11 muestra el detalle de los tipos de representación soportados.

Se observa que el tipo de representación más soportado es el categorial, presente en 11 de los 12 estudios (91.7%). Los otros tipos tienen una menor presencia: 3 estudios (25%) en el caso de la representación dimensional y 1 estudio (8.3%) la representación componencial.

#### 5.4.2. Análisis y discusión

Los resultados permiten constatar que un porcentaje significativo de los estudios de análisis de sentimiento que aplican ontologías de dominio se apoyan en ellas para identificar emociones y no sólo polaridad.

En cuanto a los modelos emocionales empleados, no se encontró ningún modelo que tenga un grado de adopción o preferencia significativo. La carencia de una estandarización en la clasificación emocional empleada dificulta la comparación entre diferentes propuestas.

Ante ello, los estudios que plantean una metodología independiente del modelo emocional subyacente tienen un mayor grado de aplicabilidad.

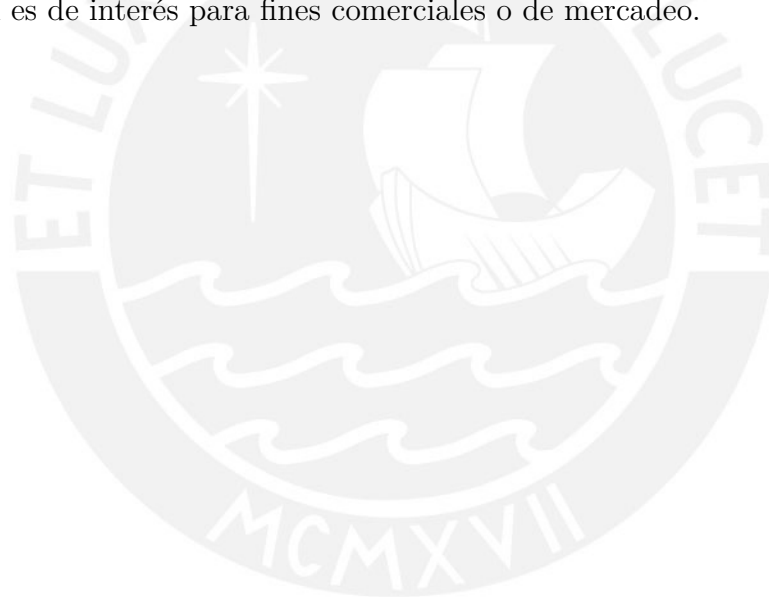
#### 5.4. REPRESENTACIÓN DE EMOCIONES (RQ4)

59

El hecho de que 91.7 % de los estudios soporten una clasificación categorial de las emociones es consistente el modo usual de entenderlas y nombrarlas en el lenguaje coloquial. Los 3 estudios que soportan una representación dimensional soportan también una representación categorial.

En este contexto es valioso el aporte de S14 [50], en el que se desarrolla la Ontología de Emociones Humanas (HEO), como esfuerzo de estandarización de la representación de las emociones humanas y soporte, mediante inferencias semánticas, para el mapeo de las emociones entre diversos modelos de representación.

S29 [81] fue el único estudio en emplear una representación componencial. En este caso la ontología aplicada modela el análisis del sentimiento a partir del modelo componencial de las emociones. En particular, modela la percepción de los usuarios y las emociones asociadas, y se predice la intención conductual (compra del producto o cambio por otro). Este tipo de aproximación es de interés para fines comerciales o de mercadeo.





## Capítulo 6

# Trabajos futuros y Conclusiones

### 6.1. Trabajos futuros

Los resultados encontrados en esta Revisión Sistemática de Literatura permiten señalar importantes y poco exploradas líneas de investigación futura, entre las cuales se puede mencionar:

1. ¿De qué manera se puede emplear ontologías de dominio para realizar un análisis de sentimiento a nivel de concepto (RQ1.1)? En S14 [50], que es la única publicación en la que se realiza un análisis de sentimiento a este nivel, se tiene un sugerente ejemplo con excelentes resultados que merecen ser profundizados y desarrollados.
2. ¿De qué manera una ontología sobre el análisis de sentimiento podría ayudar a mejorar el proceso entero de dicho análisis en todas sus fases (RQ1.2)? Las dos publicaciones que emplearon una ontología de dominio en todas las fases del análisis de sentimiento (S15 [118] y S43 [93]) se valieron de ontologías de representación de opiniones –respectivamente Marl y OntoSentilo–. ¿Sería posible representar no sólo las opiniones sino el proceso mismo de análisis de sentimiento por medio de una ontología de dominio? ¿Qué beneficios de estandarización, modularidad de algoritmos y vinculación de datos se harían posibles?
3. De manera relacionada al punto anterior, ¿qué otros beneficios (RQ1.3) se pueden derivar del empleo de una ontología de representación de opiniones como Marl u OntoSentilo? Los estudios que emplean dichas ontologías ya han mostrado que no sólo es posible representar de manera estructurada y semánticamente rica los diversos elementos de las opiniones, sino también resolverlos como *linked data* con un alto grado de precisión.

4. Las ontologías de dominio facilitan de manera natural la sumarización de resultados del análisis de sentimiento según múltiples criterios. Sin embargo, no se encontró una manera estandarizada de realizar dicha sumarización. ¿De qué manera se podría estandarizar la fase de sumarización del análisis de sentimiento con el auxilio de ontologías de dominio?
5. Como parte de la investigación revisada, se encontró numerosas ontologías de dominio sobre productos y servicios, generadas manualmente o con algún grado de automatización. Estas ontologías, a pesar de ser conceptualmente muy similares entre sí, no siguieron por lo general un estándar común, lo cual hace difícil su reutilización (RQ2) y la comparación de los resultados de los diferentes estudios. ¿De qué manera la vinculación con una ontología común como la *Product Types Ontology*<sup>1</sup> o estándares como *GoodRelations*<sup>2</sup> pueden agilizar la reutilización de las ontologías de dominio generadas y el avance de la investigación en análisis de sentimiento?
6. En RQ3 se concluyó que en las ontologías examinadas hay un bajo nivel de adopción de los elementos de mayor refinamiento semántico y se constató que sólo S39 [3] aprovechó el máximo de expresividad semántica de las ontologías difusas de tipo-2 empleadas, realizando inferencias semánticas como parte del proceso. Dicho estudio reportó notables resultados para la recuperación de opiniones relevantes en reseñas de hoteles. Una pregunta importante que queda por abordar es: ¿De qué manera el aprovechamiento de la riqueza y expresividad semántica de las ontologías de dominio, que son mucho más que taxonomías de clases, puede apoyar el análisis de sentimiento?
7. Se halló que la carencia de una estandarización en la clasificación emocional dificulta la comparación entre diferentes propuestas y aproximaciones de análisis de las emociones (RQ4). ¿De qué manera se puede integrar la interoperabilidad de modelos emocionales lograda en S14 [50] con otros modelos y aproximaciones de análisis de sentimiento? Dicha interoperabilidad, que permite integrar diversos vocabularios y tipos de representación emocional, así como aprovechar un conjunto amplio de descriptores emocionales, muestra una gran potencial que sería muy interesante profundizar.

---

<sup>1</sup><http://www.productontology.org/>

<sup>2</sup><http://purl.org/goodrelations/>

## 6.2. Conclusiones

Se ha presentado los resultados de una Revisión Sistemática de Literatura sobre la aplicación de ontologías de dominio en el análisis de sentimiento. Los objetivos del trabajo fueron realizar una síntesis de la contribución que ha tenido la aplicación de ontologías de dominio en la solución de los problemas específicos del análisis de sentimiento, reportar el grado de generalización que las investigaciones realizadas tienen para ser aplicadas en dominios diferentes a los de su experimentación inicial, verificar en qué medida viene siendo aprovechada la riqueza expresiva de las ontologías de dominio en su aplicación al análisis de sentimiento y reportar el estado del arte actual en la representación de las emociones humanas por medio de ontologías de dominio en su aplicación al análisis de sentimiento.

De los 1445 resultados obtenidos en 6 bases de datos electrónicas se seleccionó finalmente 45 estudios que cumplieron con los criterios de inclusión, exclusión y calidad definidos en el protocolo de revisión. Se extrajo y sintetizó los datos de estos estudios según las cuatro preguntas de investigación definidas.

Se confirmó el número creciente de publicaciones del área de investigación del análisis de sentimiento en las cuales se viene incluyendo la aplicación de ontologías de dominio.

En cuanto a la contribución de la aplicación de ontologías de dominio en las investigaciones de análisis de sentimiento se concluyó que todos los niveles de granularidad del análisis de sentimiento pueden ser apoyados con ellas, y que los niveles de granularidad actualmente más apoyados son los niveles de aspecto o característica y de documento.

Asimismo, se halló que la aplicación de ontologías de dominio suele apoyar una dos fases del análisis de sentimiento. Dos estudios realizan un apoyo a cuatro de las cinco fases. En ellos se hace una representación estructurada de las opiniones y se las vincula a otros conjuntos de datos (*linked data*).

Se identificó 9 distintos problemas del análisis del sentimiento a los que se aplicó ontologías de dominio y un total de 22 beneficios de dicha aplicación.

Una especial atención corresponde a los tres beneficios más reportados: (1) Soporte para una representación estructurada de las opiniones que posibilita la vinculación de los análisis de sentimiento con otros conjuntos de datos (*linked data*); (2) Mayor precisión y exhaustividad en la clasificación de la polaridad gracias a un aprovechamiento del conocimiento sobre las particularidades de cada dominio; y (3) Soporte para representar modelos emocionales y realizar un análisis de sentimiento que vaya más allá de una clasificación de la polaridad a una clasificación de emociones complejas.

En cuanto al grado de aplicabilidad de las aproximaciones reportadas a

otros dominios se concluyó que todos los estudios que emplean ontologías sobre productos o servicios son generalizables a otros dominios del mismo tipo, mientras que las ontologías sobre el dominio de opiniones, emociones o redes sociales no lo son.

Con respecto al grado de expresividad de las ontologías empleadas y el aprovechamiento de las capacidades de razonamiento de dichas representaciones de conocimiento se concluyó que no se está aprovechando los elementos de mayor refinamiento semántico propios de las ontologías y que sólo 13.3 % de estudios aplicaron inferencias semánticas, sin que se haya encontrado evidencia de ninguna estandarización en cuanto a los razonadores empleados.

Acerca de los modelos de representación emocional empleados en el análisis de sentimiento apoyados por ontologías de dominio se concluyó que 26.7 % de los estudios realizan un análisis de las emociones presentes en las opiniones, pero que no hay ninguna estandarización en los modelos de emociones empleados y que las clasificaciones son en su mayoría por medio de representaciones categoriales. No obstante, la representación dimensional, presente en 3 de los estudios, permite cuantificar las distintas representaciones categoriales y dar soporte a la interoperabilidad entre diferentes modelos emocionales.

Los resultados de esta Revisión Sistemática de la Literatura sintetizan el estado del arte sobre la aplicación de ontologías de dominio en el análisis de sentimiento, recogiendo los beneficios reportados, los niveles y fases en los que se pueden emplear, su grado de aplicabilidad y los diversos modelos emocionales presentes en los estudios realizados. Como tal, es de gran utilidad para los investigadores del análisis de sentimiento.

# Bibliografía

- [1] Basant Agarwal, Namita Mittal, Pooja Bansal, and Sonal Garg. Sentiment analysis using common-sense and context information. *Computational Intelligence and Neuroscience*, 2015:715730, jan 2015.
- [2] Khurshid Ahmad. *Affective computing and sentiment analysis: emotion, metaphor and terminology*, volume 45. Springer Science & Business Media, 2011.
- [3] Farman Ali, Eun Kyoung Kim, and Yong-Gi Kim. Type-2 fuzzy ontology-based opinion mining and information extraction: A proposal to automate the hotel reservation system. *Applied Intelligence*, 42(3):481–500, nov 2015.
- [4] Cecilia Ovesdotter Alm, Dan Roth, and Richard Sproat. Emotions from text: machine learning for text-based emotion prediction. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 579–586. Association for Computational Linguistics, 2005.
- [5] Orestes Appel, Francisco Chiclana, and Jenny Carter. Main Concepts, State of the Art and Future Research Questions in Sentiment Analysis. *Acta Polytechnica Hungarica*, 12(3), 2015.
- [6] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *LREC*, volume 10, pages 2200–2204, 2010.
- [7] Matteo Baldoni, Cristina Baroglio, Viviana Patti, and Paolo Rena. From Tags to Emotions : Ontology-driven Sentiment Analysis in the Social Semantic Web. *Diunitoit*, 771:1–12, 2011.
- [8] Renata Maria Abrantes Baracho, Gabriel Caires Silva, and Luiz Gustavo Fonseca Ferreira. Sentiment analysis in social networks. In *CEUR Workshop Proceedings*, volume 938, pages 132–143, 2012.
- [9] Andreas Beelmann, M. Petticrew, and H. Roberts. Systematic reviews in the social sciences. A practical guide. *European Psychologist*, 11(3):244–245, 2006.
- [10] Mike Bennett. The financial industry business ontology: Best practice for big data. *Journal of Banking Regulation*, 14(3-4):255–268, 2013.
- [11] Tim Berners-Lee, James Hendler, and Ora Lassila. The semantic web. *Scientific american*, 284(5):28–37, 2001.
- [12] Federico Bertola and Viviana Patti. Organizing artworks in an ontology-based semantic affective space. In *CEUR Workshop Proceedings*, volume 1096, pages 119–130. CEUR-WS, 2013.

- [13] Fernando Bobillo, Miguel Delgado, and Juan Gómez-Romero. Delorean: A reasoner for fuzzy owl 1.1. In *URSW*, 2008.
- [14] Damian Borth, Rongrong Ji, Tao Chen, Thomas Breuel, and Shih-Fu Chang. Large-scale visual sentiment ontology and detectors using adjective noun pairs. In *ACM International Conference on Multimedia (MM)*, pages 223–232. ACM, 2013.
- [15] Cristina Bosco, Viviana Patti, and Andrea Bolioli. Developing corpora for sentiment analysis: The case of irony and senti-TUT. *IEEE Intelligent Systems*, 28(2):55–63, March 2013.
- [16] Jeen Broekstra, Arjohn Kampman, and Frank Van Harmelen. Sesame: A generic architecture for storing and querying rdf and rdf schema. In *The Semantic Web—ISWC 2002*, pages 54–68. Springer, 2002.
- [17] Julian Brooke and Matthew Hurst. Patterns in the stream. In *Proceeding of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion - TSA '09*, TSA '09, page 1, New York, NY, USA, 2009. ACM.
- [18] Anaïs Cadilhac, Farah Benamara, and Nathalie Aussenac-Gilles. Ontolexical resources for feature based opinion mining: a case-study. In *23rd International conference on computational linguistics*, page 77. Citeseer, 2010.
- [19] Erik Cambria, Catherine Havasi, and Amir Hussain. SenticNet 2: A semantic and affective resource for opinion mining and sentiment analysis. In *Proceedings of FLAIRS, Marco Island*, pages 202–207, 2012.
- [20] Erik Cambria, Thomas Mazzocco, and Amir Hussain. Application of multi-dimensional scaling and artificial neural networks for biologically inspired opinion mining. *Biologically Inspired Cognitive Architectures*, 4:41–53, 2013.
- [21] Erik Cambria, Bjorn Schuller, Yunqing Xia, and Catherine Havasi. New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, 28(2):15–21, 2013.
- [22] Erik Cambria, Robert Speer, Catherine Havasi, and Amir Hussain. SenticNet : A Publicly Available Semantic Resource for Opinion Mining. In *Artificial Intelligence*, volume 10, pages 14–18, 2010.
- [23] N. Cesa-Bianchi, C. Gentile, and L. Zaniboni. Incremental algorithms for hierarchical classification. *The Journal of Machine Learning Research*, 7:54, 2006.
- [24] Marcirio Silveira Chaves, Larissa A. de Freitas, Marlo Souza, and Renata Vieira. PIRPO: an algorithm to deal with polarity in portuguese online reviews from the accommodation sector. In *Proceedings of the 17th international conference on Applications of Natural Language Processing and Information Systems*, volume 7337 LNCS, pages 296–301, 2012.
- [25] Marcirio Silveira Chaves and Cássia Trojahn. Towards a multilingual ontology for ontology-driven content mining in Social Web sites. In *CEUR Workshop Proceedings*, volume 687, 2010.
- [26] Mosha Chen and Tianfang Yao. Combining dependency parsing with shallow semantic analysis for Chinese opinion-element relation identification. In *2010 4th International Universal Communication Symposium*, pages 299–305. IEEE, oct 2010.

- [27] Xiwen Cheng and Feiyu Xu. Fine-grained Opinion Topic and Polarity Identification. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, number 1, pages 2710–2714, 2008.
- [28] Francesco Colace, Massimo De Santo, Paolo Napoletano, Carlo Becchi, and S.K. Chang. Ontological filtering for sentiment analysis. In *Proceedings: DMS 2012 - 18th International Conference on Distributed Multimedia Systems*, pages 60–66. Knowledge Systems Institute Graduate School, 2012.
- [29] Oscar Corcho, Mariano Fernández-López, Asunción Gómez-Pérez, and Angel López-Cima. Building legal ontologies with METHONTOLOGY and WebODE. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 3369 LNAI:142–157, 2005.
- [30] Liviu-Adrian Cotfas, Camelia Delcea, Ioan Roxin, and Ramona Paun. Twitter Ontology-Driven Sentiment Analysis. In *New Trends in Intelligent Information and Database Systems*, pages 131–139, 2015.
- [31] Roddy Cowie and Randolph R Cornelius. Describing the emotional states that are expressed in speech. *Speech communication*, 40(1):5–32, 2003.
- [32] Munmun De Choudhury and Scott Counts. The nature of emotional expression in social media: measurement, inference and utility. *Human Computer Interaction . . .*, 2012.
- [33] Lipika Dey and Sk. Mirajul Haque. Opinion mining from noisy text data. *International Journal on Document Analysis and Recognition (IJ DAR)*, 12(3):205–226, aug 2009.
- [34] Shengchun Ding and Ting Jiang. Comment Target Extraction Based on Conditional Random Field & Domain Ontology. In *2010 International Conference on Asian Language Processing*, pages 189–192, 2010.
- [35] Xiaowen Ding, Xiaowen Ding, Bing Liu, Bing Liu, Philip S. Yu, and Philip S. Yu. A holistic lexicon-based approach to opinion mining. *Proceedings of the international conference on Web search and web data mining - WSDM '08*, page 231, 2008.
- [36] Xiaowen Ding and Bing Liu. The utility of linguistic rules in opinion mining. *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval SIGIR 07*, page 811, 2007.
- [37] T. Drummond. *Vocabulary of Emotions*.
- [38] Fabon Dzogang, Marie-Jeanne Lesot, Maria Rifqi, and Bernadette Bouchon-Meunier. Expressions of graduality for sentiments analysis—A survey. In *Fuzzy Systems (FUZZ), 2010 IEEE International Conference on*, pages 1–7. IEEE, 2010.
- [39] Paul Ekman. *All emotions are basic*. 1994.
- [40] Andrea Esuli and Fabrizio Sebastiani. Determining the semantic orientation of terms through gloss classification. In *Proceedings of the 14th ACM international conference on Information and knowledge management*, pages 617–624. ACM, 2005.
- [41] Christiane Fellbaum. *WordNet*. Wiley Online Library, 1998.
- [42] Virginia Francisco, Pablo Gervas, and Federico Peinado. *Ontological reasoning to configure emotional voice synthesis*. Springer, 2007.

- [43] Virginia Francisco, Raquel Hervás, Federico Peinado, Pablo Gervás, and Virginia Francisco. Raquel Hervás Federico Peinado Pablo Gervás. EmoTales: Creating a corpus of folk tales with emotional annotations. *Language Resources and Evaluation*, 46(3):341–381, 2012.
- [44] Virginia Francisco, F Peinado, R Hervás, and P Gervás. Semantic Web Approaches to the Extraction and Representation of Emotions in Texts, 2010.
- [45] Nico H Frijda. The Laws of Emotion. *American Psychologist*, 43(5):349–358, 1988.
- [46] Sujit Fulse, Rekha Sugandhi, and Anjali Mahajan. A survey on multimodal sentiment analysis. In *International Journal of Engineering Research and Technology*, volume 3. ESRSA Publications, 2014.
- [47] Aldo Gangemi and Valentina Presutti. Towards a pattern science for the Semantic Web. *Semantic Web*, 1(1-2):61–68, 2010.
- [48] A García-Crespo, Ricardo Colomo-Palacios, Myriam Mencke, and Juan M Gómez-Berbís. Cusent: Social sentiment analysis using semantics for customer feedback. *Social web evolution: integrating semantic applications and web*, 2, 2008.
- [49] Ángel García-Crespo, Ricardo Colomo-Palacios, Juan Miguel Gómez-Berbís, and Belén Ruiz-Mezcua. SEMO: A framework for customer social networks analysis based on semantics. *Journal of Information Technology*, 25(2):178–188, jun 2010.
- [50] Asunción Gómez-Pérez, Mariano Fernández-López, and Oscar Corcho. *Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*. Springer-Verlag, 2004.
- [51] Marco Grassi. Developing heo human emotions ontology. In *Biometric ID Management and Multimodal Communication*, pages 244–251. Springer, 2009.
- [52] Marco Grassi, Erik Cambria, Amir Hussain, and Francesco Piazza. Sentic Web: A New Paradigm for Managing Social Media Affective Information. *Cognitive Computation*, 3(3):480–489, may 2011.
- [53] Catherine Havasi, Robert Speer, and Jason Alonso. Conceptnet 3: a flexible, multilingual semantic network for common sense knowledge. In *Recent advances in natural language processing*, pages 27–29. Citeseer, 2007.
- [54] Elis Hernandes, Augusto Zamboni, Sandra Fabbri, and André Di Thommazo. Using GQM and TAM to evaluate StArt—a tool that supports Systematic Review. *CLEI ELECTRONIC JOURNAL*, 15(1), 2012.
- [55] Ian Horrocks. Ontologies and the semantic web. *Communications of the ACM*, 51(12):58, 2008.
- [56] Minqing Hu and Bing Liu. Mining and summarizing customer reviews. *Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '04*, 04:168, 2004.
- [57] Mus'ab Husaini, Andrea Ko, Dilek Tapucu, and Yücel Saygın. Ontology Supported Policy Modeling in Opinion Mining Process. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7567 LNCS:252–261, 2012.



- [58] Carroll E Izard. Emotion theory and research: highlights, unanswered questions, and emerging issues. *Annual review of psychology*, 60:1–25, 2009.
- [59] Hyun Duk Kim, Kavita Ganesan, Parikshit Sondhi, and ChengXiang Zhai. Comprehensive review of opinion summarization. *Illinois Environment for ...*, pages 1–30, 2011.
- [60] Barbara Kitchenham and S Charters. Guidelines for performing Systematic Literature Reviews in Software Engineering. *Engineering*, 2:1051, 2007.
- [61] Efstratios Kontopoulos, Christos Berberidis, Theologos Dergiades, and Nick Basiliades. Ontology-based sentiment analysis of twitter posts. *Expert Systems with Applications*, 40(10):4065–4074, aug 2013.
- [62] Akshi Kumar and Teeja Mary Sebastian. Sentiment Analysis: A Perspective on its Past, Present and Future. *International Journal of Intelligent Systems and Applications*, 4(10):1–14, 2012.
- [63] Fleur J M Laros and Jan Benedict E M Steenkamp. Emotions in consumer behavior: A hierarchical approach. *Journal of Business Research*, 58(10):1437–1445, 2005.
- [64] Raymond Y K Lau, Dawei Song, Yuefeng Li, Terence C H Cheung, and Jin-Xing Hao. Toward a fuzzy domain ontology extraction method for adaptive e-learning. *Knowledge and Data Engineering, IEEE Transactions on*, 21(6):800–813, 2009.
- [65] Raymond Y.K. Lau, Chapmann C.L. Lai, Jian Ma, and Yuefeng Li. Automatic Domain Ontology Extraction for Context-Sensitive Opinion Mining. In *ICIS 2009 Proceedings*, pages 35–53. AIS Electronic Library, 2009.
- [66] Douglas B Lenat, Ramanathan V. Guha, Karen Pittman, Dexter Pratt, and Mary Shepherd. Cyc: toward programs with common sense. *Communications of the ACM*, 33(8):30–49, 1990.
- [67] Warit Leopairote, Athasit Surarerks, and Nakornthip Prompoon. Software quality in use characteristic mining from customer reviews. In *2012 2nd International Conference on Digital Information and Communication Technology and its Applications, DICTAP 2012*, pages 434–439. IEEE, may 2012.
- [68] Maurizio Leotta, Silvio Beux, Viviana Mascardi, and Daniela Briola. My MOoD , a Multimedia and Multilingual Ontology Driven MAS : Design and First Experiments in the Sentiment Analysis Domain. In *CEUR Workshop Proceedings*, volume 1351, pages 51–66. CEUR-WS, 2015.
- [69] Sheng Li, Lingling Liu, and Zenggang Xiong. Ontology-Based Sentiment Analysis of Network Public Opinions. *International Journal of Digital Content Technology and its Applications*, 6(23):371–380, dec 2012.
- [70] Bing Liu. Sentiment analysis and subjectivity. *Handbook of natural language processing*, 2:627–666, 2010.
- [71] Bing Liu and South Morgan Street. Opinion Observer : Analyzing and Comparing Opinions on the Web. *Proceedings of the 14th International Conference on World Wide Web*, pages 342–351, 2005.
- [72] Lizhen Liu, Hao Liu, Hanshi Wang, Wei Song, and Xinlei Zhao. Generating domain-specific affective ontology from Chinese reviews for sentiment analysis. *Journal of Shanghai Jiaotong University (Science)*, 20(1):32–37, 2015.

- [73] Lizhen Liu, Xinhui Nie, and Hanshi Wang. Toward a fuzzy domain sentiment ontology tree for sentiment analysis. In *2012 5th International Congress on Image and Signal Processing*, pages 1620–1624. IEEE, oct 2012.
- [74] De-cheng Lou and Tian-fang Yao. Semantic polarity analysis and opinion mining on Chinese review sentences [J]. *Journal of Computer Applications*, 11:30–45, 2006.
- [75] Iris B. Mauss and Michael D Robinson. Measures of emotion: A review. *Cognition & Emotion*, 23(2):209–237, 2009.
- [76] Albert Mehrabian. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology*, 14(4):261–292, 1996.
- [77] George A Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.
- [78] Boris Motik and R Studer. Kaon2—a scalable reasoning tool for the semantic web. In *Proceedings of the 2nd European Semantic Web Conference (ESWC'05), Heraklion, Greece*, volume 17, 2005.
- [79] Subhabrata Mukherjee and Sachindra Joshi. Sentiment aggregation using concept-net ontology. In *Proceedings of the 6th International Joint Conference on Natural Language Processing, IJCNLP*, 2013.
- [80] Roberto Navigli and Simone Paolo Ponzetto. Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250, 2012.
- [81] Robert Neches, Richard E Fikes, Tim Finin, Thomas Gruber, Ramesh Patil, Ted Senator, and William R Swartout. Enabling technology for knowledge sharing. *AI magazine*, 12(3):36, 1991.
- [82] R. Nithish, S. Sabarish, M.Navaneeth Kishen, A.M. Abirami, and A. Askarunisa. An ontology based sentiment analysis for mobile products using tweets. In *2013 5th International Conference on Advanced Computing, ICoAC 2013*, pages 342–347. IEEE, dec 2014.
- [83] Daniel J Olsher. Full spectrum opinion mining: Integrating domain, syntactic and lexical knowledge. In *Data Mining Workshops (ICDMW), 2012 IEEE 12th International Conference on*, pages 693–700. IEEE, 2012.
- [84] Bo Pang and Lillian Lee. Opinion Mining and Sentiment Analysis, volume 2 (1–2) of Foundations and Trends in Information Retrieval. *Now Publ*, 2008.
- [85] Eun Hee Park, Veda C. Storey, and Shane Givens. An ontology artifact for information systems sentiment analysis. In *International Conference on Information Systems (ICIS 2013): Reshaping Society Through Information Systems Design*, volume 1, pages 431–449, 2013.
- [86] W Gerrod Parrott. *Emotions in social psychology: Essential readings*. Psychology Press, 2001.
- [87] Isidro Peñalver-Martínez, Francisco García-Sánchez, Rafael Valencia-García, Miguel Ángel Rodríguez-García, Valentín Moreno, Anabel Fraga, and Jose Luis Sánchez-Cervantes. Feature-based opinion mining through ontologies. *Expert Systems with Applications*, 41(13):5995–6008, oct 2014.

- [88] Isidro I.a Peñalver-Martínez, Rafael R.a Valencia-García, and Francisco F.b García-Sánchez. Ontology-guided approach to feature-based opinion mining. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6716 LNCS:193–200, 2011.
- [89] Rosalind W Picard. Affective Computing “MIT Media Laboratory Perceptual Computing Section Technical Report No. 321, 1995.
- [90] Rosalind W Picard. *Affective computing*. The MIT Press, 1st edition, 2000.
- [91] R Plutchik. *Emotion: A Psychoevolutionary Synthesis*. 1980.
- [92] Robert Plutchik. The nature of emotions human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*, 89(4):344–350, 2001.
- [93] Jantima Polpinij and Aditya K. Ghose. An ontology-based sentiment classification methodology for online consumer reviews. In *Proceedings - 2008 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2008*, pages 518–524. IEEE, December 2008.
- [94] Ana-Maria Popescu and Orena Etzioni. Extracting product features and opinions from reviews. In *Natural language processing and text mining*, pages 9–28. Springer, 2007.
- [95] Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. Opinion Word Expansion and Target Extraction through Double Propagation. *Computational Linguistics*, 37(1):9–27, 2011.
- [96] Ankit Ramteke. Detecting Turnarounds in Sentiment Analysis : Thwarting. In *ACL 2013 - 51st Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, volume 2, pages 860–865, 2013.
- [97] Diego Reforgiato Recupero, Valentina Presutti, Sergio Consoli, Aldo Gangemi, and Andrea Giovanni Nuzzolese. Sentilo: Frame-Based Sentiment Analysis. *Cognitive Computation*, 7(2):211–225, 2015.
- [98] Ismael Rivera, Myriam Mencke, Juan Miguel Gomez, Giner Alor-Hernandez, and Angel Garcia-Crespo. Collaborative OpenSocial Network Dataset Based Email Ranking and Filtering. *Third International Conference on Systems (icons 2008)*, pages 241–246, 2008.
- [99] Kin Meng Sam and Chris R. Chatwin. Ontology-Based Sentiment Analysis Model of Customer Reviews for Electronic Products. *International Journal of e-Education, e-Business, e-Management and e-Learning*, 3(6):477–n/a, dec 2013.
- [100] J. Fernando Sánchez-Rada and Carlos A. Iglesias. Onyx: Describing emotions on the web of data. In *CEUR Workshop Proceedings*, volume 1096, pages 71–82. CEUR-WS, 2013.
- [101] J. Fernando Sánchez-Rada, Marcos Torres, Carlos A. Iglesias, Roberto Maestre, and Raquel Peinado. A Linked Data Approach to Sentiment and Emotion Analysis of Twitter in the Financial Domain. In *Second International Workshop on Finance and Economics on the Semantic Web (FEOSW 2014)*, volume 1240, pages 51–62. CEUR-WS, 2014.

- [102] Francesco Sclano and Paola Velardi. TermExtractor: a Web Application to Learn the Common Terminology of Interest Groups and Research Communities. *Proceedings of the 9th Conference on Terminology and Artificial Intelligence (TIA 2007)*, pages 8–9, 2007.
- [103] Nigel Shadbolt, Wendy Hall, and Tim Berners-Lee. The semantic web revisited. *Intelligent Systems, IEEE*, 21(3):96–101, 2006.
- [104] Rob Shearer, Boris Motik, and Ian Horrocks. Hermit: A highly-efficient owl reasoner. In *OWLED*, volume 432, page 91, 2008.
- [105] Khin Phyu Phyu Shein. Ontology based combined approach for Sentiment Classification. *Technology*, pages 3–6, 2009.
- [106] Khin Phyu Phyu K.P.P. Shein and T.T.S. Thi Thi Soe Nyunt. Sentiment Classification Based on Ontology and SVM Classifier. In *Communication Software and Networks, 2010. ICCSN '10. Second International Conference on*, pages 169–172. IEEE, IEEE, 2010.
- [107] Wei Shi, Hongwei Wang, and Shaoyi He. EOSentiMiner: an opinion-aware system based on emotion ontology for sentiment analysis of Chinese online reviews. *Journal of Experimental & Theoretical Artificial Intelligence*, 27(4):423–448, oct 2015.
- [108] Marshall S. Smith, Daniel M. Ogilvia, Philip J. Stone, Dexter C. Dunphy, and John J. Hartman. The General Inquirer: A Computer Approach to Content Analysis., 1967.
- [109] Selver Softic and Michael Hausenblas. Towards Opinion Mining Through Tracing Discussions on the Web. In *Proceedings of the First Social Data on the Web Workshop*, volume 405, 2008.
- [110] Gamgarn Somprasertsri. Mining Feature-Opinion in Online Customer Reviews for Opinion Summarization. *Computer*, 16(6):938–955, 2010.
- [111] John F Sowa. Knowledge representation: logical, philosophical, and computational foundations. 1999.
- [112] Martin D. Sykora, Thomas W. Jackson, Ann O'Brien, and Suzanne Elayan. Emotive ontology: Extracting fine-grained emotions from terse, informal messages. In *Proceedings of the IADIS International Conference Intelligent Systems and Agents 2013, ISA 2013, Proceedings of the IADIS European Conference on Data Mining 2013, ECDM 2013*, pages 19–26, 2013.
- [113] Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2):267–307, 2011.
- [114] Dandibhotla Teja Santosh and Bulusu Vishnu Vardhan. Obtaining Feature- and Sentiment-Based Linked Instance RDF Data from Unstructured Reviews using Ontology-Based Machine Learning. *International Journal of Technology*, 6(2):198, 2015.
- [115] Mike Thelwall, Kevan Buckley, and Georgios Paltoglou. Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, 63(1):163–173, 2012.

- [116] R. Thiyagu and S. Sendhilkumar. Ontology based sentiment classification. In *Proceedings of the 5th Indian International Conference on Artificial Intelligence, IICAI 2011*, pages 1158–1169, 2011.
- [117] Peter D Turney and Michael L Littman. Measuring Praise and Criticism: Inference of Semantic Orientation from Association. *ACM Transactions on Information Systems*, 21(4):315–346, 2003.
- [118] Mingqiang Wang, Mengting Liu, Shi Feng, Daling Wang, and Yifei Zhang. A Novel Calibrated Label Ranking Based Method for Multiple Emotions Detection in Chinese Microblogs. In *Natural Language Processing and Chinese Computing*, pages 238–250, 2014.
- [119] Wei Wei and Jon Atle Gulla. Sentiment learning on product reviews via Sentiment Ontology Tree. *ACL 2010 - 48th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, (July):404–413, jul 2010.
- [120] Wei Wei and Jon Atle Gulla. Sentiment analysis in a hybrid hierarchical classification process. In *Seventh International Conference on Digital Information Management (ICDIM 2012)*, pages 47–55. IEEE, aug 2012.
- [121] Albert Weichselbraun, Stefan Gindl, and Arno Scharl. Extracting and Grounding Contextualized Sentiment Lexicons. *IEEE Intelligent Systems*, 28(2):39–46, mar 2013.
- [122] Adam Westerski, Carlos A Iglesias, and Fernando Tapia Rico. Linked Opinions : Describing Sentiments on the Structured Web of Data. In *Rand Journal of Economics*, volume 30, pages 1–21, 2011.
- [123] Casey Whitelaw, Navendu Garg, and Shlomo Argamon. Using appraisal taxonomies for sentiment analysis. *Proceedings of CIKM*, pages 625–631, 2005.
- [124] Mohd Ridzwan Yaakub, Yuefeng Li, and Yanming Feng. Integration of Opinion into Customer Analysis Model. In *2011 IEEE 8th International Conference on e-Business Engineering*, pages 90–95. IEEE, dec 2012.
- [125] Liang Yang, Hongfei Lin, and Yuan Lin. Sentiment analysis based on Chinese thinking modes. *Communications in Computer and Information Science*, 333 CCIS:46–57, 2012.
- [126] Pei Yin, Hongwei Wang, and Kaiqiang Guo. Feature–opinion pair identification of product reviews in Chinese: a domain ontology modeling method. *New Review of Hypermedia and Multimedia*, 19(1):3–24, mar 2013.
- [127] Xu Linhong Lin Hongfei Pan Yu and Ren Hui Chen Jianmei. Constructing the Affective Lexicon Ontology. *Journal of the China Society for Scientific and Technical Information*, 2:6, 2008.
- [128] Chengxiang Yuan, Yi Zhuang, and Haohong Li. Semantic based Chinese sentence sentiment analysis. In *Proceedings - 2011 8th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2011*, volume 4, pages 2099–2103. IEEE, jul 2011.
- [129] Lili Zhao and Chungping Li. Ontology Based Opinion Mining for Movie Reviews. In Dimitris Karagiannis and Zhi Jin, editors, *Analysis*, volume 5914 of *Lecture Notes in Computer Science*, pages 204–214. Springer, 2009.

- [130] Lina Zhou and Pimwadee Chaovalit. Ontology-supported polarity mining. *Journal of the American Society for Information Science and Technology*, 59(1):98–110, jan 2008.

