



COLECCIÓN CONOCIMIENTO CONTEMPORÁNEO

La palabra a través: diálogos entre el pensamiento, la palabra y el cuerpo

Coords.

Manuel Bermúdez Vázquez

Javier Martín Párraga

Dykinson, S.L.

LA PALABRA A TRAVÉS:
DIÁLOGOS ENTRE EL PENSAMIENTO,
LA PALABRA Y EL CUERPO



COLECCIÓN CONOCIMIENTO CONTEMPORÁNEO

LA PALABRA A TRAVÉS:
DIÁLOGOS ENTRE EL PENSAMIENTO,
LA PALABRA Y EL CUERPO

Coords.

MANUEL BERMÚDEZ VÁZQUEZ
JAVIER MARTÍN PÁRRAGA

Dykinson, S.L.

2023

LA PALABRA A TRAVÉS: DIÁLOGOS ENTRE EL PENSAMIENTO,
LA PALABRA Y EL CUERPO

Diseño de cubierta y maquetación: Francisco Anaya Benítez

© de los textos: los autores

© de la presente edición: Dykinson S.L.

Madrid - 2023

N.º 124 de la colección Conocimiento Contemporáneo

1ª edición, 2023

ISBN: 978-84-1170-145-7

NOTA EDITORIAL: Los puntos de vista, opiniones y contenidos expresados en esta obra son de exclusiva responsabilidad de sus respectivos autores. Dichas posturas y contenidos no reflejan necesariamente los puntos de vista de Dykinson S.L., ni de los editores o coordinadores de la obra.

Los autores asumen la responsabilidad total y absoluta de garantizar que todo el contenido que aportan a la obra es original, no ha sido plagiado y no infringe los derechos de autor de terceros. Es responsabilidad de los autores obtener los permisos adecuados para incluir material previamente publicado en otro lugar. Dykinson S.L. no asume ninguna responsabilidad por posibles infracciones a los derechos de autor, actos de plagio u otras formas de responsabilidad relacionadas con los contenidos de la obra. En caso de disputas legales que surjan debido a dichas infracciones, los autores serán los únicos responsables.

ÍNDICE

INTRODUCCIÓN.....	10
MANUEL BERMÚDEZ VÁZQUEZ	
JAVIER MARTÍN PÁRRAGA	

SECCIÓN I

LA PALABRA Y EL PENSAMIENTO

CAPÍTULO 1. DE LA AUTOFICCIÓN A LA NOVELA DE NO-FICCIÓN: DESPLAZAMIENTOS EN LA LITERATURA VASCA CONTEMPORÁNEA... 14	
BEÑAT SARASOLA	
CAPÍTULO 2. NUEVAS TENDENCIAS EN LAS REPRESENTACIONES AUDIOVISUALES: PERSONAJES MIGRANTES DENTRO DE LA COMUNIDAD LGBTI+. EL CASO DE <i>SMILEY</i> (GUILLEM CLUA, 2022) .. 31	
LUCÍA LERMA PARRA	
MARTA ZORNOZA MADRID	
CAPÍTULO 3. LA NARRATIVA TRANSMEDIA COMO RECURSO DIDÁCTICO PARA FOMENTAR LA LECTURA EN EDUCACIÓN SUPERIOR.....	47
ÁLVARO CLAVIJO CORCHERO	
CAPÍTULO 4. A EDUCAÇÃO DOS MAIS VELHOS: OPORTUNIDADES E DESAFIOS	65
LUÍSA CARVALHO	
ISABEL FERREIRA	
CAPÍTULO 5. EL CÓMIC EN L2: LA APLICACIÓN BDNF, LA FABRIQUE À BD PARA MEJORAR LAS DESTREZAS ESCRITAS EN ELE Y FLE	83
ÁLVARO CLAVIJO CORCHERO	
NURIA CABELLO ANDRÉS	
CAPÍTULO 6. EL MANIQUÍ EN LA NARRATIVA BREVE DE CORTE FANTÁSTICO DE HENRI DE RÉGNIER	101
NURIA CABELLO ANDRÉS	
CAPÍTULO 7. EL RETORNO DE MIGRANTES A TRAVÉS DE LA REPRESENTACIÓN FÍLMICA: EL CASO DE <i>THE RETURN</i> (MALENE CHOI, 2018)	118
LUCÍA LERMA PARRA	
CAPÍTULO 8. PERCEPCIÓN DE LOS DOCENTES DE MÚSICA SOBRE LAS METODOLOGÍAS QUE INCORPORAN ELEMENTOS TECNOLÓGICOS EN EDUCACIÓN PRIMARIA.....	134
PALOMA BRAVO FUENTES	

CAPÍTULO 9. ANIMALIDAD Y BIOPOLÍTICA EN LA LITERATURA ARGENTINA DEL SIGLO XX.....	148
XIMENA VENTURINI	
CAPÍTULO 10. CÓMO OPTIMIZAR LOS RESULTADOS DE UN SISTEMA DE RECUPERACIÓN DE INFORMACIÓN CON LA ONTOLOGÍA DE FUNGRAMKB. DISEÑO PILOTO DE UN BUSCADOR DE TÉRMINOS CULINARIOS	160
NOELIA ESTÉVEZ-RIONEGRO	
CAPÍTULO 11. TRAYECTOS JUVENILES HACIA EL EMPLEO: CAUSAS Y CONSECUENCIAS DEL DESENGANCHE EDUCATIVO.....	182
JUAN GARCÍA-FUENTES	
MARÍA NAVARRO-GRANADOS	
CAPÍTULO 12. ¿RESPONDE LA EDUCACIÓN COLOMBIANA A LOS PROBLEMAS DEL PRESENTE?	202
ANDRÉS FELIPE SANMARTÍN SANMARTÍN	
MATEO GIL DÍAZ	
CAPÍTULO 13. EL JUEGO CON LO REAL Y EL ESPACIO MADRILEÑO: CLAVES DE NO-FICCIÓN EN LA OBRA DE ELÍAS LEÓN SIMINIANI..	217
ELIOS MENDIETA	
LUIS DELTELL ESCOLAR	
CAPÍTULO 14. REVISIÓN DE RECURSOS LITERARIOS PARA TRABAJAR LA DIVERSIDAD FAMILIAR EN EDUCACIÓN PRIMARIA (6-12 AÑOS)	232
PAULA PEREGRINA NIEVAS	
CARMEN DEL PILAR GALLARDO MONTES	
JESÚS ESTEBAN MORA	
CAPÍTULO 15. PROFESORADO DE PRIMARIA Y SUS PERCEPCIONES SOBRE DIVERSIDAD EFECTIVO-SEXUAL	250
JESÚS ESTEBAN MORA	
PAULA PEREGRINA NIEVAS	
CARMEN DEL PILAR GALLARDO MONTES	
CAPÍTULO 16. EVALUACIÓN DEL ANÁLISIS DE RIESGOS EN EL PATRIMONIO ECLESIAÍSTICO. LA IGLESIA DE SANTO DOMINGO EN SANLÚCAR DE BARRAMEDA.....	265
INÉS FLORES FERNÁNDEZ	
CAPÍTULO 17. EL APRENDIZAJE DE LA ACENTUACIÓN Y EL USO DE LA TILDE EN LA ENSEÑANZA DEL ESPAÑOL COMO SEGUNDA LENGUA LE/L2: UN PROYECTO DE INVESTIGACIÓN-ACCIÓN EN AULAS UNIVERSITARIAS CON UN ENTORNO HÍBRIDO, VIRTUAL O EN LÍNEA	289
ÓSCAR RUIZ HERNÁNDEZ	

CAPÍTULO 18. LA CULTURA DE LA VIOLACIÓN EN LA FICCIÓN AUDIOVISUAL NORTEAMERICANA DE FINALES DEL SIGLO XX.....	305
JAVIER MARTÍN PÁRRAGA	
CAPÍTULO 19. EL DESAFÍO DE LA EDUCACIÓN EN EQUIDAD.....	317
MARTA ROJANO SIMÓN	
CAPÍTULO 20. EL CASO DE <i>POLLY</i> : LA CARA Y LA CRUZ DE LA TRADUCCIÓN OPERÍSTICA	329
MIRIAN SOLEDAD TRIGUEROS-DE-LA-FUENTE	
CAPÍTULO 21. APRENDIZAJE EXPERIENCIAL UNIVERSITARIO BASADO EN LA IDENTIFICACIÓN DE RECURSOS TERRITORIALES PARA DIVERSIFICAR LA OFERTA TURÍSTICA DE LA CIUDAD DE GRANADA (ESPAÑA)	344
BELÉN PÉREZ-PÉREZ	
LAURA PORCEL-RODRÍGUEZ	
CAPÍTULO 22. EL LIDERAZGO COGNITIVO EN EL/LA DOCENTE DE EDUCACIÓN SUPERIOR A TRAVÉS DE LA COMUNICACIÓN DIDÁCTICA.....	362
MARÍA DEL MAR CAMÚS FERRI	
MARCOS JESÚS IGLESIAS MARTÍNEZ	
INÉS LOZANO CABEZAS	
CAPÍTULO 23. HABILIDADES Y COMPETENCIAS DOCENTES FRENTE A LOS NUEVOS RETOS EDUCATIVOS.....	380
CRISTINA PINEL MARTÍNEZ	
MARÍA DOLORES PÉREZ ESTEBAN	
JOSEFA MARTÍNEZ TALAVERA	
CAPÍTULO 24. LAS MÚLTIPLES CARAS DE LA LITERATURA EN LA OBRA DE JONÁS TRUEBA.....	398
ELIOS MENDIETA	
CAPÍTULO 25. EDUCACIÓN PARA LA SOSTENIBILIDAD A TRAVÉS DE LAS ARTES	412
SILVANA LONGUEIRA MATOS	
NELLY FORTES GONZÁLEZ	
STEFANY SANABRIA FERNÁNDES	
CAPÍTULO 26. LA EDUCACIÓN CÍVICA Y LA EDUCACIÓN INTERCULTURAL: DOS TIPOLOGÍAS EDUCATIVAS CLAVES EN LA CONFIGURACIÓN DE LA CONVIVENCIA PACÍFICA.....	426
MARÍA GONZÁLEZ BLANCO	
SILVANA LONGUEIRA MATOS	

CAPÍTULO 27. SUBTITULAR LOS MEDIOS AUDIOVISUALES HACE QUE EL APRENDIZAJE DE LENGUAS EXTRANJERAS SEA MÁS FÁCIL Y AMENO: EL CASO DEL APRENDIZAJE DE CHINO PARA LOS HISPANOHABLANTES	442
JIA LIN	
CAPÍTULO 28. FORMACIÓN DEL PROFESORADO Y ALFABETIZACIÓN MEDIÁTICA EN EL GRADO DE EDUCACIÓN INFANTIL.....	459
SILVIA MEDINA QUINTANA	
MIGUEL JESÚS LÓPEZ SERRANO	
CAPÍTULO 29. LA MEMORIA Y EL OLVIDO EN LA NARRATIVA DE JORDI SOLER.....	473
JOAQUÍN JUAN PENALVA	
MARÍA SAMPER CERDÁN	
CAPÍTULO 30. LAS CANCIONES DEL CINE INFANTIL EN EL AULA DE EDUCACIÓN MUSICAL	487
PALOMA BRAVO FUENTES	
CAPÍTULO 31. LA COMPETENCIA DIGITAL DEL DOCENTE UNIVERSITARIO ESPAÑOL.....	505
JUAN-FRANCISCO ÁLVAREZ-HERRERO	
CAPÍTULO 32. EL USO DE LA GAMIFICACIÓN COMO HERRAMIENTA PARA EDUCAR HÁBITOS SALUDABLES EN ESTUDIANTES UNIVERSITARIOS DE EDUCACIÓN PRIMARIA	519
RUBEN CAMACHO-SÁNCHEZ	
JOSÉ MIGUEL RODRÍGUEZ-FERRER	
ANA MANZANO-LEÓN	
PAULA RODRÍGUEZ-RIVERA	

SECCIÓN II CUERPO Y DEPORTE

CAPÍTULO 33. TEACHING BASKETBALL THROUGH THE TAXONOMY OF OBJECTIVES.....	540
PABLO CAMACHO LAZARRAGA	
JOSÉ HORNO TOMÉ	
CAPÍTULO 34. “NEUROCIENCIA Y ACTIVIDADES EN LA NATURALEZA”	558
JOSE J. HORNO TOMÉ	
JULIO FUENTESAL GARCÍA	
ALBERTO MARTÍN BARRERO	
CAPÍTULO 35. EL ENTORNO SOCIAL EN LA FORMACIÓN DEL JUGADOR DE FÚTBOL. INFLUENCIA DE LOS PADRES.....	573
ALBERTO MARTÍN BARRERO	
PABLO CAMACHO LAZARRAGA	

CAPÍTULO 36. PROYECTO DE INTERVENCIÓN EDUCATIVA: “REDUCIR LA DESHIDRATACIÓN EN LA ESCUELA”	586
<p style="text-align: center;">ÁLVARO MORENTE MONTERO JUAN DE DIOS BENÍTEZ SILLERO CRISTINA CASTEJÓN RIBER</p>	
CAPÍTULO 37. PERCEPCIÓN DEL USO DE JUEGOS Y ACTIVIDADES FÍSICO-EXPRESIVAS PARA LA EDUCACIÓN ALIMENTARIA Y NUTRICIONAL EN CLASES DE EDUCACIÓN FÍSICA DEL FUTURO PROFESORADO	605
<p style="text-align: center;">CRISTINA CASTEJÓN-RIBER ÁLVARO MORENTE JUAN DE DIOS BENÍTEZ-SILLERO</p>	
CAPÍTULO 38. PREVENCIÓN DEL CONSUMO DE DROGAS EN LOS ADOLESCENTES Y JÓVENES. PUEBLO JOVEN MIGUEL GRAU, PAUCARPATA, AREQUIPA - PERÚ	630
<p style="text-align: center;">JESSICA YESENIA GARCÍA APAZA LUIS DANIEL HUAMÁN ASILLO GLORIA ISABEL MONZÓN ÁLVAREZ</p>	
CAPÍTULO 39. LAS PELÍCULAS EN LA CLASE DE CULTURA PARA ESTUDIANTES DE TRADUCCIÓN: ANÁLISIS DE LA REPRESENTACIÓN DE LAS MASCULINIDADES NEGRAS EN EL CINE ESTADOUNIDENSE.....	648
<p style="text-align: center;">ELISA SERNA MARTÍNEZ</p>	
CAPÍTULO 40. LOS ORÍGENES DEL ACTUAL MODELO EDUCATIVO ESPAÑOL	661
<p style="text-align: center;">ADRIÁN GARCÍA SERRANO</p>	
CAPÍTULO 41. FOMENTO DE LA ACTITUD CRÍTICA Y PENSAMIENTO VISIBLE EN EDUCACIÓN PRIMARIA. EDUCACIÓN AMBIENTAL Y REDES SOCIALES	679
<p style="text-align: center;">MARÍA DEL MAR MONTALVO GARCÍA</p>	
CAPÍTULO 42. COORDINACIÓN INTERDISCIPLINAR EN EL GRADO DE EDUCACIÓN INFANTIL: MEJORAS Y OPORTUNIDADES.....	694
<p style="text-align: center;">VIRGINIA MORCILLO LORO MAR OLIVER-BARCELÓ MARIA FERRER-RIBOT MARÍA ROSA ROSSELLÓ RAMON</p>	
CAPÍTULO 43. SUICIDE: KNOWLEDGE AND FALSE MYTHS IN EDUCATION STUDENTS.....	715
<p style="text-align: center;">NOELIA NAVARRO GÓMEZ</p>	

CÓMO OPTIMIZAR LOS RESULTADOS DE UN SISTEMA DE RECUPERACIÓN DE INFORMACIÓN CON LA ONTOLOGÍA DE FUNGRAMKB. DISEÑO PILOTO DE UN BUSCADOR DE TÉRMINOS CULINARIOS

NOELIA ESTÉVEZ-RIONEGRO
Universidade de Santiago de Compostela

1. INTRODUCCIÓN

La mayor fuente de información que existe en el mundo, en el siglo XXI, sigue siendo de carácter textual, si bien en gran parte está digitalizada o surge ya en un entorno digital. Esto lleva al diseño de nuevos motores de búsqueda y a la selección de nuevos escenarios de aplicación (desde bibliotecas y archivos, hasta depósitos documentales de instituciones y empresas, o cualquier otra fuente en la que se almacenen conjuntos de textos).

Con la finalidad de mejorar la calidad de las consultas y las respuestas del sistema, los Sistemas de Recuperación de Información (en lo sucesivo, SRI) se encuentran en un permanente proceso de desarrollo y optimización. Con el auge de la Lingüística Computacional, cada vez son más estrechas las sinergias entre profesionales de una y otra disciplina, que contribuyen a fortalecer herramientas TIC de forma conjunta. De estas sinergias surgen nuevas herramientas como FunGramKB⁴⁰, concebida con una finalidad lingüística, pero pensada para la implementación y optimización de recursos informáticos. Tomándola como base, se tratará de demostrar, a lo largo de este trabajo, su utilidad como soporte de un SRI.

⁴⁰ Consultable en <http://www.fungramkb.com>

2. OBJETIVOS

El SRI que proponemos pretende ser una herramienta de búsqueda de términos que designan alimentos y que constituyen ingredientes principales de recetas de cocina. El propósito del buscador es que un usuario pueda recuperar documentos a partir de consultas léxicas, pero la particularidad de su diseño y funcionamiento es que se apoya en una base ontológica que permite discriminar casos de polisemia, homonimia o cualquier otra coincidencia gráfica a partir de una sencilla anotación conceptual de los términos.

El escenario de aplicación elegido, por tanto, para ilustrar cómo una base ontológica conceptual puede contribuir a mejorar considerablemente los resultados de un SRI, es un conjunto de recetarios digitales alojado en un blog de actualidad especializado en cocina. Si bien un buscador común permitiría realizar consultas por coincidencias de palabra, lo que se propone en este trabajo es poder realizar búsquedas a partir de sustantivos que designan alimentos. Para ello, el buscador ha de diseñarse teniendo en cuenta determinados parámetros semánticos y conceptuales, además de otros de índole gramatical. Por eso, resulta fundamental el empleo de una base ontológica en la que pueda sustentarse el SRI.

La FunGramKB, con sus múltiples funciones multilingües, es una herramienta óptima para este propósito porque, como se tratará de demostrar, permite optimizar la calidad de las consultas y la eficacia en la recuperación de la información. De este modo, y a grandes rasgos, la principal diferencia con respecto a un buscador común es que el que aquí se presenta realiza búsquedas conceptuales (es decir, de palabras que contengan determinado concepto en su esquema semántico), en lugar de búsquedas de palabras exactas. El SRI es capaz, por tanto, de discriminar coincidencias léxicas que no pertenezcan al campo conceptual del término buscado (por ejemplo, *kiwi* ‘pájaro’ vs. *kiwi* ‘fruta’).

3. MARCO TEÓRICO

De acuerdo con autores como Cleverdon (1991) o Liddy (2005), la recuperación de información mediante búsquedas por ordenador es una

práctica que comienza en torno a 1940 y que, desde muy pronto, suscita gran interés y demanda, especialmente para la recuperación automática de documentos. Sin embargo, las limitaciones de la época (como la escasa disponibilidad de ordenadores o de investigaciones científico-técnicas) no permitían mucho más que la recuperación de documentos a partir de los autores, los títulos o las palabras clave. Los avances en cuanto a la recuperación de texto completo se producen más adelante y de forma progresiva.

Bush (1945) contribuye a este propósito con una inspiradora idea: la creación de un dispositivo de almacenaje de todos los libros, registros y comunicaciones de un individuo, que esté mecanizado, de modo que se puedan hacer consultas de forma rápida y flexible. A estos dispositivos, a modo de archivos o bibliotecas privadas mecanizadas, el autor los denomina, al azar, *memex*. Aunque existía como concepto, el término Information Retrieval no aparece, sin embargo, hasta 1948-1950, cuando es acuñado por Mooers (1950).

Desde finales de los años 50 y principios de los 60 se produce un intenso debate en torno a las tecnologías óptimas para el diseño de los sistemas de recuperación de información. La controversia vino suscitada a raíz de las demostraciones de unas máquinas de autoindexación creadas por IBM a partir del trabajo de Luhn (creador del algoritmo o fórmula de Luhn), que atrajeron el interés comercial hacia los sistemas de recuperación booleanos, a la vez que provocaron la discrepancia de algunos especialistas de la talla de Mooers, que consideraban el diseño basado en el álgebra de Boole (1847) completamente erróneo.

Más recientemente, Witten, Paynter, Frank, Gutwin y Nevill-Manning (1999) realizan una detallada comparación sobre la eficiencia espacial y temporal del índice invertido frente a otras estructuras de datos. En la misma línea, aunque más actualizado, se encuadra el trabajo de Zobel y Moffat (2006).

En la actualidad, de acuerdo con Nadkarni, Ohno-Machado y Chapman (2011), la tecnología RI forma parte de la oferta de los sistemas de gestión de bases de datos relacionales (RDBMS) de los principales proveedores, que ofrecen, además, una amplia variedad de métodos que

permiten tanto la consulta integrada como la recuperación de información de bases de datos que contienen datos estructurados y no estructurados. Sin embargo, los RDBMS todavía no alcanzan el grado de evolución de los RI, que permiten la indexación de textos por proximidad y por sinónimos, y en una mayor variedad de formatos que los anteriores.

Paralelamente al debate en torno al diseño de los sistemas de recuperación de información, otros autores tratan de cubrir algunas lagunas en la optimización de los patrones de búsqueda, como es el uso práctico de las expresiones regulares (*regex*), que aporta Friedl (2006). Se trata de un sistema que todavía se utiliza hoy en día para buscar, por ejemplo, subcadenas concretas o para corregir palabras mal escritas (ya que permite hacer sustituciones de caracteres). Ruano-Ordás, Fdez-Riverola y Méndez (2018), por su parte, proponen el empleo de la computación evolutiva para identificar automáticamente patrones comunes en conjuntos de textos. El procedimiento consiste en procesar textos de una categoría (A) y textos de otra categoría (B), de modo que se generen patrones que sean comunes solo en textos de A y no de B, y viceversa. Generar un patrón no significa que este aparezca en todos los documentos, sino que solo se generan patrones que permiten discernir correctamente las dos categorías. Sin embargo, los autores señalan la problemática de que, actualmente, se aplica al SPAM (categoría A) y al HAM (categoría B), por lo que presenta la limitación de que solo puede hacerse para dos clases.

El hecho de que la principal fuente de información existente en el mundo sea todavía textual propicia la evolución e innovación de las herramientas de RI para optimizarlas y lograr que su aplicación sea más productiva. Por ejemplo, la disponibilidad de grandes bases de datos como Pub-Med o de vocabularios como UMLS ha facilitado mucho la labor de los investigadores en RI a la hora de idear y probar enfoques escalables que resulten útiles a los investigadores de laboratorio y contribuyan a su labor.

La aplicación de las tecnologías de RI es muy rentable en la actualidad para la minería de datos, en general, y el procesamiento del lenguaje natural, en particular. No obstante, todavía quedan mejoras por realizar que suponen los retos del futuro. Todos los modelos clásicos de RI, esto

es, el booleano, el probabilístico y el vectorial, presentan ventajas e inconvenientes, como describe sucintamente Cacheda (2008). De acuerdo con el autor, el modelo booleano, basado en la teoría de conjuntos, tiene como principal ventaja la sencillez, pero impide ordenar los resultados obtenidos y no tiene en cuenta el número de ocurrencias de una palabra en un documento. El modelo probabilístico, basado en un proceso iterativo, permite representar el proceso de recuperación de información y ordenar los documentos según su probabilidad de importancia; sin embargo, presenta una importante desventaja, porque requiere partir siempre de una primera estimación del conjunto de documentos relevantes. Además, como el modelo booleano, tampoco permite conocer el número de ocurrencias de un término en un documento. Por su parte, el modelo vectorial, que representa los documentos a modo de vectores de términos y los recupera en función de su similitud con el vector de consulta, resulta eficiente para colecciones de documentos de gran magnitud. Permite aciertos parciales, porque puede seleccionar como relevantes documentos que no incluyen todos los términos de la consulta. Sin embargo, este modelo pierde parte de la información sintáctica y semántica del documento, pues se fundamenta en la independencia de los términos dentro del mismo.

En definitiva, las mejoras en los sistemas de recuperación de información realizadas hasta el momento son innegables, pero las desventajas señaladas evidencian la necesidad de seguir innovando y trabajando por perfeccionar y optimizar estas herramientas. Concretamente, en lo que concierne a las funciones de búsqueda, cabe prestar atención a la conexión entre la computación y la lingüística y, a través de sólidas sinergias entre especialistas de ambas disciplinas, tratar de mejorar los patrones en base a criterios no solo informáticos sino también léxicos, gramaticales y discursivos. Como señalan Perrián-Pascual y Mairal-Usón (2010), los sistemas informáticos de comprensión de lenguaje natural necesitan “una base de conocimiento provista de representaciones conceptuales que reflejen el sistema cognitivo de los seres humanos” (2010: 11) En la actualidad, existen herramientas lingüísticas informatizadas que pueden resultar muy útiles para este propósito, como FunGramKB, que contiene una base de conocimiento conceptual en la que podrían

fundamentarse los buscadores automáticos para recuperar información a partir de búsquedas semánticas o conceptuales.

En realidad, la herramienta FunGramKB surge como un proyecto más ambicioso, que pretende dar soporte conceptual a diferentes tareas de procesamiento del lenguaje natural (*vid.* Perrián-Pascual y Arcas-Túnez, 2010) y servir como recurso para optimizar los resultados de otras herramientas o realizar implementaciones sobre ellas (como los mencionados sistemas de recuperación de información, así como lematizadores, desambiguadores y otras herramientas para el tratamiento automático de textos). Una descripción completa y detallada de la herramienta puede ser la que sigue:

FunGramKB Suite es un entorno computacional destinado a la construcción semiautomática de una base de conocimiento léxico-conceptual multipropósito para el desarrollo de sistemas automatizados del procesamiento del lenguaje natural (PLN). Por una parte, FunGramKB es multipropósito en el sentido de que es tanto multifuncional como multilingüe. En otras palabras, FunGramKB ha sido diseñado con el fin de ser potencialmente reutilizado en diversas tareas del PLN (recuperación y extracción de información, traducción automática, sistemas basados en el diálogo, etc.) y con diversas lenguas. Por otra parte, nuestra base de conocimiento comprende tres niveles principales de conocimiento (i.e. léxico, gramatical y conceptual), cada uno de los cuales está constituido por diversos módulos independientes aunque interrelacionados. (Perrián-Pascual y Mairal-Usón, 2010: 12-13)

Los esquemas conceptuales en los que se sustenta la base de conocimiento de FunGramKB están estructurados en tres módulos: la *Ontología* (que recoge el conocimiento semántico representado por un conjunto de proto-microestructuras), el *Cognición* (que recoge el conocimiento procedimental representado por un conjunto de proto-macroestructuras) y el *Onomasticón* (que recoge el conocimiento episódico representado por un conjunto de bio-estructuras). Lo que interesa, pensando en la mejora y optimización de un sistema de recuperación de información, es la información conceptual que se puede extraer de la ontología, esto es, las denominadas proto-macroestructuras o postulados de significado.

Teniendo en cuenta los postulados de Tulving (1985), que considera que el modo de facilitar la representación y la recuperación de la información es que las memorias semántica, procedimental y episódica trabajen

conjuntamente, una de las claves a la hora de desarrollar un motor de razonamiento en un sistema para el procesamiento del lenguaje natural es el empleo de un único lenguaje de interfaz para representar cada uno de los esquemas conceptuales señalados. Por esa razón, “FunGramKB postula COREL (*Conceptual Representation Language*) como lenguaje de lenguaje de representación conceptual común a los tres tipos de conocimiento” (Periñán-Pascual y Mairal-Usón, 2010). Pero, además, FunGramKB implementa este lenguaje (a diferencia de otras bases de conocimiento, como EuroWordnet, donde los conceptos se definen en relación a otros conceptos), de modo que no solo se definen las unidades conceptuales, sino que estas definiciones se ven enriquecidas al aplicarse sobre las proto-microestructuras los mecanismos de herencia e inferencia (*vid.* Periñán-Pascual y Arcas-Túnez, 2005).

Como todo lenguaje de representación, COREL distingue dos componentes: una colección de términos (unidades conceptuales jerárquicamente organizadas, lo que permite la herencia) y un sistema de anotación (la sintaxis a la que han de ajustarse las interpretaciones interlingüísticas). Como describen Periñán-Pascual y Mairal-Usón (2010), FunGramKB contiene tres tipos de unidades, cada una de las cuales se representa con un sistema de anotación diferente: los metaconceptos (unidades de nivel superior que coinciden con las de otras ontologías) se representan precedidos por el símbolo #; los conceptos básicos (identificados, primero, a partir del *Longman Dictionary of Contemporary English* de Procet y, posteriormente, conceptualizados, jerarquizados, remodelados y refinados) se representan con el símbolo +, y los conceptos terminales (nodos finales de la estructuración jerárquica) se representan con el símbolo \$. Así, de esta anotación, resultaría la siguiente jerarquía ontológica de la palabra *revision*:

```
#ENTITY > #PHYSICAL > #PROCESS > +OCCURRENCE_00 > $REVISION_00
```

Para el trabajo que planteamos, como se mostrará en el siguiente apartado, interesan, principalmente, los conceptos básicos. Estos, junto con los terminales, presentan tres características en la ontología de FunGramKB que resultan de especial relevancia (*vid.* Periñán-Pascual y Mairal-Usón, 2010): (i) la mayoría de los conceptos se entienden como

“conceptos de motivación léxica” (entendiendo la semántica léxica como el estudio de la estructuración conceptual), (ii) no son primitivos semánticos (como lo eran, por ejemplo, los formulados por Goddard y Wierzbicka (2002) en el Metalenguaje Semántico Natural) y (iii) tienen dos propiedades semánticas: los marcos temáticos y los postulados de significado, y ambas poseen una base conceptual, puesto que no se construyen con unidades léxicas sino con conceptos.

4. METODOLOGÍA

El SRI diseñado está basado en el modelo booleano, ya que es el más comúnmente adoptado, además de ser fácil de implementar y fundamentarse en conceptos intuitivos. Además, resulta útil para recuperar tanto un gran número de documentos como una pequeña cantidad de ellos, por lo que es adecuado para la magnitud del *dataset* empleado en nuestra propuesta de SRI.

Para realizar el SRI, se ha partido de un conjunto de datos (*dataset*) en formato JSON, que está compuesto por más de 125000 recetas distintas, extraídas del blog Eight Portions⁴¹ de Lee (2021). Se ha elegido este *dataset* por varias razones: (i) sigue un formato estándar JSON que permite extraer la información de forma sencilla, pues es soportado por todos los lenguajes de programación; (ii) su magnitud resulta lo suficientemente representativa, dado el número de recetas que contiene, lo que permite generar una base de conocimiento eficaz; (iii) está actualizado y contiene ejemplos de la lengua contemporánea en uso; (iv) al estar escrito en lengua inglesa permite hacer búsquedas en FunGramKB sin recurrir a una traducción previa.

Cada entrada del fichero JSON representa una receta. Al mismo tiempo, y según la definición del estándar del JSON, cada receta está formada por un identificador único (formado por 31 caracteres alfanuméricos) y tres pares atributo-valor: (i) nombre de la receta o *title*, (ii) ingredientes necesarios o *ingredients*, y (iii) descripción con los pasos a seguir para

⁴¹ Consultable en <https://eightportions.com/>

realizar el plato o *instructions*. Veamos un ejemplo en la imagen que sigue:

TABLA 1. Ejemplo de entrada del fichero JSON

```
"rmK12Uau.ntP510KelmX506H6Mr6jTu": {
  "title": "Slow Cooker Chicken and Dumplings",
  "ingredients": [
    "4 skinless, boneless chicken breast halves",
    "2 tablespoons butter",
    "2 (10.75 ounce) cans condensed cream of chicken soup",
    "1 onion, finely diced",
    "2 (10 ounce) packages refrigerated biscuit dough, torn into pieces"
  ],
  "instructions": "Place the chicken, butter, soup, and onion in a slow cooker, and fill with enough water to cover.\nCover, and cook for 5 to 6 hours on High. About 30 minutes before serving, place the torn biscuit dough in the slow cooker. Cook until the dough is no longer raw in the center.\n",}
```

Fuente: elaboración propia

El objetivo es poder realizar búsquedas en las recetas en función de los ingredientes. Por ello, los campos de título y descripción se almacenan en bruto (*raw data*), mientras que la información contenida en el campo de los ingredientes, que será utilizada como base de conocimiento para la creación del SRI, debe ser previamente procesada.

De este modo, la base de conocimiento está formada por una matriz (tabla) de n-filas por m-columnas (NxM), donde la primera y la última columna contienen el nombre y la descripción, mientras que las restantes m-2 columnas almacenan la existencia (o no) de un determinado ingrediente en la n-ésima receta. El proceso de construcción de la matriz se fundamenta en recorrer iterativamente todas las recetas incluidas en el *dataset* (sección) con el objetivo de obtener el nombre de la receta (valor del atributo “title”), los ingredientes (valor del atributo “ingredients”) y la descripción (valor del atributo “instructions”).

Como se puede observar, el atributo “ingredients” no solo contiene sustantivos que designan los alimentos que se van a utilizar para la realización de los platos (*cebolla, zanahoria...*), sino que también contiene sustantivos cuantificadores (*cucharada, bol, pizza...*) y verbos en participio (*cortado, triturado*), que no son necesarios para el tipo de búsqueda que se quiere realizar. Por ejemplo, en la imagen anterior, puede leerse

“tablespoons” (*cucharada*) o “onion, finely diced” (*cebolla cortada fina*). Este tipo de palabras son discriminadas porque no tienen el contenido semántico y conceptual que requiere el tipo de buscador que se pretende diseñar, donde solo interesa el léxico, cuyo valor semántico puede enmarcarse en el concepto “food”, de acuerdo con la jerarquía ontológica de la FunGramKB. De este modo, solo se pre-procesan aquellos sustantivos que designan alimentos.

En la fase de preprocesamiento de los datos, se ha utilizado el método de tokenización (usando tanto espacios y caracteres de puntuación como delimitadores de palabras), que permite aislar cada una de las palabras y tratarlas como entidades únicas. Seguidamente, es necesario eliminar los *tokens* que contienen dígitos o signos de puntuación dobles (comillas, paréntesis, etc.). Esto puede lograrse con la siguiente expresión regular: `/^[a-z']*/i`

La expresión regular construida está dividida en dos partes, tal y como se determina en el estándar PCRE2⁴². En la primera parte, se define el patrón de coincidencia (*regex matching pattern*), mientras que la segunda sirve para establecer comportamientos especiales de la expresión regular. Concretamente, en el caso que nos ocupa, el patrón de coincidencia permite seleccionar todas aquellas palabras que contengan caracteres alfabéticos o apóstrofes intercalados entre ellos. Se ha incluido el argumento *y* en la expresión regular para indicar que no se diferencien letras mayúsculas de minúsculas (*case insensitive*).

A continuación, es necesario realizar una selección de las palabras léxicas o, lo que es lo mismo, eliminar las palabras que solo tienen valor gramatical (como determinantes, preposiciones, etc.). Una forma de llevar a cabo esta selección de palabras es a través de un *stopwords removal*, para el que se pueden emplear diccionarios como el elaborado por Savand (2018)⁴³, que posee un listado con 1298 *stopwords*.

Seguidamente, se lematizan los sustantivos que se hayan localizado, de modo que las palabras flexionadas (por ejemplo, en plural) o derivadas

⁴² Consultable en <https://www.pcre.org/current/doc/html/>

⁴³ Consultable en <https://github.com/Alir3z4/stop-words/blob/master/english.txt/>

(por ejemplo, con un sufijo) puedan relacionarse con su forma canónica o lema. De este modo, los vocablos *tomates* o *tomatito* estarían asociados al lema *tomate*, y podrían recuperarse al hacer una búsqueda basada en este ingrediente. Cabe señalar, no obstante, que la lematización sería más productiva en castellano que en inglés, ya que no es habitual encontrar en esta lengua sustantivos derivados por sufijación o prefijación referidos a los alimentos. De este modo, y aplicado a la lengua inglesa, la lematización sería más útil para las palabras flexionadas que para las derivadas.

Una vez seleccionado y lematizado el léxico, se emplea la FunGramKB para poder discriminar todos aquellos *tokens* que no sean sustantivos. Sin embargo, tal y como se mencionó con anterioridad, no todos los sustantivos son válidos para este cometido, pues solo se necesitan aquellos que designan alimentos propiamente dichos. Consultar la FunGramKB permite conocer, gracias a la ontología que incorpora, si un determinado sustantivo contiene, en su esquema semántico, la entrada conceptual **+FOOD_00**. En caso afirmativo, el sustantivo se almacena en la base de conocimiento y, en caso contrario, se ignora.

Cuando un término es válido (es decir, designa un alimento), se comprueba si está ya incluido en la base de conocimiento o si es un término nuevo. En caso de que el término (atributo) ya exista previamente, se registra su presencia indicando un 1 en la casilla intersección entre la instancia y el atributo. Por el contrario, si se trata de un término nuevo, se añade poniendo un 1 en la casilla correspondiente (intersección instancia-atributo), así como 0 en las casillas correspondientes a las *i*-ésimas instancias anteriores. Con esto, se asegura que todas las instancias contengan la misma cantidad de atributos.

Finalmente, una vez que la totalidad del *dataset* se ha indexado en la base de conocimiento del SRI, ya se dispone de un sistema habilitado para la realización de consultas. Con el fin de obtener resultados flexibles y personalizados, el SRI propuesto permitirá realizar consultas que engloben múltiples ingredientes unidos mediante los conectores booleanos AND, OR y el conector de negación NOT.

5. RESULTADOS

A lo largo del apartado, tratará de ilustrarse cómo se preprocesa y almacena la información de cada receta a partir de los seis pasos que se describen a continuación:

1. **Tokenización:** se basa en dividir el texto en las unidades que lo conforman, entendiendo por unidad el elemento más sencillo con significado propio para el análisis en cuestión, en este caso, las palabras. La división se realiza teniendo en cuenta los caracteres de delimitación de palabra estándar, como los signos gráficos de puntuación (punto, coma, dos puntos, espacio o punto y coma).
2. **Expresión regular:** se utiliza para eliminar los caracteres que no son alfabéticos.
3. **Stopwords removal:** se eliminan las palabras que solo tienen valor gramatical, y se mantienen las palabras léxicas.
4. **Lematización:** las palabras flexionadas y derivadas se lematizan bajo su forma canónica (por ejemplo, en la segunda receta, *potatoes* se engloba bajo el lema *potato*). Se eliminan, también, los nombres propios y/o de marcas comerciales, pues no se asocian a ningún lema (salvo que estén lexicalizados).
5. **Jerarquía ontológica de los términos en la FunGramKB:** a partir de las búsquedas que permite realizar la herramienta *Navigator*, se seleccionan los sustantivos que contienen, en su esquema semántico, la entrada conceptual (concepto *básico*, de acuerdo con la mencionada estructuración de la Ontología de FunGramKB) +FOOD_00, lo que permite extraer aquellos que hacen referencia a alimentos (por ejemplo, el concepto *onion* presenta la siguiente jerarquía ontológica: #ENTITY > #PHYSICAL > #OBJECT > #SELF_CONNECTED_OBJECT > +NATURAL_OBJECT_00 > +CORPUSCULAR_00 > +SOLID_00 > +**FOOD_00** > +VEGETABLE_00 > +ONION_00).
6. **Base de conocimiento:** está formada por una matriz donde se indexan las recetas. Las columnas de la matriz (atributos) representan cada uno de los ingredientes; las filas (instancias),

las recetas y las celdas (intersección fila-columna) contienen una codificación binaria (1 o 0) donde se indica la presencia o ausencia del atributo en la receta, respectivamente.

Teniendo en cuenta que las fases descritas se repiten en el tratamiento de cada uno de los textos, en lo sucesivo, se tratará de ilustrar esquemáticamente cómo se realizaría el pre-procesamiento y almacenamiento de la información en la base de conocimiento del SRI, a partir de tres recetas que permiten explicar el procedimiento de forma esquemática.

TABLA 2. Primera receta de ejemplo de preprocesamiento

RECETA 1	
<pre>"nqBvmVkWIXKP2YacMkfdM8wfCMZDEYO": { "title": "Oven Roasted Red Potatoes", "ingredients": ["1 (1 ounce) envelope dry onion soup mix", "2 pounds red potatoes, halved", "1/3 cup olive oil",], "instructions": "Preheat oven to 450 degrees F (230 degrees C).\nIn a large plastic bag, combine the soup mix, red potatoes and olive oil. Close bag, and shake until potatoes are fully covered.\nPour potatoes into a medium baking dish; bake 40 minutes in the preheated oven, stirring occasionally.\n"} </pre>	

Fuente: elaboración propia

1. Tokenización:

1	(1	ounce)	envelope	dry	onion	soup	mix
2	pounds	red	potatoes	halved			
1/3	cup	olive	oil				

2. Expresión regular:

1	(1	ounce)	envelope	dry	onion	soup	mix
2	pounds	red	potatoes	halved			
1/3	cup	olive	oil				

3. Stopwords removal:

	ounce	envelope	dry	onion	soup	mix
pounds	red	potatoes	halved			
cup	olive	oil				

4. Lematización:

	ounce	envelope	dry	onion	soup	mix
pounds	red	potatoes	halved			
cup	olive	oil				

5. Jerarquía ontológica de los términos en la FunGramKB:

onion						
#ENTITY > #PHYSICAL > #OBJECT > #SELF_CONNECTED_OBJECT > +NATURAL_OBJECT_00 > +CORPUSCULAR_00 > +SOLID_00 > +FOOD_00 > +VEGETABLE_00 > +ONION_00						
soup						
#ENTITY > #PHYSICAL > #OBJECT > #SELF_CONNECTED_OBJECT > +ARTIFICIAL_OBJECT_00 > +SUBSTANCE_00 > +LIQUID_00 > +FOOD_00 > +SOUP_00						
potato						
#ENTITY > #PHYSICAL > #OBJECT > #SELF_CONNECTED_OBJECT > +NATURAL_OBJECT_00 > +CORPUSCULAR_00 > +SOLID_00 > +FOOD_00 > +VEGETABLE_00 > +POTATO_00						
oil						
#ENTITY > #PHYSICAL > #OBJECT > #SELF_CONNECTED_OBJECT > +ARTIFICIAL_OBJECT_00 > +SUBSTANCE_00 > +LIQUID_00 > +FOOD_00 > +OIL_00						
	ounce	envelope	dry	onion	soup	mix
pounds	red	potato	halved			
cup	olive	oil				

6. Base de conocimiento:

	Onion	Soup	Potato	Oil	Descripción
R1	1	1	1	1	D1

TABLA 3. Segunda receta de ejemplo de preprocesamiento

RECETA 2
<pre>"DvnbTZ2ePbVTHXmrc.n13YopTheo9WO": { "title": "Creamy Au Gratin Potatoes", "ingredients": ["4 russet potatoes, sliced into 1/4 inch slices", "1 onion, sliced into rings", "salt and pepper to taste", "3 tablespoons butter", "3 tablespoons all-purpose flour", "1/2 teaspoon salt", "2 cups milk", "1 1/2 cups shredded Cheddar cheese",], "instructions": "Preheat oven to 400 degrees F (200 degrees C). Butter a 1 quart casserole dish.\nLayer 1/2 of the potatoes into bottom of the prepared casserole dish. Top with the onion slices, and add the remaining potatoes. Season with salt and pepper to taste.\nIn a medium-size saucepan, melt butter over medium heat. Mix in the flour and salt, and stir constantly with a whisk for one minute. Stir in milk. Cook until mixture has thickened. Stir in cheese all at once, and continue stirring until melted, about 30 to 60 seconds. Pour cheese over the potatoes, and cover the dish with aluminum foil.\nBake 1 1/2 hours in the preheated oven.\n"} }</pre>

Fuente: elaboración propia

1. Tokenización:

4	russet	potatoes	sliced	into	1/4	inch	slices
1	onion	sliced	into	rings			
salt	and	pepper	to	taste			
3	tablespoon	butter					
3	tablespoon	all-purpose	flour				
1/2	teaspoons	salt					
2	cups	milk					
1	1/2	cups	shredded	Cheddar	cheese		

2. Expresión regular:

4	russet	potatoes	sliced	into	1/4	inch	slices
1	onion	sliced	into	rings			
salt	and	pepper	to	taste			
3	tablespoon	butter					
3	tablespoon	all-purpose	flour				
1/2	teaspoons	salt					
2	cups	milk					
1	1/2	cups	shredded	Cheddar	cheese		

3. Stopwords removal:

	russet	potatoes	sliced	into		slices
	onion	sliced	into	rings		
salt	and	pepper	to	taste		
	tablespoon	butter				
	tablespoon		flour			
	teaspoons	salt				
	cups	milk				
		cups	shredded	Cheddar	cheese	

4. Lematización:

	russet	potatoes	sliced			slices
	onion	sliced		rings		
salt		pepper		taste		
	tablespoon	butter				
	tablespoon		flour			
	teaspoons	salt				
	cups	milk				
		cups	shredded	Cheddar	cheese	

5. Jerarquía ontológica de los términos en la FunGramKB:

salt						
#ENTITY > #PHYSICAL > #OBJECT > #SELF_CONNECTED_OBJECT > +ARTIFICIAL_OBJECT_00 > +SUBSTANCE_00 > +SOLID_00 > +FOOD_00 > +SEASONING_00 > +SALT_00						
pepper						
#ENTITY > #PHYSICAL > #OBJECT > #SELF_CONNECTED_OBJECT > +ARTIFICIAL_OBJECT_00 > +SUBSTANCE_00 > +SOLID_00 > +FOOD_00 > +SEASONING_00 > +PEPPER_00						
butter						
#ENTITY > #PHYSICAL > #OBJECT > #SELF_CONNECTED_OBJECT > +ARTIFICIAL_OBJECT_00 > +SUBSTANCE_00 > +SOLID_00 > +FOOD_00 > +BUTTER_00						
cheese						
#ENTITY > #PHYSICAL > #OBJECT > #SELF_CONNECTED_OBJECT > +ARTIFICIAL_OBJECT_00 > +SUBSTANCE_00 > +SOLID_00 > +FOOD_00 > +CHEESE_00						
	russet	potato	slice			slice
	onion	sliced		ring		
salt		pepper		taste		
	tablespoon	butter				
	tablespoon		flour			

	teaspoon	salt				
	cup	milk				
		cup	shred		cheese	

6. Base de conocimiento:

	Onion	Soup	Potato	Oil	Salt	Pepper	Butter	Cheese	Descripción
R1	1	1	1	1	0	0	0	0	D1
R2	1	0	1	0	1	1	1	1	D2

TABLA 4. Tercera receta de ejemplo de preprocesamiento

RECETA 3
<pre>"H65OtOXu5WMe6Tuo0loyhP6xZrYCY3i": { "title": "Perfect Baked Potato", "ingredients": ["1 medium baking potato", "1 teaspoon olive oil", "1/2 teaspoon salt", "2 teaspoons butter", "1 pinch freshly ground black pepper", "1/4 cup shredded Cheddar cheese",], "instructions": "Preheat the oven to 300 degrees F (150 degrees C). Scrub the potato, and pierce the skin several times with a knife or fork. Rub the skin with olive oil, then with salt.\nPlace the potato in the preheated oven, and bake for 90 minutes, or until slightly soft and golden brown. Slice the potato down the center, and serve with butter and black pepper. Sprinkle shredded Cheddar cheese over the top, if desired.\n" }</pre>

Fuente: elaboración propia

1. Tokenización:

1	medium	baking	potato		
1	teaspoon	olive	oil		
1/2	teaspoon	salt			
1	pinch	freshly	ground	black	pepper
1/4	cup	shredded	Cheddar	cheese	

2. Expresión regular:

1	medium	baking	potato		
1	teaspoon	olive	oil		
1/2	teaspoon	salt			

1	pinch	freshly	ground	black	pepper
1/4	cup	shredded	Cheddar	cheese	

3. Stopwords removal:

	medium	baking	potato		
	teaspoon	olive	oil		
	teaspoon	salt			
	pinch	freshly	ground	black	pepper
	cup	shredded	Cheddar	cheese	

4. Lematización:

		baking	potato		
	teaspoon	olive	oil		
	teaspoon	salt			
	pinch	freshly	ground	black	pepper
	cup	shredded	Cheddar	cheese	

5. Jerarquía ontológica de los términos en la FunGramKB:

			potato		
			oil		
		salt			
					pepper
				cheese	

6. Base de conocimiento:

	Onion	Soup	Potato	Oil	Salt	Pepper	Butter	Cheese	Descripción
R1	1	1	1	1	0	0	0	0	D1
R2	1	0	1	0	1	1	1	1	D2
R3	0	0	1	1	1	1	0	1	D3

A continuación, trataremos de ilustrar el funcionamiento del sistema de consulta, a partir de tres ejemplos de posibles búsquedas que podría realizar un usuario y los resultados que obtendría de ellas, de acuerdo con el SRI que hemos diseñado.

7. Hipótesis de consulta 1

Suponiendo que un usuario desea buscar recetas de cocina que contengan cebolla y patata o cebolla y pimienta, la consulta booleana sería la siguiente:

(onion) AND (potato OR pepper)

Con ella, obtendremos, de la matriz de conocimiento, los documentos que contienen cada uno de los términos:

$$U^{44} = \{1,2,3\}$$

$$\text{Onion} = \{1,2\}$$

$$\text{Potato} = \{1,2,3\}$$

$$\text{Pepper} = \{2,3\}$$

Por lo que:

$$A = (\text{onion}) \text{ AND } (\text{potato}) = \{1,2\} \cap \{1,2,3\} = \{1,2\}$$

$$B = (\text{onion}) \text{ AND } (\text{pepper}) = \{1,2\} \cap \{2,3\} = \{2\}$$

$$C = (\text{onion AND potato}) \text{ OR } (\text{onion AND pepper}) = \{1,2\} \cup \{2\} = \{1,2\}$$

De este modo, la respuesta del sistema serían las recetas $\{1,2\}$.

8. Hipótesis de consulta 2

Ante un usuario que busque recetas que contengan queso o manjetilla junto con sal y caldo, la consulta booleana sería

(cheese OR butter) AND salt AND soup

Con ella, se obtendrán, de la matriz de conocimiento, los documentos que contienen cada uno de los términos:

$$U = \{1,2,3\}$$

$$\text{Cheese} = \{2\}$$

$$\text{Butter} = \{2\}$$

$$\text{Salt} = \{2,3\}$$

$$\text{Soup} = \{1\}$$

⁴⁴ Universo de conocimiento.

Por lo que:

$$\begin{aligned}A &= \text{cheese AND salt AND soup} = \{2\} \cap \{2,3\} \cap \{1\} = \{\emptyset\} \\B &= \text{butter AND salt AND soup} = \{2\} \cap \{2,3\} \cap \{2\} = \{2\} \\C &= (\text{cheese AND salt AND soup}) \text{ OR } (\text{butter AND salt AND soup}) = \\&\{\emptyset\} \cup \{2\} = \{2\}\end{aligned}$$

De este modo, la respuesta del sistema sería la receta $\{2\}$.

9. Hipótesis de consulta 3

Si la búsqueda del usuario consistiese en recuperar las recetas que contienen pimienta y queso y que no llevan cebolla, la consulta booleana sería la que sigue:

$$\text{pepper AND cheese AND NOT onion}$$

Con ella, se obtendrán, de la matriz de conocimiento, los documentos que contienen cada uno de los términos:

$$\begin{aligned}U &= \{1,2,3\} \\ \text{Pepper} &= \{2,3\} \\ \text{Cheese} &= \{2,3\} \\ \text{Onion} &= \{1,2\}\end{aligned}$$

Por lo que:

$$A = \text{pepper AND cheese AND NOT onion} = \{2,3\} \cap \{2,3\} \cap \{3\} = \{3\}$$

La respuesta del sistema, en esta ocasión, sería la receta $\{3\}$.

6. CONCLUSIONES

El sistema de recuperación de información diseñado permite discriminar los sustantivos no válidos (“cucharada”, “porción”, etc.) y seleccionar solo aquellos que contengan el concepto FOOD_00 en su esquema semántico. Esto se ha logrado empleando la ontología de FunGramKB como soporte conceptual, de modo que el SRI no identifica la palabra, sino el concepto. Así, por ejemplo, la palabra polisémica *kiwi*, que puede designar una fruta (esto es, un alimento) o un tipo de ave, sería diferenciada por nuestro buscador, de modo que, al consultar *kiwi* en el *dataset*,

el SRI solo devolvería los documentos que contuviesen este sustantivo con el valor semántico de alimento.

Aunque, para el diseño piloto presentado, el método booleano resulta adecuado, los resultados podrían optimizarse empleando un modelo probabilístico o vectorial, pues permitirían ordenar los resultados de acuerdo con su grado de similitud con respecto a la búsqueda realizada. Esta posibilidad podría tomarse en cuenta, como propuesta de mejora, para trabajos futuros sobre SRI aplicados a recetarios de cocina digitalizados u otros escenarios web.

En definitiva, lo que ha tratado de evidenciarse, a lo largo de este trabajo, es la utilidad de la Ontología de FunGramKB como herramienta de soporte para otros recursos informáticos, particularmente para los SRI aplicados a escenarios especializados. Aunque se ha empleado como escenario de aplicación un conjunto de recetas de cocina, podría plantearse cualquier otro cuya terminología estuviese incluida en FunGramKB (que se encuentra, como es esperable, en continuo proceso de construcción).

8. REFERENCIAS

- Boole, G. (1847). *The Mathematical Analysis of Logic*. Macmillan.
- Bush, V. (1945). *As we may think*. *Atlantic Monthly*, 176 (1), 101-108.
- Cacheda, F. (2008). Introduction to the Classic Models of Information Retrieval. *Revista General de Información y Documentación*, 18, 365-374.
- Cleverdon, C. W. (1991). The significance of the Cranfield tests on index languages. *SIGIR '91: Proceedings of the 14th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 3-12). ACM Digital Library. <https://doi.org/10.1145/122860.122861>
- Friedl, J. E. F. (2006). *Mastering Regular Expressions*. O'Reilly.
- Goddard, C. y Wierzbicka, A. (2002). *Meaning and universal grammar*. John Benjamins.
- Lee, Ryan. 2021. *Eight Portions*. Consultado el 20 de febrero de 2023. <https://eightportions.com/>
- Liddy, E. D. (2005). *Automatic Document Retrieval*. *Encyclopedia of Language and Linguistics*. Elsevier Press.
- Mooers, C. (1950). *Making Information Retrieval Pay*. Zator Company.

- Nadkarni, P. M., Ohno-Machado, L. y Chapman, W. W. (2011). Natural language processing: an introduction. *J Am Med Inform Association*, 18(5), 544-51.
- Periñán-Pascual, C. FunGramKB. Consultado el 20 de febrero de 2023. <http://www.fungramkb.com>
- Periñán-Pascual, C. y Mairal-Usón, R. (2010). Enhancing UniArab with FunGramKB. *Procesamiento del Lenguaje Natural*, 44, 19-26.
- Periñán-Pascual, C. y Arcas-Túnez, F. (2010). The architecture of FunGramKB. *Proceedings of the Seventh International Conference on Language Resources and Evaluation* (pp. 2667-2674). ELRA.
- Periñán-Pascual, C. y Arcas-Túnez, F. (2005). Microconceptual-Knowledge Spreading in FunGramKB. *Proceedings of the 9th IASTED International Conference on Artificial Intelligence and Soft Computing* (pp. 239-244). ACTA Press.
- Procter, P. (1978). *Longman Dictionary of Contemporary English*. Longman.
- Ruano-Ordás, D., Fdez-Riverola, F. y Méndez, R. (2018). Using evolutionary computation for discovering spam patterns from e-mail samples. *Information Processing & Management*, 54, 303-317. <https://doi.org/10.1016/J.IPM.2017.12.001>
- Savand, A. (2018). Alir3z4/stop-words. Consultado el 20 de febrero de 2023. <https://github.com/Alir3z4/stop-words/blob/master/english.txt>
- Tulving, E. (1985). How many memory systems are there? *American Psychologist*, 40, 385-398.
- Witten, I., Paynter, G. W., Frank, E., Gutwin, C. y Nevill-Manning, C. G. (1999). KEA: Practical Automatic Keyphrase Extraction. *DL '99: Proceedings of the fourth ACM conference on Digital libraries* (pp. 254-255). ACM Digital Libraries. <https://doi.org/10.1145/313238.313437>
- Zobel, J. y Moffat, A. (2006). Inverted files for text search engines. *ACM Computing Surveys*, 38, 1-56.