

Estructuras conceptuales como modelos de representación en la extracción de conocimiento almacenado en texto no estructurado*

Gustavo Pérez Reyes**

Resumen

De las herramientas para facilitar el acceso al cúmulo de información que se genera diariamente, una de las más utilizadas es la comúnmente llamada minería de texto. Aquí proponemos utilizar y analizar dos métodos: mapas y grafos conceptuales. Estos se caracterizan por ser una encarnación más completa del contenido almacenado con el objetivo de representar la información estructural de un texto. Ello implica un avance significativo a un nivel conceptual respecto del nivel temático propio de los modelos descriptivo y predictivos. Debido al volumen de información a procesar puede requerir una gran carga computacional, un objetivo secundario pero no menos importante es la aplicación del modelo de agentes para controlar de manera dinámica el espacio de búsqueda y optimizar los tiempos de procesamiento de los textos. El presente es un trabajo en desarrollo, centrado sobre el eje conceptual, mientras el aspecto práctico forma parte del trabajo subsecuente.

Palabras clave: Minería de texto. Mapas conceptuales. Grafos conceptuales. Teoría de agentes.

Abstract

Of the tools to facilitate access to the accumulation of information that is generated daily, one of the most used is the one commonly called text mining. The proposal here consists of the use and subsequent analyzing of two methods: maps and concept graphs. These are characterized by being a more complete incarnation of the stored content in order to represent the structural information of a text. This implies a significant advance at a conceptual level with respect to

* Enviado: 1-03-2021. Aceptado: 22-06-2021.

** Analista de Sistemas Informáticos por el Instituto Tecnológico ORT. Licenciado en Ciencias de la Computación en Universidad de Palermo. Investigador Independiente. E-mail: greyes1@palermo.edu

the thematic level of the descriptive and predictive models. Due to the volume of information to be processed, it can require a large computational load, a secondary but no less important objective is the application of the agent model to dynamically control the search space and optimize text processing times. This is a work in progress, in this first stage it will focus on the conceptual axis, while the practical aspect is part of the subsequent work.

Key words: Text mining. Concept maps. Concept graphs. Agent theory.

Resumo

Das ferramentas que facilitam o acesso ao acúmulo de informações geradas diariamente, uma das mais utilizadas é a comumente chamada text mining. A proposta aqui consiste na utilização e posterior análise de dois métodos: mapas e gráficos conceituais. Estes são caracterizados por serem uma encarnação mais completa do conteúdo armazenado para representar a informação estrutural de um texto. Isso implica um avanço significativo a nível conceitual no que diz respeito ao nível temático dos modelos descritivos e preditivos. Devido ao volume de informações a serem processadas, pode exigir uma grande carga computacional, um objetivo secundário, mas não menos importante, é a aplicação do modelo de agente para controlar dinamicamente o espaço de busca e otimizar os tempos de processamento do texto. Este é um work in progress, esta primeira fase centra-se no eixo conceptual, enquanto o aspecto prático faz parte do trabalho subsequente.

Palavras-chaves: Mineração de texto. Mapas conceituais. Gráficos conceituais. Teoria dos agentes.

Introducción

Hay autores que identifican el proceso de descubrimiento de conocimiento en texto, KDT por sus siglas en inglés, con la minería de textos, sin embargo otros se inclinan por definir cabalmente a la minería como fase principal del proceso global de descubrimiento. “Hay autores que ven la minería de textos como un equivalente de la minería de datos pero aplicado a información textual y los hay que opinan que la minería de textos debería realizar algo más. El

concepto de Minería de Textos es relativamente reciente. No es hasta finales de los 90 cuando se produce el despegue de las investigaciones en esta área. El trabajo de Hearst marca un hito en este campo y sienta las bases tanto para diferenciar los conceptos de Descubrimiento de Conocimiento en Datos y en Textos como para definir que es Minería de Textos” (Justicia de la Torre, 2017: 23).

De todas las definiciones seleccionadas (Delgado, 2002; Feldman & Dagan, 1995; Frawley et al., 1992; Hearst, 1999; Montes y Gomez et al., 2002; Nahm & Mooney, 2002; Tan, 1999; Xu et al., 2002) me quedo con esta que considero más precisa y ajustada para el presente trabajo: el descubrimiento de conocimiento en texto constituye el área de conocimiento dentro de la que se estudia y se generan soluciones para el descubrimiento de conocimientos potencialmente útiles, y no explícito, en una colección de textos, a partir de la identificación y exploración de patrones interesantes (Feldman et al., 1998) “combinando métodos de aprendizaje con métodos de procesamiento de textos. Está asociado principalmente al descubrimiento de patrones interesantes como clusters, asociaciones, desviaciones, similitudes, y diferencias” (Ale & Federico, 2007: 1).

Antecedentes y Planteamiento

Información y conocimiento

En la actualidad el concepto clásico de información, como agrupación de datos organizados y presentados en un contexto, se ha visto superado por una nueva visión de la misma como recurso que genera valor para el negocio sobre la base de la creación de productos más competitivos o de procesos más eficaces.

En términos más concretos, “el conocimiento surge cuando una persona considera, interpreta y utiliza la información de manera combinada con su propia experiencia y capacidad” (Mazo I, Ortiz de Frutos E., 1998: 32) En consecuencia, se puede afirmar que el conocimiento está determinado por la interpretación que las personas realizan con la información disponible, condicionada por el contexto en el que se desenvuelven y la experiencia que poseen.

“El conocimiento es un recurso estratégico para el desarrollo económico y social

contemporáneo; la información es el elemento básico principal en el proceso de adquisición, generación, gestión y transmisión del conocimiento” (Febles Rodríguez J, González Pérez A., 2002: 1) En este sentido, es la expresión material del conocimiento con fines de uso, está destinada a resolver determinados problemas. Debe estar, entonces, disponible y servir para el desarrollo individual y corporativo.

Los volúmenes de datos, informaciones y conocimientos que se almacenan en bases de datos y en los que se han dado en llamar grandes almacenes de datos o data warehouse, hoy — como resultado del desarrollo de las TIC— permiten una exploración dirigida al descubrimiento del conocimiento almacenado en ellos.

Descubrimiento de Conocimiento en Textos

Se han creado nuevas herramientas para facilitar el acceso al cúmulo de información que se genera diariamente. Una de las más utilizadas a nivel organizacional es la comúnmente llamada minería de texto que ofrece a la organización la posibilidad de explorar grandes cantidades de textos, no organizados en forma de datos, establecer patrones y extraer conocimientos útiles. (Botta-Ferret Eleazar, Cabrera-Gato Jania E., 2007: 1)

La gente a menudo habla de minería de texto y datos al mismo tiempo, pero estrictamente hablando la minería de texto es un proceso del descubrimiento de conocimiento en texto que a su vez es una forma específica de minería de datos que se relaciona con el texto. El Descubrimiento de Conocimiento en Textos o KDT (Knowledge Discovery in Text en inglés) intenta resolver una particularidad del texto escrito en lenguaje natural: la ausencia de estructura del mismo.

En la literatura está mucho más extendido el termino Minería de Textos que el termino Descubrimiento de Conocimiento en Textos. Esto es debido a que la mayoría de las veces se utiliza la fase Minería de Textos para identificar al proceso de descubrimiento completo. Este fenómeno nace de una metonimia semántica: la fase principal del proceso identifica al proceso genérico.

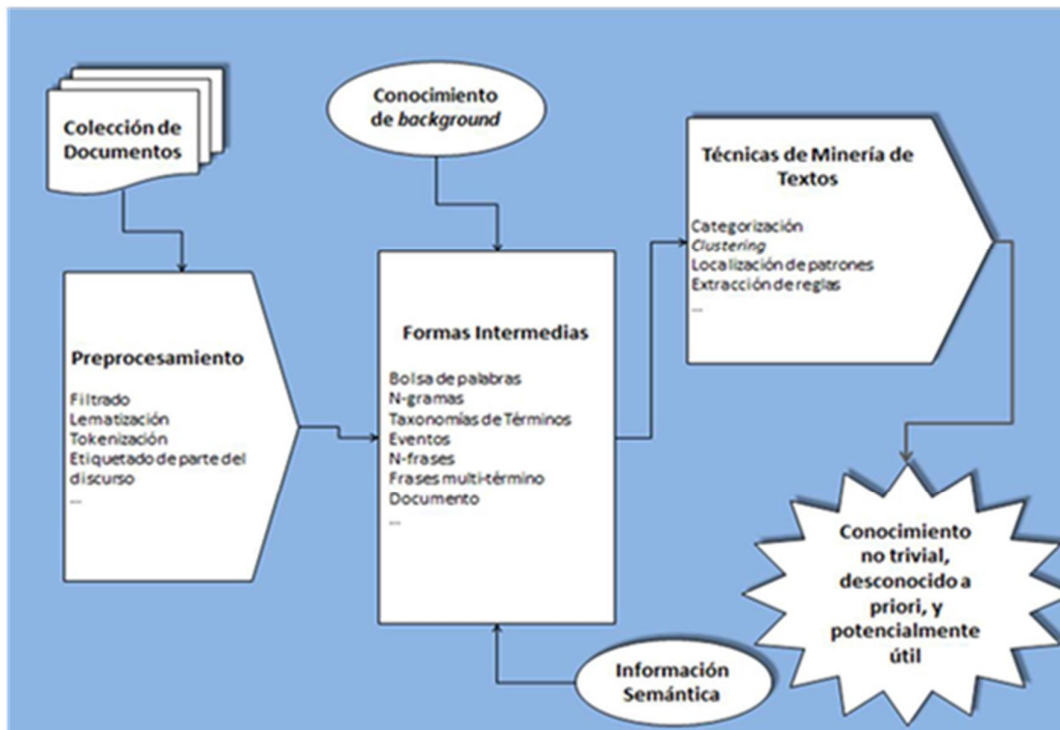


Figura 1. Etapas del proceso de Descubrimiento de Conocimiento en Textos.

Como se observa en la figura 1, el Descubrimiento de Conocimiento en Textos puede dividirse en tres pasos e implicara a diferentes ámbitos de conocimiento, principalmente: Recuperación de Información (para filtrar y reunir documentos adecuados), Extracción de Información (que selecciona hechos específicos sobre tipos de entidades y relaciones de interés), Procesamiento del Lenguaje Natural (para realizar el pre-procesamiento y etiquetado de los textos) y Minería de Datos (para descubrir asociaciones desconocidas entre hechos desconocidos). Es un método muy eficiente para generar nueva información y conocimiento a la vez que permite reducir el tiempo dedicado a la lectura de textos extensos. Esto significa que la información clave se puede encontrar con mayor rapidez y eficacia. También permite a los usuarios obtener nueva información que de otro modo sería difícil de encontrar. De tal modo que una gran ayuda en la toma de decisiones y permite responder rápidamente a las consultas de los clientes en diversos procesos. Algunas de las técnicas más comunes usadas son:

- Árboles de decisión
- Métodos de clasificación y regresiones no-lineales
- Métodos basados en ejemplos prototípicos
- Modelos gráficos de dependencias probabilísticas
- Modelos relacionales

La minería de textos lleva las cosas un paso más allá al extraer información precisa basada en mucho más que palabras clave. En su lugar, busca entidades o conceptos, relaciones, frases y/o oraciones; intenta determinar el significado real basado en algoritmos de Procesamiento del Lenguaje Natural (NLP), que le permiten reconocer conceptos similares. Una búsqueda utilizando la minería de texto puede identificar hechos, relaciones e inferencias que no son del todo obvios.

Mapas Conceptuales

El término *mapa conceptual* se emplea para abarcar una amplia gama de representaciones esquemáticas del conocimiento. Definido por Novak como una forma de instrumentar la teoría de aprendizaje significativo, la que se sustenta en que el nuevo conocimiento es adquirido a partir de lo que ya se conoce y esto se realiza a través de un proceso constructivista (Ausubel y Novak, 1989). En este escenario, Novak define un Mapa Conceptual como una técnica que representa, simultáneamente, una estrategia de aprendizaje, un método para captar lo más significativo de un tema y un recurso esquemático para representar un conjunto de significados conceptuales incluidos en una estructura de proposiciones (Novak y Gowin, 1988). Otra definición más general es la que los reconoce como una herramienta para organizar y representar el conocimiento (Novak y Cañas, 2008), en forma de grafo dirigido y etiquetado.

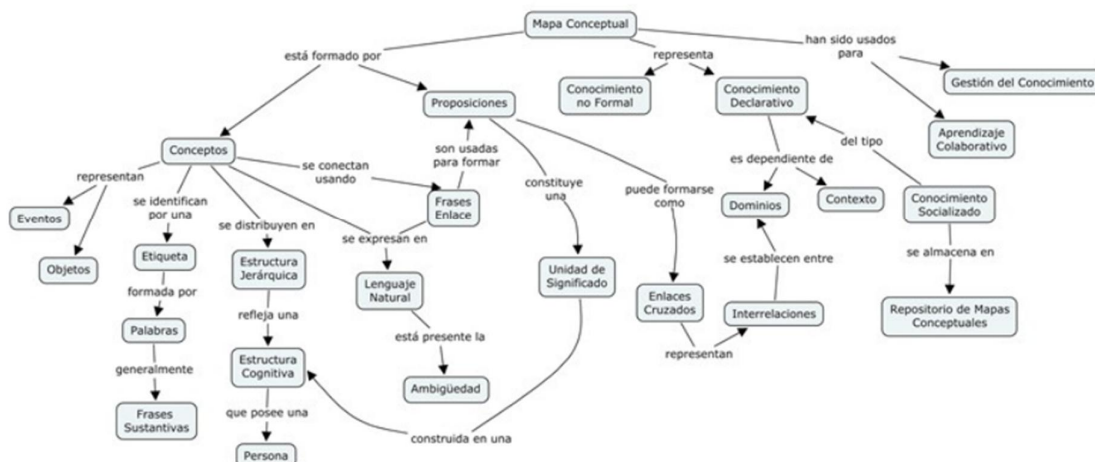


Figura 2. Ejemplo de MC representando los conceptos significativos que lo caracterizan como forma de representación de conocimiento. Fuente: <http://scielo.sld.cu/pdf/rcci/v7n1/rcci07113.pdf>

Un mapa conceptual constituye una forma de representación de conocimiento basada en conceptos significativos y sus relaciones en una estructura proposicional, tal como se ve en la figura 2. Se compone de conceptos y relaciones formando proposiciones. Los conceptos son definidos como regularidades percibidas en eventos u objetos, evidencias de ellos, especificados por una etiqueta, generalmente formada por una palabra (simple), pero es posible usar más de una (compuesta):

Los conceptos representan eventos u objetos, o evidencias de ellos, especificados por una etiqueta y las relaciones están etiquetadas por una frase que establece el tipo de relación entre los conceptos. Las proposiciones se forman por dos o más conceptos interconectados mediante una frase enlace, representando expresiones significativas, [...] y en ocasiones son consideradas como unidades semánticas o de significado. (Rodríguez Blanco y Simón Cuevas, 2013: 2)

Grafos Conceptuales

Los elementos básicos que invariablemente encontramos en todo esquema de representación de conocimiento formal, comúnmente denominados redes semánticas, son estructuras de datos en nodos, que representan conceptos, unidas por arcos que representan las relaciones entre los conceptos; y un conjunto de procedimientos de inferencia que operan sobre las estructuras de datos (Chantal Pérez Hernández, 2002).

La representación mediante red semántica permite representar relaciones entre conceptos. Intenta proporcionar una descripción para el significado de las palabras y las condiciones bajo las que los significados interactúan para ser compatibles con otros aspectos del lenguaje. Este tipo de representación es muy compleja pero contiene gran carga semántica. (Justicia de la Torre, 2017: 51)

Siendo utilizados por Dubois (Dubois and Quafafou, 2002) y Gelfand (Gelfand *et al.*, 1998).

Podemos distinguir tres tipos tradicionales de redes semánticas: redes IS-A, grafos conceptuales y redes de marcos. Los tres esquemas comparten ciertas características fundamentales: la representación de conocimiento basada principalmente en el uso de nodos y de relaciones al mismo tiempo que utilizan la inferencia como procedimiento para operar sobre la información representada en los nodos que pueden ser conceptos, estructuras o marcos.

La principal diferencia radica en las características representadas por los enlaces entre

los nodos, así, mientras que en las redes IS-A, los enlaces no están etiquetados, en las redes de marcos, los enlaces forman parte de la etiqueta que se le proporciona al nodo, mientras que los grafos conceptuales presentan diferencias importantes como el poseer dos tipos de nodos: de conceptos y de relaciones, ya que en los grafos conceptuales las relaciones también son un tipo de nodo (Chantal Pérez Hernández, 2002; Justicia de la Torre, 2017).

Como afirma John Sowa, los grafos conceptuales son un sistema de representación del conocimiento basado en las redes semánticas de la Inteligencia Artificial y en la lógica de Peirce, son un sistema de lógica orientado a la representación de la semántica del lenguaje natural (Sowa, 1984). Es un grafo finito, dirigido y bipartito, donde cada nodo o es un nodo concepto o un nodo relación (Mishne, 2003), donde cada relación se une a un número de conceptos y viceversa y que se utiliza para la representación de conocimiento.

Según Sowa y Way (1986) un grafo conceptual sirve para expresar significado de forma lógicamente precisa, legible por un humano y extensible computacionalmente. Los elementos básicos de un grafo conceptual son el tipo de concepto, el concepto en sí, y las relaciones conceptuales.

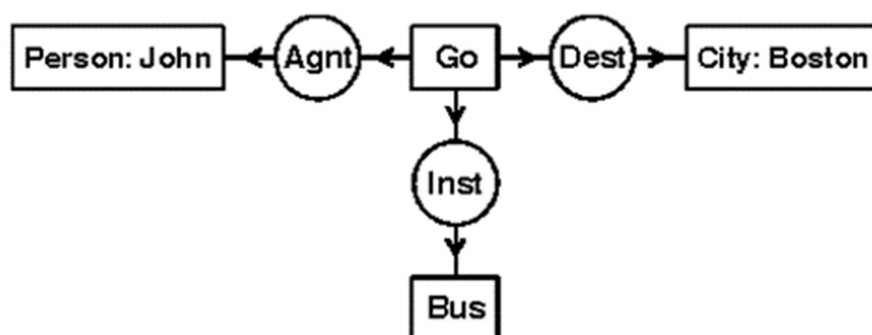


Figura 3. Ejemplo de grafo conceptual. Fuente: SOWA, John F. Conceptual Graph Examples. <http://www.jfsowa.com/cg/cgexamp.htm>

Los conceptos representan entidades, acciones y atributos, y tienen un tipo conceptual y un referente. El tipo conceptual indica la clase de elemento representado por el concepto, y el referente indica el elemento específico (instancia de la clase) referido por éste. Por ejemplo, para el concepto [gato:Félix], su tipo es gato y su referente es Félix (Montes, 2014).

En la Figura 3 se muestra un grafo conceptual con 4 conceptos [Go], [Person: John], [City: Boston], y [Bus]. Éste tiene 3 relaciones conceptuales: (Agnt) se refiere [Go] al agente

John, (Dest) se refiere [Go] al destino Boston, e (Inst) se refiere [Go] al instrumento bus. Desde el concepto [Go] se agregan 3 relaciones conceptuales.

Justificación del tema

La tendencia al crecimiento acelerado del volumen de información disponible no se puede obviar, lo que plantea nuevos problemas y retos para la recuperación de la información. Las diversas áreas de la organización que interactúan con esta información almacenada en texto no estructurado, y apenas indexado, no poseen el tiempo para la consulta y análisis de los grandes volúmenes de información existentes.

El conocimiento almacenado en texto “contiene una base de conocimiento completa que incorpora, además de hechos, relaciones de diversos tipos, con la desventaja de estar expresados en un lenguaje de representación (el lenguaje natural) que, aunque de una expresividad máxima, resulta totalmente inadecuado para llevar a cabo procesos automáticos de inferencia” (Justicia de la Torre, 2017: vi).

El procesamiento de esta información suele ser complejo debido a la gran cantidad de documentos (Tan, 1999), su heterogeneidad y su falta de estructura (Feldman & Dagan, 1995), la mayoría de la información de la que disponen las organizaciones está en formato textual en la intranet corporativa, páginas web, informes de trabajo, publicaciones, correos electrónicos, etc. (Justicia de la Torre, 2017).

Es precisamente atento que tratamos con datos textuales y no de datos estructurados en bases de datos lo que hace emerger una serie de problemas que necesitan un tratamiento adicional, requiriendo un proceso que los trate convenientemente. Se estima que alrededor del 80% de esa información está incluida en textos en lenguaje natural y en forma no estructurada (Gupta & Leha, 2009; Kumar et al., 2014; Ramanathan & Meyyappan, 2013; Sumathy & Chidambaram, 2013).

La posesión real de todo ese conocimiento dependerá de la habilidad para hacer ciertas operaciones con la información, por ejemplo:

- Buscar la información necesaria

- Comparar fuentes de información diferentes, y obtener conclusiones
- Manejar los textos, por ejemplo, traducirlos, editarlos, etc.

Ayudar al usuario a tener éxito en la búsqueda y utilización de la información es un proceso extremadamente complejo cuando desde un comienzo no se ha planificado recolectar toda la información necesaria en un único sitio de manera eficaz y con los medios efectivos para su recuperación.

Las tecnologías de la información y el conocimiento (TIC), que comprenden diferentes ciencias y áreas del conocimiento, incluyen herramientas para la gestión del conocimiento capaces de buscar, recuperar, filtrar, personalizar e, incluso, analizar datos estructurados o no.

Una de las disciplinas importantes de la ciencia de datos a este respecto es el Descubrimiento de Conocimiento en Textos o KDT (Knowledge Discovery in Text), siendo la Minería de Texto aquel proceso que descubre información nueva a partir de los datos y que constituye la fase central del proceso de descubrimiento.

KDT es un área multidisciplinaria que en base a un conjunto de técnicas y tecnologías busca extraer información útil e importante de formatos de documentos heterogéneos, tales como páginas web, correos electrónicos, medios sociales, artículos de revistas, etc. Esto se hace mediante la identificación de patrones dentro de los textos, tales como tendencias en el uso de palabras, estructura sintáctica, etc.

¿Qué podría aparejar el uso de estructuras conceptuales frente a otras herramientas? La mayoría de los procesos de minería de texto emplean representaciones sencillas del contenido de los textos, por ejemplo, listas de conceptos o palabras clave facilitando el análisis de los textos, pero a la vez limitando la variedad y utilidad de los patrones descubiertos, limitándose a un nivel temático.

¿Qué tan importante serían esos beneficios? Los métodos propuestos permiten procesar documentos de diferentes formatos, y combinar el análisis sintáctico superficial y profundo o de dependencias, el reconocimiento de entidades, patrones lingüísticos y conocimientos de referencia.

La propuesta aquí consiste en analizar dos métodos para la extracción de información estructurada que se caracterizan por ser una representación más completa del contenido almacenado con el objetivo de representar la información estructural de un texto. Ello implica

un avance significativo a un nivel conceptual respecto del nivel temático propio de los modelos descriptivo y predictivos de la Minería de Datos.

La búsqueda de nuevas alternativas para la representación de textos, que permitan no solo la inclusión de la semántica propia del lenguaje, y que esta representación sea exacta, sino que faciliten operaciones que conduzcan al descubrimiento de conocimiento, representa aún tema abierto a la investigación (Rodríguez Blanco *et al.*, 2015). En específico, en este trabajo se plantea el uso de mapas y grafos conceptuales para la representación del contenido de los textos.

¿Quiénes podrían valorar más esos beneficios? Existen diversas técnicas de Minería de Textos que se pueden aplicar a problemas del mundo real. Weiss (2010) proporciona una detallada lista de aplicaciones.

Justicia De la Torre (2017) aporta una recopilación de aplicaciones adicionales que se reproduce a continuación, la que no pretende ser exhaustiva, sino proporcionar un catálogo de casos que demuestren la utilidad real de las herramientas desarrolladas en este área:

Área de Aplicación	Uso
Web Semántica	Servicios Web de Descubrimiento Aprendizaje de ontologías
Redes Sociales	Filtrado de E-mail Personalización de perfiles web Detección de Comunidades Web Teorías Sociales
<i>Opinion Mining, Sentimental Analysis</i>	Clasificación de Opiniones <i>Hotspots</i> en Foros Predicciones Detección de Comportamiento Antisocial Encuestas de Opinión
Síntesis, Organización	Revisiones sistemáticas Obtención de Titulares Extracción de Ideas Útiles y Nuevas Discusión de Temas Principales Organización de Documentos Detección de Plagios Búsqueda de Contradicciones
Minería de Fuentes Abiertas, Tesoros	Identificación de Temas Exploración de Tesoros
<i>e-Commerce</i>	Toma de Decisiones
Marketing	Localización de "trozos" de información Análisis de Mercado
<i>e-Learning</i>	Herramientas Colaborativas
<i>Help Desk</i>	Generación de Casos Modelo Detección de Fallos

Tabla 1: Aplicaciones de la Minería de Textos.

Objetivos

Un aspecto importante en el desarrollo de soluciones de descubrimiento de conocimiento en texto es la representación intermedia que se utiliza para la estructuración y almacenamiento de los contenidos extraídos, puesto sobre esta se aplican las técnicas de análisis con el fin de alcanzar el descubrimiento de conocimiento. Existen diferentes enfoques en minería de textos distinguiéndose principalmente por las técnicas que se utilicen para obtener la representación intermedia y por el tipo de conocimiento que se pretenda extraer a partir del texto.

Se propone analizar dos modelos de representación, por un lado la representación basada en grafos y por otro en mapas conceptuales. Dichas estructuras constituyen la representación intermedia para la estructuración de contenidos textuales.

El objetivo general es definir que podría aparejar el uso de una u otra herramienta en la localización, manipulación y recuperación de la información almacenada en texto, establecer las ventajas y desventajas de las mismas para materializar y estructurar el contenido conceptual de la información almacenada, tal que se generen datos, se establezcan patrones y se extraigan conocimientos útiles.

Debido a que el volumen de información a procesar puede requerir una gran carga computacional, un objetivo secundario pero no menos importante es la aplicación del modelo de agentes (Ale & Federico, 2007; Federico, 2008; Henríquez Miranda, 2009; Azabache Paredes, 2013) en los algoritmos desarrollados para controlar de manera dinámica el espacio de búsqueda y en consecuencia optimizar los tiempos de procesamiento de los textos. La aplicación de la Teoría de Agentes permite definir una estrategia de refinado definida que posibilita reducir la complejidad de la representación resultante.

Desafío y planteamiento del problema

En el presente trabajo se abordada la problemática referente a la representación de contenidos textuales en el proceso de descubrimiento de conocimiento en textos. Uno de los

requisitos principales del proceso de Descubrimiento de Conocimiento en Textos es proporcionar una representación intermedia al texto:

Este paso, crucial y complicado, debe lidiar con datos procedentes de fuentes multilingües, debe identificar relaciones semánticas entre conceptos nuevos y antiguos o relacionar identificadores, todo ello procurando que el texto original no pierda su integridad. Esta representación intermedia del texto es conocida también como Forma Intermedia. (Justicia de la Torre *et al.*, 2008: 40).

Existe una enorme variedad de estas representaciones intermedias, en un abanico que varía tanto en complejidad como en eficiencia. Y de acuerdo a la carga computacional que requiere trabajar con la semántica y significado de un texto dado no siempre resultará rentable uno u otro método. En pos de dicho objetivo han sido empleados entre otras técnicas: Modelo Booleano (Valero Moreno, 2017: 31), Análisis de Semántica Latente (Valero Moreno, 2017: 32), el Modelo de Espacio Vectorial o MEV (Valero Moreno, 2017; 40), Taxonomía de Términos (Feldman, Aumann *et al.*, 2002) y diversas implementaciones de la teoría de grafos.

Aun cuando MEV es el más comúnmente usado (Shekhar *et al.*, 2014; Chang y Kim, 2014), los modelos basados en redes semánticas están teniendo una gran demanda. Algunos de estos modelos son: Grafos Conceptuales (Montes, Gelbukh *et al.*, 2001; Thavamani & Rengarajan, 2014), Redes Semánticas (Shekhar *et al.*, 2014), y Mapas Conceptuales (Palmeira *et al.*, 2012). De estas estructuras dependen los algoritmos y métodos de descubrimiento de conocimiento que se apliquen y sobre dos de ellas me ocupo aquí.

En este sentido, el desafío es realizar una comparativa entre un modelo de representación de texto basado en grafos (Rodríguez Blanco *et al.*, 2015; Montes y Gomez *et al.*, 2002; Montes, Gelbukh *et al.*, 2001; Montes, 2014) y otro inspirado en los fundamentos de los mapas conceptuales (Novak y Cañas, 2006; Novak y Cañas, 2008; Rodríguez Blanco y Simón Cuevas, 2013), optimizados a fin de que sean manejables computacionalmente.

Antes de proceder a describir su construcción, es oportuno recordar similitudes y diferencias de estas estructuras. Ambas tienen un aspecto gráfico y un aspecto semántico. Con respecto al aspecto gráfico, puede decirse que los mapas y las redes de grafos se parecen mientras que es en el aspecto semántico donde se diferencian sustancialmente.

Mientras que los mapas conceptuales poseen una jerarquía (Novak y Gowin, 1988; Ontoria, 1995); es decir, los conceptos más inclusivos se explicitan en la parte superior del

mismo y, descendiendo por el mapa, se encuentran los conceptos de jerarquía intermedia y luego aquellos que menos abarcan. La lectura de un mapa conceptual es, entonces, de arriba hacia abajo (Galagovsky y Ciliberti, 1994).

Los grafos conceptuales no sólo no requieren de una disposición jerarquizada de los conceptos, no hay otro criterio más allá del que sea establecido con miras a procurar claridad en la interpretación; por lo tanto, las conexiones entre nodos, en vez de líneas, son flechas que orientan el sentido de la lectura (Galagovsky, 1993, 1996; Galagovsky y Ciliberti, 1994). Mientras que es en el aspecto semántico donde los mapas y las redes de grafos se diferencian fundamentalmente. Es decir, importa con qué tipo de palabras está permitido llenar los nodos y completar las leyendas sobre los nexos.

A partir de los trabajos de investigación que se derivan de la descripción de las *redes conceptuales*, encontramos que las exigencias semánticas requeridas para la confección de estos instrumentos obligan a su constructor a realizar un análisis verdaderamente conceptual del contenido. Este nivel es claramente diferente del nivel semántico ambiguo –y muchas veces superficial– requerido para la construcción de los mapas (Ciliberti y Galagovsky, 1999: 1).

En la tabla 2 se muestran las diferencias fundamentales entre un mapa y una red de grafos conceptual.

	Mapa Conceptual	Grafos Conceptuales
Nodos	Se completan con sustantivos verbos o adjetivos. Se admiten expresiones matemáticas.	Se completan con sustantivos o sustantivos más adjetivos, que sean conceptos relevantes del tema. La repetición de nodos está prohibida. No se aceptan fórmulas matemáticas, excepto unidas por la leyenda «se simboliza mediante».
Leyendas que unen nodos	Se utiliza cualquier clase de palabras para formar proposiciones entre nodos. La extensión de una oración puede abarcar más de dos nodos consecutivos.	Se utilizan palabras y verbos muy precisos que completan una oración nuclear de óptimo significado entre dos nodos consecutivos (Galagovsky, 1993, 1996). La oración nuclear no puede extenderse consecutivamente a más de dos nodos. Las oraciones nucleares se leerán siguiendo el recorrido de una flecha.
Jerarquía Gráfica	Es absolutamente necesaria. Esta jerarquía vertical debe reflejar la jerarquía conceptual específica del tema.	No es necesaria. Los conceptos más importantes son los más relacionados.

Tabla 2. Diferencias fundamentales entre los *aspectos semántico y gráfico* necesarios para la construcción de un mapa o de una red de grafos conceptual.

En cuanto a la aplicación de la Teoría de Agentes, también conocidos como Agentes Inteligentes o Racionales, al ámbito del lenguaje y el descubrimiento de conocimiento en texto, se han llevado a cabo investigaciones que propician su aplicación (Ciravegna *et al.*, 2001; Federico, 2008; Feldman, 1999; Henríquez Miranda, 2009; Tsuyoshi Kitani *et al.*, 1994). Ahora bien, ¿cuáles son los beneficios que otorga la incorporación de agentes racionales que justifica su implementación? Con esta incorporación se busca tener mayor dominio sobre el volumen del grafo y mapa resultante mediante la reducción de los caminos irrelevantes y una mayor velocidad en la detección de patrones mediante la definición del espacio de búsqueda con la reducción consecuente en los tiempos de proceso y recursos computacionales, simplificando el manejo de información.

Siendo de importancia para el entendimiento del presente trabajo, previo a pasar al desarrollo de la construcción de los grafos y mapas conceptuales, es necesario introducir el concepto y los elementos asociados a la teoría de agentes.

Teoría de Agentes

Dentro de la Inteligencia Artificial, un área que ha ganado relevancia es el modelo de agentes, impulsado por la necesidad de hallar abordajes alternativos para problemas complejos que son más difíciles de tratar mediante sistemas monolíticos o que son de naturaleza distribuida; y una creciente corriente de la IA que considera que ciertos comportamientos inteligentes más que ser modelados como conductas complejas individuales, es preferible un enfoque distribuido basado en la interacción de conductas simples. De hecho, racional resulta más adecuado como adjetivo que inteligente, lo cual sugeriría cierta capacidad de comprensión, aprendizaje y resolución de problemas, para describir su comportamiento.

Precisemos: ¿de qué hablamos cuando hablamos de agentes? Russell y Norvig, en su reconocido libro sobre Inteligencia Artificial (2020), señalan que un agente es un sistema capaz de percibir a través de sensores la información que proviene del entorno donde está inserto, procesar tales entradas y responder o actuar a través de efectores también llamados actuadores (ver Figura 4), por lo que se lo puede definir como una entidad que exhibe un comportamiento autónomo, situado en un ambiente a partir del cual percibe los cambios y en el cual es capaz de

realizar acciones para alcanzar sus objetivos, en tanto Jennings (1999) y Wooldridge (1999) lo expresan, más sencillamente, como un sistema informático situado en algún entorno, que es capaz de una acción autónoma y flexible en ese entorno para alcanzar sus objetivos de diseño.

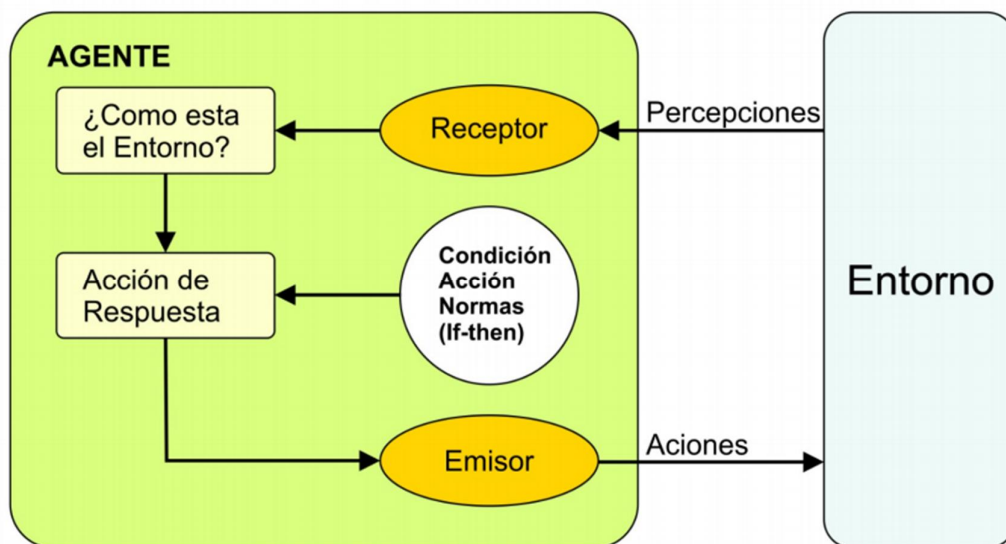


Figura 4. Agente Racional Simple.

De acuerdo al IBM Technology and Research, un agente racional es un “software que ayuda a las personas y actúa en su nombre, funcionan permitiendo que las personas deleguen el trabajo que podrían haber realizado en el software del agente, estos pueden, al igual que los asistentes, automatizar tareas repetitivas, recordar cosas que olvidó, resumir inteligentemente datos complejos, aprender de usted e incluso hacerle recomendaciones (Don Gilbert, 1997), o dicho de otro modo son entidades de software que ejecutan un conjunto de operaciones en nombre de un usuario u otro programa con cierto grado de independencia o autonomía, utilizando algún conocimiento o representación de las metas y deseos del usuario (Franklin y Graesser, 1996).

Como hemos visto, algunos prefieren considerar la conducta dirigida a objetivos, más aún unas elecciones que tiendan a maximizar un resultado, como aquello que es esencial y hacen foco en la racionalidad mientras otras definiciones hacen énfasis en la autonomía. ¿En qué consiste un agente? La definición dependerá del contexto donde el mismo sea aplicado.

Otros enfoques son más restrictivos, se basan en agentes cuyo conocimiento y capacidad deductiva son limitados, afirmando que es igualmente eficaz y más eficiente un conjunto de agentes simples para la ejecución de un sistema inteligente de forma más sencilla

(Baray y Wagner, 1999). Estas aproximaciones no son antagonistas, de hecho pueden ser fusionadas con el fin de pensar en ellos como un conjunto de entidades software autónomas que interaccionan mutuamente y con su entorno, que ejecutan una variedad de operaciones sencillas a partir de una representación de objetivos deseados (Federico, 2008). El hecho de ser autónomos significa que los agentes son entidades activas que pueden tomar sus propias decisiones, decidirá si realiza o no una operación solicitada, para lo cual tendrá en cuenta sus objetivos y prioridades, así como el contexto en que crea encontrarse (Pavón Mestras *et al.*, 2006).

A fin de poder concluir en una definición significativa vale preguntar: ¿qué constituye la esencia de ser un agente? Todas las definiciones tienen en cuenta unos conceptos fundamentales sobre los agentes racionales, a saber:

- Son sistemas autónomos. Operan independiente de los usuarios y no requiere del accionar de otro sistema. El concepto de autonomía implica que el agente posea cierto grado de inteligencia para la toma de decisiones.
- Actúan racionalmente tendiendo a maximizar un resultado esperado. Esta capacidad se puede llevar a cabo mediante los llamados motores de inferencia que poseen reglas para deducir ciertas acciones considerando la información disponible, otras posibilidades son el razonamiento basado en casos, o una red neuronal. Dependerá de las necesidades de la aplicación.
- Forman parte de un ambiente del cual obtienen información. El ambiente estimula al agente y le provee información para poder operar y tomar decisiones.
- Son colaborativos. Tienen la habilidad de trabajar junto con otros agentes para alcanzar un objetivo en común mucho más complejo. Se manejan diversos protocolos de comunicación para que un agente entienda a otro y pueda compartir los datos que posee.
- Tienen una meta o propósito que alcanzar.
- Actúan (o pueden actuar) en lugar de otro, pero no actúan de forma reactiva, aunque sí, a veces, proactiva; capaces de mostrar un comportamiento dirigido por objetivos.
- Sus acciones repercuten en el medio.

La definición de un agente es bastante amplia y se la puede confundir con la definición de un programa, sin embargo como se ha podido observar, existen fundamentales diferencias. Los agentes racionales poseen diversas propiedades según el propósito que deban alcanzar (Franklin y Graesser, 1996), lo cual deriva en la siguiente clasificación:

Propiedad	Significado
Reactivo	Responde inmediatamente a cambios en el ambiente
Autónomo	Tiene control sobre sus acciones.
Orientado a metas	No actúa en respuesta al ambiente, sino que sigue un plan para alcanzar su propósito.
Temporalmente continuo	Es un proceso que está continuamente corriendo
Comunicativo, sociable	Tiene la capacidad de comunicarse con otros agentes, tal vez también con usuarios.
Adaptativo	Cambia su comportamiento de acuerdo a sus experiencias.
Móvil	Se puede transportar autónomamente hacia otro lado
Flexible	Sus acciones no están determinadas de ante mano
Personaje	Con personalidad y "estado emocional".

Tabla 3. Categorización de agentes.

La categorización brindada anteriormente no es única, se puede clasificar a un agente de acuerdo a las metas que posea, a su arquitectura de control o al rango y sensibilidad de su percepción del medio (Federico, 2008). De acuerdo a la definición aquí utilizada un agente debe satisfacer las cuatro primeras características.

Construcción de Grafos conceptuales

El modelo escogido es el desarrollado por Rodríguez Blanco *et al.* (2015), el cual está inspirado en los fundamentos de los mapas conceptuales (Novak y Cañas, 2006) y el grafo de asociación (Medina *et al.*, 2005). Se basa en la representación de frases conceptuales formadas por una o varias palabras, y las relaciones entre ellos, según su proximidad en el texto, como rasgos representativos del contenido textual. La representación de conceptos, en lugar de términos simples, ofrece la posibilidad de capturar mayor información semántica, además de reducir la ambigüedad y llevar a cabo procesos discriminatorios más efectivos teniendo en cuenta la relevancia (Zhong y Wu, 2012).

El análisis de la proximidad entre conceptos en un texto, para la identificación y representación de las relaciones entre ellos, así como el uso de medidas de semejanza, posibilita descubrir patrones interesantes sobre la base del análisis de los vínculos contextuales en el

contenido.

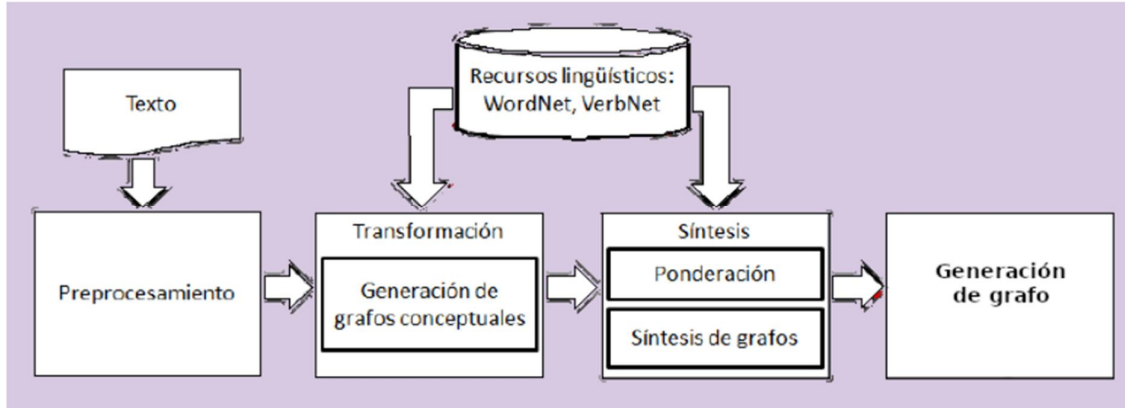


Figura 5. Esquema de construcción del modelo de representación con grafos. Basado en trabajo de Miranda *et al.* (2014)

Este análisis es soportado por métricas definidas para el cálculo de la distancia entre conceptos en un texto mediante técnicas tales como los algoritmos de agrupamiento (Medina *et al.*, 2005). El método propuesto se concibió en fases, en consonancia con el trabajo de Rodríguez Blanco y Simón Cuevas (2013), como se puede observar en la Figura 5.

Pre-procesamiento

En esta primera etapa se aplican técnicas de procesamiento de lenguaje natural. Dicho proceso es soportado por el analizador y etiquetador morfo sintáctico *Freeling* (diseñado por la Universitat Politecnica de Catalunya). El objetivo es capturar la información sintáctica de las oraciones del texto y los elementos que la componen. Es decir, etiquetar todas las palabras del texto con información que expresan la categoría gramatical y una serie de informaciones adicionales (ver Figura 6).

Yo	bajo	con	el	hombre	bajo	a
yo	bajar	con	el	hombre	bajo	a
PP1CSN00	VMIP1S0	SPS00	DA0MS0	NCMS000	AQ0MS0	SPS00
tocar	el	bajo	bajo	la	escalera	.
tocar	el	bajo	bajo	el	escalera	.
VMN0000	DA0MS0	NCMS000	SPS00	DA0FS0	NCFS000	Fp

Figura 6. Esquema morfo sintáctico de la oración “Yo bajo con el hombre bajo a tocar el bajo la escalera” llevada a cabo con *Freeling* (Oliver González, 2016).

Freeling posee las siguientes ventajas: 1. ser Software Libre; 2. puede hacer análisis a varios niveles: morfológico, morfo sintáctico, sintáctico superficial, sintáctico completo, de dependencias y semántico; 3. funcionar para muchas lenguas (asturiano, catalán, español, francés, gales, gallego, inglés, portugués, ruso) aunque lamentablemente no todas disponen de todos los niveles de análisis descriptos. Posee versiones en línea así como ejecutables para plataformas Linux y Windows.

Transformación y extracción

En esta etapa se combinan patrones léxicos definidos, con recursos lingüísticos de dominio público para la desambiguación y la identificación de conceptos, y se definen dos criterios referentes a la evaluación de la proximidad (distancia más cercana y ventana contextual), así como reglas las asociadas a cada uno de ellos para identificar los tipos de relaciones entre los conceptos. Durante el proceso de generación de los grafos conceptuales se agrega información semántica y sintáctica de fuentes externas. En esta etapa se realiza la integración de varios recursos lingüísticos de dominio público, como bases de datos léxica y diccionarios con información sintáctica y semántica.

Síntesis y generación

Se realiza la ponderación y refinado de los grafos. En la etapa de síntesis, los grafos se reducen de acuerdo a un conjunto de operaciones. Son eliminados los conceptos que tengan menor relevancia, en cuanto a su vínculo contextual con el resto de los conceptos, a partir de la evaluación del peso de asociatividad acumulada; la métrica definida con este propósito.

Construcción del Mapa Conceptual

Varios autores han tratado el tema de la construcción automática de mapas conceptuales a partir de textos (Valerio y Leake, 2006; Valerio *et al.*, 2008; Kowata *et al.*, 2010; Estrada, 2011), y se han hecho propuestas, caracterizadas en su mayoría por ser soluciones

dirigidas a textos en idioma inglés.

Valerio y Leake (2006) proponen un algoritmo de extracción de información a partir de documentos. Inicialmente el documento es segmentado y cada sentencia o segmento es analizada sintácticamente usando el algoritmo de Charniak y Johnson (2005). En la extracción de conceptos, se determina que una palabra forma una frase conceptual, si es sustantivo o adjetivo, abordándose primero las frases más simples, que son las más cercanas a las hojas en un árbol de dependencias, luego las más complejas.

Kowata presenta un método de construcción de mapas conceptuales a partir de texto en el que se incluyen las siguientes etapas:

1. Extracción de Texto Plano
2. Segmentación del Texto
3. Extracción de tokens
4. PoS (Part-of-speech) Tagging
5. Reconocimiento de elementos centrales candidatos (conceptos y enlaces candidatos)
6. Intérprete de dependencias
7. Constructor del Mapa Conceptual

El uso de métodos lingüísticos para la construcción de Mapas Conceptuales a partir de texto también es considerada por Valerio y Leake (2006), donde también se usan frases sustantivas y verbales para extraer conceptos y relaciones.

La variante escogida (Rodríguez Blanco y Simón Cuevas, 2013), que será detallada a continuación, toma en consideración algunos aspectos tenidos en cuenta en otras propuestas similares (Valerio y Leake, 2006; Valerio *et al.*, 2008; Kowata *et al.*, 2010), aunque con la característica de estar dirigida al procesamiento de textos en idioma español (Simón *et al.*, 2004; Estrada, 2011). En el proceso de extracción y estructuración de información se parte de una información textual en lenguaje natural no estructurado y se ejecutan un conjunto de tareas que pueden ser agrupadas en tres etapas, a saber: 1) pre-procesamiento; 2) extracción de información (frases conceptuales y relaciones); 3) refinado y construcción.

Pre-procesamiento

Se compone de las siguientes etapas: extracción de texto plano, segmentación de texto, extracción de tokens, análisis morfo-sintáctico, análisis sintáctico superficial y análisis

sintáctico Profundo (o análisis de dependencias).

Extracción de información

Representa el núcleo del método propuesto, dado que aquí es donde se extraen los contenidos de información claves para la estructuración y construcción del Mapa Conceptual, tal es el caso de las frases que expresan conceptos potenciales, como las relaciones que se establecen entre ellos y sus correspondientes etiquetas.

En este proceso se tiene en cuenta toda la información sintáctica obtenida en la fase anterior y se dispone de un mecanismo de trabajo con un corpus de mapas conceptuales, el cual es conformado según el tipo de conocimiento que se pretenda extraer a partir del texto. Aquí se realiza la extracción de relaciones explícitas e implícitas.

Refinado y construcción

Luego de haber extraído los conceptos y relaciones explícita y no explícita entre ellos y formada las proposiciones, se procede a eliminar posibles errores que puedan existir en dichas proposiciones, para lo cual se ejecutan una serie de reglas, las cuales indicaran que proposiciones deben ser eliminadas. Por ejemplo, son eliminadas las proposiciones en las que el concepto origen es igual al concepto destino, o en las que alguno de los dos conceptos contenga a la frase de enlace. Además, se eliminan las proposiciones repetidas, que son las que cada concepto en una proposición sea igual o contenga a su correspondiente en otra proposición y que las frases de enlace sean iguales o una incluya a la otra.

Luego de concluido el refinado de las proposiciones, se procede con la normalización de los conceptos, que no es más que un proceso de unificación de conceptos, a partir de un algoritmo de comparación sintáctica, unificándose aquellos conceptos muy similares sintácticamente.

Por último, luego que se ha refinado la información extraída y estructurada en proposiciones se procede a la última tarea y es la integración de todas las proposiciones para construir finalmente el Mapa Conceptual y visualizarlo. Como resultado se estaría obteniendo de forma automática un Mapa Conceptual que representa el conjunto de frases conceptuales

(nodos) que fueron identificadas y las relaciones entre ellas.

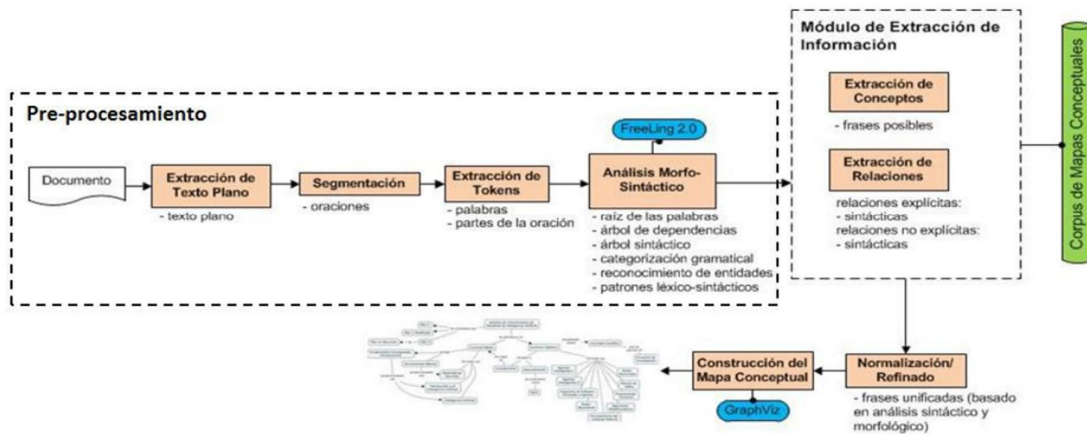


Figura 7. Esquema Funcional del Método de Extracción de Información Estructurada en forma de Mapa Conceptual (Rodríguez Blanco y Simón Cuevas, 2013).

En el caso de las relaciones que se identifican, a diferencia del resto de las propuestas reportadas, es posible identificar relaciones no explícitas en el texto, las cuales se pueden detectar a partir del uso de un corpus de Mapa Conceptual como recursos de conocimiento con el que se puede dotar al procesamiento que se ejecuta en este método. El corpus de mapa Conceptuales también facilita la identificación de frases conceptuales. En la Figura 7 se muestra el esquema general del método.

Agentes racionales

Aspectos organizativos del sistema

Maes (1991: 115) definió a la arquitectura de agentes como

una metodología específica para la construcción de los agentes. En él se especifica la forma en la que el agente se puede descomponer en la construcción de un conjunto de módulos o componentes y cómo estos módulos deben interactuar. El conjunto de los módulos y sus interacciones tiene que dar una respuesta a la cuestión de cómo los datos de los sensores y el estado mental actual del agente determinan las acciones [...] y el futuro estado mental del agente.

Esencialmente pretende sentar los principios básicos para esquematizar los estados internos de un agente. La arquitectura de agentes representa un punto medio entre la

especificación y la implementación (Ponce *et al.*, 2014). Se trata de especificaciones del comportamiento expresadas como propiedades que los agentes deben satisfacer y las principales funciones que determinan el comportamiento, así como las interdependencias que existen entre ellas.

Por otra parte, la arquitectura de la organización determina el marco en el que los agentes actúan. Son definidas las relaciones estructurales (roles, grupos, jerarquías), las limitaciones y formas en el comportamiento de los agentes y sus interacciones, y como colaboran en el cumplimiento de sus objetivos. Puede haber varias formas de estructurar una organización, por ejemplo, de acuerdo a necesidades funcionales (Pavón Mestras *et al.*, 2006).

Existen numerosos trabajos sobre los aspectos organizativos de los sistemas de agentes que plantean distintos modelos teóricos y arquitecturas, clasificados en: arquitecturas basadas en la lógica (agentes deductivos), arquitecturas reactivas (agentes reactivos), arquitecturas en capas (agentes híbridos) y arquitecturas deliberativas (agentes de razonamiento procedural y agentes BDI) (Wooldridge, 2009).

Arquitectura BDI

De estos, el modelo más extendido se basa en los sistemas intencionales, el denominado Deseos Creencias Intenciones (en inglés, *Believes Desires Intentions* o BDI), formulado por Bratman (Bratman, 1987) y formalizado por Kinny, Georgeff y Rao (Kinny *et al.*, 2000), el que representa el accionar de un agente a partir del conocimiento que posea, las metas y sus intenciones. La importancia del modelo no se suscribe al marco teórico, sino que se han presentado diferentes plataformas que implementan esta arquitectura.

La arquitectura BDI se inspira en un modelo cognitivo humano (Bratman, 1987), donde las creencias de los agentes están dadas por la información que reciben del entorno, los deseos están constituidos por aquellos estados que el agente quiere que se verifiquen en su mundo y las intenciones son un subconjunto de los deseos que el agente se propone alcanzar (Cataldi *et al.*, 2006).

En este modelo, un agente BDI es un tipo de agente deliberativo, éste debe decidir ante cada entrada que acción seguir en orden a satisfacer los objetivos. Los objetivos que decide alcanzar son sus intenciones, las cuales juegan un papel crucial en el proceso de razonamiento

práctico. Esta arquitectura, la describiré acorde a la conceptualización de Ponce, Torres, Aguilera y otros (Ponce *et al.*, 2014), esta incluye dos importantes procesos:

- Deliberación: decide en cada caso que objetivos perseguir.
- Razonamiento de medios-fines: determina la forma en que se alcanzarán dichos. objetivos

Y se caracteriza porque los agentes están dotados de estados mentales, incluyendo las correspondientes estructuras de datos:

- Creencias (*Beliefs*): representan el conocimiento que el agente tiene sobre sí mismo y sobre el entorno.
- Deseos (*Desires*): son los objetivos que el agente desea cumplir.
- Intenciones (*Intentions*): se puede considerar como un subconjunto de deseos consistentes entre sí que el agente decide alcanzar. Las intenciones derivan en las acciones que ejecutará el agente en cada momento.

El proceso de un agente BDI, además de los tres componentes principales del modelo (*Believes, Desires e Intentions*) posee los siguientes componentes:

- Un conjunto de creencias actuales: *Bel*
- Una función de revisión de creencias (*belief revision function* o *brf*), que recoge la entrada y el estado actual de creencias del agente para determinar un nuevo conjunto de creencias (donde \wp representa el conjunto de partes de un conjunto):

$$brf: \wp(Bel) \times P \rightarrow \wp(Bel)$$

- Una función generadora de opciones (*generate options*), que determina qué opciones están disponibles para el agente (sus deseos), en base a sus creencias e intenciones actuales:

$$opciones : \wp(Bel) \times \wp(Int) \rightarrow \wp(Des)$$

- Una función de filtrado (*filter*), que representa el proceso de deliberación del agente en la que el agente determina sus intenciones, basado en sus actuales creencias, deseos e intenciones:

$$filtrado : \wp(Bel) \times \wp(Des) \times \wp(Int) \rightarrow \wp(Int)$$

- Una función selectora de acciones (*actions*), la cual determina en cada caso qué acción llevar a cabo en base a las intenciones vigentes:

$$ejecutar : \wp(Int) \rightarrow A$$

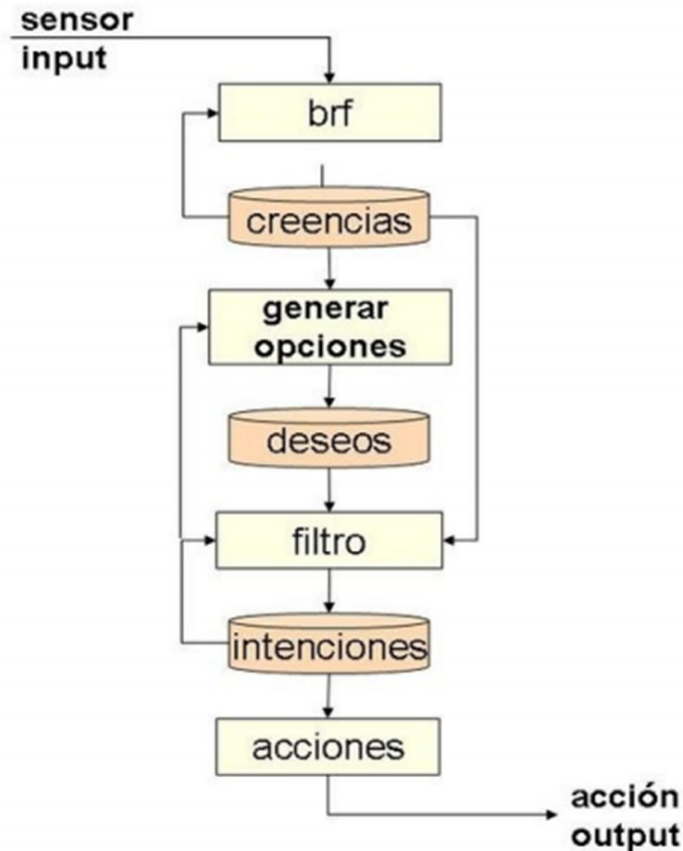


Figura 8. Diagrama de una arquitectura BDI genérica.
(Ponce *et al.*, 2014: 63)

“El proceso resultante de un agente BDI es una función acción: $P \rightarrow A$, la cual para cada percepción (elemento del conjunto P) decide una acción (del conjunto A), y puede definirse con el siguiente pseudo código” (Ponce *et al.*, 2014: 64):

```

function accion (p:P):A
begin
  B:= brf (B, p)
  D:= opciones (B, I)
  I:= filtrado (B, D, I)
  return ejecutar(I)
end function accion
  
```

Desde la perspectiva temporal, un agente puede modelar el mundo como es o como que fuera. En términos de la taxonomía BDI, las creencias son modelos del mundo tal como es ahora, mientras que los deseos y las intenciones se refieren a un estado futuro del mundo (Parunak, 1998).

Características de los agentes individuales

Al describir los agentes individuales, es importante comprender no sólo como se asignan los agentes al dominio del problema y modelan su entorno, es igualmente importante cómo se organizan internamente. Al comparar diferentes alternativas de construcción, es útil distinguir no sólo las particularidades de la organización en general sino también las propiedades de los agentes individuales.

Si bien los agentes tienen características similares, se diferencian entre sí por su estructura interna. Estos varían en complejidad (que tan sofisticada es la representación del conocimiento de su entorno) y los algoritmos que implementan (el razonamiento que utilizan para modelar su mundo). Van Dyke Parunak (1998) estableció cuatro dimensiones donde se pueden observar las diferencias:

- Semejanza (*similarity*): Los agentes que no se comunican directamente entre sí, sino que simplemente interactúan a través de sus efectos en el mundo, pueden diferir ampliamente en su estructura interna. Sin embargo, si se comunican entre sí, deben hablar el mismo lenguaje. A veces los agentes comparten el mismo código y solo difieren en los parámetros de estado. Sin embargo, lo más usual es que comparten los encabezados (*modular heads*) incluidas al menos interfaces de comunicación y quizás otras capacidades, pero el código interno (*bodies*) es arquitectónicamente distinto. Así podemos distinguir entre agentes diferentes, idénticos y los cuerpo-cabeza (*body-head agents*). Éste último representa el caso en que los encabezados de los agentes son iguales pero su cuerpo no.
- Modularidad (*modularity*): El modelo que distingue los encabezados de los cuerpos de los agentes (*body-head*) permite la reutilización de los códigos de comunicación y modelado entre los agentes que son disímiles.
- Memoria (*memory*): Los agentes pueden o no retener un histórico de los cambios que sufrieron en su estado según sus experiencias.
- Mutabilidad (*mutability*): En algunos sistemas, el código del agente puede cambiar a lo largo de su vida. En este contexto, por código se entiende una estructura de datos que se está manipulando en un momento determinado. Una secuencia lineal simple de instrucciones no cuenta; debe haber alguna ramificación o toma de decisiones. La modificación puede ser impuesta al agente desde fuera o iniciada internamente.

Sistema multiagente

Un agente racional es una entidad que hace lo correcto para cumplir con sus metas (...) buscan lograr la cooperación de un conjunto de agentes autónomos para la realización de una tarea. La cooperación depende de las interacciones entre los agentes e incorpora tres elementos: la colaboración, la coordinación y la resolución de conflictos. (Federico, 2008: 23).

Existen dos enfoques para construir un SMA: el enfoque clásico que consiste en dotar de los agentes de la mayor inteligencia posible utilizando descripciones formales del problema a resolver y de hacer reposar el funcionamiento del sistema en las capacidades cognitivas individuales. Por ejemplo, el encuadre típico de un modelo conductista de este tipo que aplica métodos de inteligencia artificial que hacen de los agentes capaces de aprender por refuerzo (Bosello, 2020; Feliu, 2013). Y el enfoque constructivista, que persigue la idea de brindarle inteligencia al conjunto, para que a través de los mecanismos de interacción entre agentes, el sistema mismo genere comportamiento inteligente (Azabache Paredes, 2013; Federico, 2008).

El segundo enfoque está basado en lo que se denomina un *dumb agent* (agente tonto), propone la formulación de sistemas basados en agentes que realizan tareas simples y que no recurren a métodos de IA, sino que dicha inteligencia se genere a partir del accionar del conjunto. El comportamiento de cada agente en el sistema puede ser el mismo o distinto y el sistema se evalúa de acuerdo al cumplimiento del objetivo común. En general,

los dumb agentes no tienen memoria, ni son capaces de aprender, sin embargo, son útiles al momento de desarrollar sistemas que impliquen un gran número de agentes y donde la toma de decisiones no es el principal objetivo. En estos casos, [...] resultan beneficiosos en tiempo y uso de memoria, (sin embargo) no deben ser vistos como un simple procedimiento que ejecute tareas sino que puede ser implementado de diversas maneras dependiendo del objetivo que se desee alcanzar. (Federico, 2008: 25).

Existe una amplia variedad de plataformas y técnicas, algunas muy recientes y novedosas, para el desarrollo de sistemas de agentes inteligentes. Esta un área de investigación en desarrollo continuo, de madurez creciente, sin mencionar la creación de nuevos y avanzados modelos. El trabajo de Constantin-Valentin Pal (2020) al respecto nos permite tener una revisión pormenorizada y actualizada de las plataformas existentes, incluyendo también una perspectiva histórica de los casi treinta años en que se han venido desarrollado dicho campo de investigación.

Modelo de agente

Uno de los objetivos propuestos es la aplicación del modelo de agentes en pos de una estrategia que busca alcanzar un mayor refinamiento y control en la operación de descubrimiento de conocimiento en texto, a la vez que logra reducir la complejidad de la representación resultante, con ello deviene una mayor velocidad en la detección de patrones y menor consumo de recursos.

Tomando como modelo el trabajo de Henríquez Miranda y Fernando Federico, se propone la aplicación del modelo de agente a los nodos, planteando sus *creencias, deseos e intenciones*. Desde un punto de vista general, la idea es que tanto en caso del mapa como el grafo sea entendido como un conjunto de agentes simples interconectados entre sí. Al definir estos como “simples”, se habla de agentes en términos del enfoque constructivista dentro de los sistemas multiagentes, materia que se abordó previamente. Vale decir

que cada nodo (agente) no lleva a cabo un grupo complejo de acciones sino que actúa a partir de primitivas sencillas que permiten, a escala global, desarrollar una actividad compleja. La interconectividad de los agentes se asocia principalmente a los arcos del grafo (o la frase o palabra de enlace en caso del mapa conceptual), sin embargo, el accionar de cada vértice es independiente de resto por lo que, los nodos, no comparten información sobre el medio que los circunda. (Federico, 2008: 53)

El funcionamiento interno de los agentes se especifica a continuación.

Creencias

Cada agente (en tanto nodo único) tiene una visión distinta de su entorno, por tanto un conjunto diferente de creencias. Es por ello que las variables asociadas al conocimiento del agente estarán asociadas a un nodo en particular, no poseen carácter global. En principio, pueden diferenciarse *en variables de definición de frecuencias y variables de peso*. Las primeras están relacionadas al manejo del tiempo mientras que las segundas conciernen a la comparación de arcos o enlaces, según corresponda.

Existen dos variables de frecuencia. Por un lado, aquella que define el número de instancias del nodo y por otro la que especifica el número de instancias que se debe aguardar para poder realizar una eliminación de arcos o enlaces. En lo concerniente a estos, cada nodo

tiene conocimiento del conjunto que parte de él. A partir de este conjunto se define el segundo tipo: las variables de peso, que depende de la instancia del nodo y es la utilizada para la comparación de los arcos.

La comunicación entre agentes se lleva a cabo sólo una vez terminado el procesamiento del texto a fin de dar a conocer los arcos o enlaces más relevantes que tienen en posesión, puntapié inicial para el proceso de reducción de caminos irrelevantes.

Deseos

A diferencia de las creencias, todos los nodos (agentes) poseen la misma meta. El objetivo principal radica en la maximización del promedio de los pesos que salen de él. [...] Una buena heurística para el cumplimiento de este objetivo radica en la selección de las mejores instancias (del conjunto de arcos salientes). De esta manera se busca eliminar los caminos irrelevantes [...] y poder, por lo tanto, detectar los patrones del texto de forma más rápida. (Ale y Federico, 2007: 10).

Intenciones

El plan de cada agente se encuentra regido principalmente por sus variables de frecuencia. El procedimiento básico se resume en la eliminación de arcos irrelevantes según lo determinado en el [...] proceso de selección. [...] Una vez finalizado el documento a procesar, cada nodo [...] comunicará sus relaciones fuertes al resto para la determinación de los patrones existentes. (Federico, 2008: 54)

Conclusiones

La mayoría de los sistemas KDT emplean estructuras intermedias sencillas para significar el contenido de los textos lo que facilita el análisis, pero a la vez limita la variedad, expresividad y utilidad de los patrones descubiertos, las representaciones se limitan a un nivel temático.

Una gran parte de los métodos actuales de minería de texto utilizan simples representaciones del contenido de los textos, especialmente como vectores de frecuencias de términos. Estas son relativamente fáciles de construir a partir de los textos, pero no pueden

expresar varios detalles de sus significados, presentando una pobre capacidad de descripción. Para lograr una mejor interpretación del conocimiento contenido en los textos, se han propuesto formatos más complejos (Guevara Martínez *et al.*, 2006).

En este trabajo de investigación se plantea la necesidad de definir una forma intermedia superadora, una que considere la información estructural de los textos procesados. Para ello se propone analizar dos métodos para el agrupamiento conceptual de una colección de textos. Más aún, sean estos capaces de trasladar los descubrimientos de conocimiento almacenado en texto del actual nivel temático a un nivel de mayor detalle, un nivel conceptual. Dos modelos a ser aplicados en los procesos del Descubrimiento de Conocimiento en Textos, a través del cual se puede representar los conceptos y las relaciones entre ellos según sea su proximidad dentro del texto.

En específico, se expone el uso de los grafos y mapas conceptuales como la representación del contenido de los textos. Se presentaron varios conceptos acerca de la construcción automática a partir de texto de estas representaciones, seleccionando en cada caso una propuesta para dicho proceso. Las mismas se ilustran de manera breve y sencilla, omitiendo de momento todo detalle algorítmico. Si el lector está interesado en estos detalles debe referirse a las publicaciones citadas.

Asimismo, la inclusión de la teoría de agentes tiene por objeto reducir el volumen del grafo o mapa mediante la reducción de caminos irrelevantes, mayor flexibilidad en el tipo de patrón que se desea detectar y la disminución de los tiempos de procesamiento.

Trabajo a futuro

El trabajo subsecuente se centrará en el eje práctico, como parte del mismo se intentará definir el diseño de una arquitectura genérica, la implementación de los algoritmos de IA y el desarrollo de un *framework* (a nivel prototipo) potencialmente implementable. Respecto de los agentes, una meta en estudio es la aplicación a un escenario concreto tanto del modelo *dumb agent* como el método de *reinforcement learning*.

Otro posible objetivo podría ser la definición, perfeccionamiento (por ejemplo mediante

la integración de mayor información como son las relaciones de hiperonimia, hiponimia y meronimia) o creación de mejores recursos léxicos de dominio público en lengua castellana, como son los diccionarios electrónicos y los corpus de entrenamiento, para mejorar la efectividad de los sistemas de extracción de información estructurada desde textos escritos en idioma castellano.

Sin duda la lingüística computacional tiene una gran utilidad práctica inmediata, ya que se relacionan con la toma de decisiones, la búsqueda e intercambio de conocimiento, y toda clase de operaciones relacionadas con la publicación y uso de los documentos. Desgraciadamente, la mayoría de los logros actuales de la lingüística computacional se orientan al inglés (Montes, 2014).¹

Bibliografía

- Ale, J. M. & Federico, F. C., (2007), “Aplicación de la teoría de Agentes al modelo de grafos para la detección de patrones en Textos”, XIII Congreso Argentino de Ciencias de la Computación, pp. 1347-1359.
- And, A., Kinny, D., Georgeff, M. & Rao, A., (2000), “A Methodology and Modelling Technique for Systems of BDI Agents”, Technical Report 59, Australian Artificial Intelligence Institute, Melbourne, Australia.
- Ausubel, D. y Novak, J. D., (1989), *Psicología Educativa*, México, Trillas.
- Azabache Paredes, J. E., (2013), *Desarrollo de un agente inteligente ontológico para la clasificación y recuperación de información textual de documentos*, Escuela Académico Profesional de Informática, Facultad de Ciencias Físicas y Matemáticas Universidad Nacional de Trujillo, Perú.
- Baray, C. y Wagner, K., (1999), *Where Do Intelligent Agents Come From?*, Crossroad ACM.
- Bordón, L, D’Avanzo, E., (2004), “Perspectivas para la integración de la minería de textos y la gestión del conocimiento”, The IPTS Report. Número 85.
- Bosello, M., (2020), “Integrating BDI and Reinforcement Learning: the Case Study of Autonomous Driving”, 10.13140/RG.2.2.13072.23044.
- Botta-Ferret, E., Cabrera-Gato, J. E., (2007), “Minería de textos: una herramienta útil para mejorar la gestión del bibliotecario en el entorno digital”, ACIMED. Disponible en http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352007001000005
- Bratman, M., (1987), *Intentions, plans and Practical Reasons*, Harvard University Press,

1 De momento se tienen en consideración los siguientes: Base de datos de morfología del español (BDME) en <https://morfogen.iatext.ulpgc.es/bdme/index.html> y <https://bdme.iatext.es>; Red semántica y ontológica multilingüe lexicalizada en <https://babelnet.org>; Multilingual Central Repository de Euskal Herria Unibertsitatea y Universitat Politècnica de Catalunya en <https://adimen.si.ehu.es/web/MCR>; MultiWordNet de la Fondazione Bruno Kessler, Center for Communication and Information Technology, Human Language Technology Group en <http://multiwordnet.fbk.eu/english/home.php>; UWN/MENTA. Base de conocimientos léxica plurilingüe automática que extiende WordNet en <https://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/uwn>; Spanish WordNet 3.0 de la Red Temática en Tratamiento de la Información Multilingüe y Multimodal en <http://timm.ujaen.es/recursos/spanish-wordnet-3-0/>; LAS-WordNet. Una WordNet para el español obtenida con traducción automática en <https://www.datos.gov.co/Ciencia-Tecnolog-a-e-Innovaci-n/LAS-WordNet-una-WordNet-para-el-espa-ol-obtenida-c/8z8d-85m7>; AncoraNet. Léxico multilingüe que combina información sintáctico-semántica y conceptual en <http://clic.ub.edu/corpus/es/ancoranet>.

- Cambridge, MA.
- Cataldi, Z., Salgueiro, F. A., Costa, G., Calvo, P., Méndez, P., Rendon Zander, J., & Lage, F. J., (2006), *Sistemas tutores inteligentes basados en agentes*, Laboratorio de Informática Educativa y Medios Audiovisuales, Facultad de Ingeniería, Universidad de Buenos Aires.
- Chang, J. Y., & Kim I. M., (2014), “Research Trends on Graph Based Text Mining”, *Int. J. of Smart Home*, 8(4), page 37-50.
- Chantal Pérez Hernández, M., (2002), “Explotación de los corpórea textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento”, *Estudios de Lingüística del Español (ELiEs)*, Volumen 18, Universidad de Málaga, España. Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=1219421>
- Charniak, E., Johnson, M., (2005), “Coarse-to-fine n-best parsing and Maximum Entropy discriminative reranking”, *ACL'05*.
- Ciliberti, N.; Galagovsky, Lydia R., (1999), “Las redes conceptuales como instrumento para evaluar el nivel de aprendizaje conceptual de los alumnos. Un ejemplo para el tema de dinámica”. *Enseñanza de las ciencias: revista de investigación y experiencias didácticas*, [en línea], 1999, Vol. 17, nro. 1, pp. 17-29. Disponible en: <https://www.raco.cat/index.php/Ensenanza/article/view/21557>
- Ciravegna et al., (2001), Ed. Proc. of the 17Th International Joint Conference on Artificial Intelligence (IJCAI-2001), “Workshop of Adaptive Text Mining”, Seattle, WA.
- Constantin Valentin Pal, Florin Leon, Marcin Paprzycki and Maria Ganzha, (2020), “A Review of Platforms for the Development of Agent Systems”, Faculty of Automatic Control and Computer Engineering, “Gheorghe Asachi” Technical University of Iași, Romania.
- Delgado, M., Martini Bautista, M., Sanchez, D., and Vila M., (2002), *Mining text data: Special features and patterns, pattern detection and discovery*, In *Proceedings ESF Exploratory Workshop*.
- Don Gilbert, (1997), *Intelligent Agents: The Right Information at the Right Time*, IBM Corporation Research, Triangle Park, NC, USA.
- Dubois, V. and Quafafou, M., (2002), *Incremental and dynamic text mining*, *ISMIS*, 2002 page 265-273.
- Estrada, E., (2011), *CMAG: Herramienta para la Construcción Automática de un Mapa Conceptual a Partir de un Texto No Estructurado en Lenguaje Natural*, Tesis de Diploma, Facultad de Ingeniería Informática, Instituto Superior Politécnico “José Antonio Echeverría”.
- Febles Rodríguez, J., González Pérez, A., (2002), “Aplicación de la minería de datos en la bioinformática”, *Acimed*, 202; 10(2). Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352002000200003 [Consultado: 16 de febrero de 2021].
- Federico, F. C., (2008), *Aplicación de la teoría de Agentes al modelo de grafos para la detección de patrones en Textos*, Disponible en: <http://materias.fi.uba.ar/7500/federico-tesisdegradoingenieriainformatica.pdf>
- Feldman, E., (1999), Proc. of The 16th International Joint Conference on Artificial Intelligence (IJCAI-1999), “Workshop on Text Mining: Foundations, Techniques and Applications”, Stockholm, Sweden.
- Feldman & Dagan, (1995), “Knowledge Discovery in Textual databases (KDT)”, Proc. of the 1st International Conference on Knowledge discovery (KDD_95).
- Feldman, R., Fresko, M., Kinar, Y., Lindell, Y., Liphstat, O., Rajman, M., Schler, Y., Zamir, O., (1998), “Text Mining at the Term Level”, Proc. of the 2nd European Symposium on Principles of Data Mining and Knowledge Discovery (PKDD'98).
- Feldman, R., Aumann, Y., Schler, J., Landau, D., Lipshtat, O., and Ben-Yehuda, Y., (2002), “Term-level text with mining with taxonomies”, *US Patent* 6,442,545.
- Feliu, J. L., (2013), *Use of Reinforcement Learning (RL) for plan generation in Belief Desire Intention (BDI) agent systems*. Disponible en <https://digitalcommons.uri.edu/theses/160/>
- Franklin Stan y Graesser Art, (1996), “Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents”, *Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages*, Springer Verlag.
- Frawley, W. J., Piatetsky Shapiro, G., and Matheus, C. J., (1992), “Knowledge discovery in databases: An overview”, *AI magazine*, 13(3):57.

- Galagovsky, L.R., (1993), “Redes conceptuales: su base teórica e implicancias para el proceso de enseñanza-aprendizaje de las ciencias”, *Enseñanza de las Ciencias*, 11, pp. 301-307.
- Galagovsky, L.R. y Ciliberti, N., (1994), “Redes conceptuales: su aplicación como instrumento didáctico en temas de física”, *Enseñanza de las Ciencias*, 12, pp. 338-349.
- Galagovsky, L.R., (1996), *Redes conceptuales. Aprendizaje, comunicación y memoria*, Argentina: Lugar Editorial.
- Gelfand, B., Wulfekuhler, M., and Punch, W., (1998), “Discovering concepts in raw texts: Building semantic relationship graphs”, In ICML/AAAI workshop on learning for text categorization.
- Gómez Sanz, Jorge J., (2003), “Metodologías para el desarrollo de sistemas multiagente”, *Inteligencia artificial: Revista Iberoamericana de Inteligencia Artificial*, ISSN 1137-3601, N° 18, pp. 51-64.
- Guevara Martínez, E., Medina Pagola, J.E., Hernández Palancar, J., (2006), “Nuevo modelo de representación de textos: grafos de asociación”, *Ingeniería Industrial*. pp. 27.
- Gupta, V., & Lehal, G. S., (2009), “Survey of Text Mining Techniques and Applications”, *Journal of Emerging Technologies in Web Intelligence*, 1(1), pp. 60-76.
- Hearst, M. A., (1999), “Untangling text data mining”, In Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics, pp 3–10, Association for Computational Linguistics.
- Henríquez Miranda, C. N., (2009), “Sistema multiagentes para la descripción, localización, recuperación y acceso a documentos”, *PROSPECTIVA*, vol. 7, núm. 1, enero-junio, 2009, pp. 31-40. Universidad Autónoma del Caribe. Disponible en: <https://www.redalyc.org/articulo.oa?id=4962/496250975005>
- Jennings, N. R., (1999), “On agent-based software engineering”, Department of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, UK
- Justicia de la Torre, M., Sanchez Fernandez, D., Blanco Medina, I. J., y Martin Bautista, M. J., (2008), *Text knowledge mining: An approach to text mining*, ESTYLE, España.
- Justicia de la Torre, M., (2017), *Nuevas técnicas de minería de textos: Aplicaciones*, Granada: Universidad de Granada. Disponible en: <http://hdl.handle.net/10481/46975>
- Kinny D., George M., Rao A., (1996), “A Methodology and Modelling Technique for Systems of BDI Agents”, Proceedings of the 7th European workshop on Modelling autonomous agents in a multi-agent world, , 56-71, ISBN:3-540-60852-4 Environmental Research Institute of Michigan (ERIM).
- Kowata J. H., Cury D. y Boeres M. C. S., (2010), “Concept Maps Core Elements Candidates Recognition From Texts”, en Proc. of Fourth International Conference on Concept Mapping (CMC’10), Viña del Mar, Chile.
- Kumar, S., Agrawal, M., Rajput, S., (2014), “An Information Retrieval (IR) Techniques for text Mining on web for Unstructured data”, *Int. J. of Adv. Research in Comp. Science and Software Engineering*, 4(2), 67-70.
- Lamarca Lapuente, M. J., (2018), *Hipertexto: El nuevo concepto de documento en la cultura de la imagen*, Universidad Complutense de Madrid. Disponible en: <http://www.hipertexto.info>
- Maes, Pattie, (1991), “The Agent network architecture (ANA)”. *SIGART Bull.* 2, 4 (Aug. 1991), 115–120. Disponible en: <https://doi.org/10.1145/122344.122367>
- Mazo I, Ortiz de Frutos E., (1998), *KMAT como herramienta de análisis de gestión del conocimiento*, Clúster del conocimiento, Bilbao: PMP.
- Medina, J., Guevara, E., Hernández, J., Hechavarría, A., & Hernández, R., (2005), “Similarity Measures in Documents using Association Graphs”, Proceedings of CIARP’05, LNCS, vol. 3773, pp 741–751, Springer Berlin Heidelberg.
- Miranda, S., Gelbukh, A., & Sidorov, G., (2014), “Generación de resúmenes por medio de síntesis de grafos conceptuales”, *Revista signos*, 47(86), 463-485. <https://dx.doi.org/10.4067/S0718-09342014000300006>
- Mishne, G., (2003), *Source code retrieval using conceptual graphs*, Master of logic thesis, Institute for Logic, Language and Computation, University of Amsterdam.
- Montes, M., Gelbukh, A., López López, A., Baeza Yates, R., (2001), “Un método de agrupamiento de grafos conceptuales para minería de texto”, *Procesamiento del lenguaje natural*, Vol. 27, pp. 115-

122.

- Montes y Gomez, M., Gelbukh, A., and Lopez Lopez, A., (2002), "Text mining at detail level using conceptual graphs", *Lecture Notes in Artificial Intelligence* 2393.
- Montes, M., (2014), "Minería de Texto: Un nuevo reto computacional", Laboratorio de Lenguaje Natural, Centro de Investigación en Computación, Instituto Politécnico Nacional. México. Disponible en: <https://ccc.inaoep.mx/~mmontesg/publicaciones/2001/MineriaTexto-md01.pdf>
- Nahm, U. Y. & Mooney, R. J., (2002), "Text mining with information extraction", AAAI. In Spring Symposium on Mining Answers from Texts and Knowledge Bases.
- Novak, J. D., Gowin, D. B., (1988), *Aprendiendo a aprender*, Martínez Roca, España.
- Novak, J. D., Cañas, A. J., (2006), "The Origins of the Concept Mapping Tool and the Continuing Evolution of the Tool", *Information Visualization Journal*, 5 (3), 175-184.
- Novak J. D., Cañas A. J., (2008), "The Theory Underlying Concept Maps and How to Construct Them", Technical Report IHMC CmapTools 2006-01 (Rev 2008-01), Florida Institute for Human and Machine Cognition, Pensacola FL, 32502, USA.
- Oliver González, A., (2016), *Herramientas Tecnológicas para Traductores*, Editorial UOC.
- Ontoria, A., (1995), *Mapas conceptuales. Una técnica para aprender*, 5ª ed. España: Narcea.
- Palmeira, C., Chaves, R., Cavalcante H., & Favero, E., (2012), "A Requirements Elicitation and Analysis Aided by Text Mining", *International Journal of Computer Science and Network Security*, 12(6), 122-128.
- Parunak, Van Dyke, (1998), *Practical and Industrial Applications of Agent Based Systems*, Industrial Technology Institute.
- Pavón Mestras, J. L., Menéndez, M. A., Collado, S. H., & Sansores, C. E., (2006), *Simulación de sistemas sociales con agentes software*, Universidad Complutense Madrid, Facultad de Informática, Madrid, España.
- Ponce, J., & Torres, A., & Aguilera, F., & Silva Sprock, A., & Flor, E., & Casali, A., & Scheihing, E., & Tupac, Y., & Torres, D., & Ornelas, F., & Hernández, J. A., & Crizpín & Vakhnia, N., & Pedreño, O., (2014), "Inteligencia Artificial". 10.13140/2.1.3720.0960.
- Ramanathan, V., & Meyyappan, T., (2013), "Survey of Text Mining", *Proceedings of International Conference on Technology and Business Management*, pp. 508-514.
- Rodríguez, H., (1999), "Tutorial de extracción y recuperación de información", *Sepln* 99.
- Rodríguez Blanco, A., & Simón Cuevas, A. J., (2013), "Método para la extracción de información estructurada desde textos", *Revista Cubana de Ciencias Informáticas*, 7(1), 55-67. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2227-18992013000100007
- Rodríguez Blanco, A., Simón Cuevas, A. J., Guevara Martínez, E., Hojas Mazo, W., (2015), "Modelo de representación de textos basado en grafo para la minería de texto", *Ciencias de la Información*, vol. 46, núm. 1, enero-abril, pp. 63-71. Instituto de Información Científica y Tecnológica, La Habana, Cuba.
- Russell, S. J. & Norvig, P., (2020), *Artificial Intelligence: A Modern Approach*, (4th Edition), Prentice Hall.
- Shekhar, Ch., Sharan, A., & Lata, M., (2014), "Semantic Graph Based Approach for Text Mining", *Proceedings of 2014 Int. Conf. Issues and Challenges Intelligent Computing Techniques (ICICT)*, pp. 596-601. IEEE Press.
- Simón A. J., Rosete A., Panucia K. y Ortiz A., (2004), "Aproximación a un método para la representación en Mapas Conceptuales del conocimiento almacenado en textos, con beneficios para la Minería de Texto", I Simposio Cubano de Inteligencia Artificial (SiCIA'04), 10ma Convención y Feria Internacional Informática, C. Habana, Cuba.
- Sowa, (1984), *Conceptual Structures: Information Processing in Mind and Machine*, Addison-Wesley, reading, M.A.
- Sowa & Way, (1986), "Implementing a semantic interpreter using conceptual graphs", *IBM Journal of Research and Development*, 30:1.
- Sumathy, K. L., & Chidambaram, M., (2013), "Text Mining: Concepts, Applications, Tools and Issues An Overview", *International Journal of Computer Applications*, 80(4), 29-32.
- Tan, A. H., (1999), "Text mining: Promises and challenges", In *Pacific Asia Conference on Knowledge Discovery and Data Mining PAKDD'99 workshop on Knowledge Discovery from Advanced Databases*, pages 65-70.

- Thavamani, C. & Rengarajan, A., (2014), “Mining Conceptual Relations from Textual Web Content Using Leximancer”, *IOSR Journal of Computer Engineering*, 16(5), 24-27.
- Tsuyoshi Kitani, Yoshio Eriguchi, Masami Hara, (1994), “Pattern Matching and Discourse Processing in Information Extraction from Japanese Text”, *Journal of Artificial Intelligence Research* 2, pages 89-110.
- Valerio, A., Leake, D. B., (2006), “Jump-Starting Concept Map Construction with Knowledge Extracted from Documents”, en *Proc. Of Second International Conference on Concept Mapping (CMC'06)*, Vol. 1, San José, Costa Rica, 2006, pp. 296-303.
- Valerio, A., Leake, D. B. y Cañas A. J., (2008), “Associating Documents To Concept Maps In Context”, en *Proc. Of Third International Conference on Concept Mapping (CMC'08)*, Vol. 1, Tallinn University, Tallinn, Estonia, pp. 114-121.
- Valero Moreno, A. I., (2017). “Técnicas estadísticas en Minería de Textos”. Universidad de Sevilla. España. Disponible en: [https://idus.us.es/bitstream/handle/11441/63197/Valero Moreno Ana Isabel TFG.pdf](https://idus.us.es/bitstream/handle/11441/63197/Valero%20Ana%20Isabel%20TFG.pdf)
- Weiss, S. M., Indurkha, N., & Zhang, T., (2010), “Fundamentals of predictive text mining”, volume 41, Springer.
- Wooldridge, M., & Jennings, N., (1994), *Intelligent Agents: Theory and Practice*. The Knowledge Engineering Review, 10(2), June 1995, pp. 115 – 152, Cambridge University Press. Disponible en: <https://doi.org/10.1017/S0269888900008122> y <http://www.cs.ox.ac.uk/people/michael.wooldridge/pubs/ker95/ker95-html.html>
- Wooldridge, M. J., & Jennings, N. R., (1995), *Agent Theories, Architectures and Languages: A survey*, M. J. Wooldridge and N. R. Jennings (Eds.), *Intelligent Agents*, Volume 890 of LNAI. pp. 1–39. Springer Verlag.
- Wooldridge, M. J., (1997), “Agent-based software engineering”, *IEE Proc, Software Engineering* 144 (1) (1997) 26–37.
- Wooldridge, M. J., (2009), *Introduction to Multiagent Systems*, 2nd Edition, John Wiley and Sons Ltd.
- Xu, F., Kurz, D., and Piskorski, J., (2002), *Term extraction and mining of term relations from unrestricted texts in the financial domain*, *Proceedings of BIS*.
- Zhong, N., Li, Y., & Wu, S. T., (2012), *Effective Pattern Discovery for Text Mining*, *IEEE Transactions*.