

Método complementario de obtención de los Conceptos Nucleares para la Teoría de los Conceptos Nucleares. El Índice de Nuclearidad.

Juan Ángel Contreras, Juan Arias, Ricardo Luengo y Luis Manuel Casas

Grupo de Investigación CIBERDIDACT,
Departamento de Ingeniería de Sistemas Informáticos y Telemáticos,
Centro Universitario de Mérida, Universidad de Extremadura
C/ Santa Teresa Jornet, 38, 06800-Mérida, Badajoz, España
{jaconvas, juanaria, rluengo, luisma}@unex.es

Resumen: Se refleja en este artículo una nueva forma de obtener el grado de importancia que tienen los conceptos a los que se hace referencia en la teoría de los Conceptos Nucleares. Esta teoría, como herramienta, utiliza las Redes Asociativas Pathfinder para establecer redes cognitivas de conceptos, observando y cuantificando la relación de cada uno de los conceptos con respecto al conjunto de todos los demás conceptos, siendo considerados, como conceptos nucleares, sólo aquellos conceptos que tienen tres o más enlaces. Ello permite discriminar cuáles son los conceptos fundamentales de entre todos los conceptos analizados. Como complemento a ello, y utilizando la matriz triangular, característica de entrada al algoritmo Pathfinder, también se utiliza la misma para crear este nuevo índice que se llamará Índice de Nuclearidad y del que tendremos dos versiones, Completa y Reducida.

Palabras clave: Índice de Nuclearidad; Redes Asociativas Pathfinder; Teoría de los “Conceptos Nucleares”.

Abstract: In this article is reflected a new way to obtain the degree of importance of the concepts referred to in the Nuclear Concepts theory. This theory, use as a tool, the Networks Associative Pathfinder to establish cognitive networks concepts, observing and quantifying the relationship of each of the concepts relative to the total of all other concepts being considered as nuclear concepts, only those concepts with three or more links. This allows to discriminate what are the fundamental concepts of all the concepts are analyzed. To complement this, and using the triangular matrix input feature Pathfinder algorithm, it is also used to create the new index is called Index Nuclearity and we will have two versions, full and reduced.

Key words: Nuclarity Index; Pathfinder Associative Networks; Theory of the “Nuclear Concepts”.

1. Introducción.

En este artículo se añaden nuevos elementos y, por tanto, se introducen mejoras en lo referido a la definición de los conceptos más importante, o nucleares, que se dan en la Teoría de los Conceptos Nucleares [Luengo 13] (TCN en adelante), creada por los doctores D. Ricardo Luengo y D. Luis Manuel Casas en la tesis doctoral de este último [Casas 02] en la Universidad de Extremadura.

Según la TCN, se considera a un concepto como nuclear aquel concepto importante que los usuarios tienen más anclados en su mente y que utilizan para poder realizar definiciones de otros conceptos, es decir, conceptos a través de los cuales se organizan los demás y sirven de base a la estructura cognitiva. Para los autores de TCN, un concepto se considera nuclear cuando tiene 3 o más enlaces utilizando las Redes Asociativas Pathfinder [Shvaneveldt 90] (en adelante

RAP), la cual se utiliza como herramienta para representar las redes cognitivas.

Algunos autores como Fenker [Fenker 75], Jonassen [Jonassen 90], Preece [Preece 76] y Shavelson [Shavelson 72] están de acuerdo en que procedimientos o técnicas similares a las Redes Asociativas Pathfinder, como son: el Análisis de Componentes Principales, el Análisis de Cluster o el Escalamiento Multidimensional, sirven para representar las redes cognitivas que se estudien.

Las RAP se puede crear y representar utilizando distintos programas software, entre los que se encuentran, el programa Goluca [Casas et al. 11]. Es decir, en la representación de la RAP, si un concepto está relacionado con al menos tres conceptos diferentes, o más, será considerado como un concepto importante o nuclear.

La representación de las RAP permiten visualizar de un simple vistazo cuales son las relaciones entre conceptos y cuáles son los conceptos nucleares, cuando existen pocos conceptos. Pero cuando el número de conceptos comienza a ser mayor en la RAP, distinguir cuales son los conceptos nucleares y sobre todo cuáles son más importantes o más nucleares que otros es muy complicado y se hace bastante tedioso y engorroso. Este fue uno de los problemas y conclusiones obtenidas en el trabajo que podemos encontrar en [Contreras et al. 14a]. En el mismo trabajo, utilizando una metodología cualitativa, se estudiaron las memorias de verificación de los títulos de Grado en Informática de 22 universidades españolas, con el objetivo de buscar los conceptos básicos o fundamentales que deben estudiar los alumnos en la materia de Bases de Datos. Se obtuvieron como resultado un total de 90 conceptos y se analizaron utilizando una metodología cuantitativa que proporcionan los indicadores en las redes RAP; determinándose que 70 de todos estos conceptos eran los conceptos básicos o nucleares. Al ser tantos los conceptos estudiados, se necesitaba graduar cuales eran más importantes que otros en aquellos casos en el que el número de enlaces en la RAP fuese el mismo.

Por otro lado, es necesario también diferenciar, por alguna otra justificación, en el caso de tener distintos conceptos el mismo número de enlaces, cuál de ellos tiene más importancia que otro. Lo cual posibilitará la

creación de jerarquías de importancia entre todos los conceptos.

Es por ello, por lo que en este artículo proponemos otra forma de identificar la Nuclearidad de los conceptos de tal forma que se pueda distinguir, ya no sólo, cuáles son los conceptos nucleares, sino definir también el grado de Nuclearidad entre ellos, permitiendo así calibrar la importancia de cada uno con respecto a los demás. Continuando el trabajo realizado en [Contreras et al. 14a] se va a definir un nuevo elemento de discriminación de los conceptos que se llamará Índice de Nuclearidad para los conceptos nucleares.

A continuación, se verán los fundamentos de la TCN junto con los cuantificadores que proporcionan las RAP. Posteriormente se definirán los elementos necesarios para generar las formulaciones de los índices de Nuclearidad. Se formularan dos índice de Nuclearidad, uno en la versión completa y otro en su versión reducida, tal que pueden ser utilizados en distintos entornos en función de la información disponible. En el apartado 5 pondremos una ejemplificación práctica de su uso. Y por último, incluiremos unas conclusiones finales.

2. Antecedentes, la TCN y las RAP.

2.1. Antecedentes.

Los fundamentos teóricos de la TCN [Casas y Luengo 05] se basan en el marco teórico general de la Ciencia Cognitiva, en sintonía con [Neisser 76], [Piaget 78], [Ausubel et al. 83], y más concretamente, con [Shavelson 72], en cuanto a la noción de estructura cognitiva, y con muchas de las ideas de [Rumelhart 84], en cuanto a organización de la memoria en base a esquemas o representaciones mentales relacionadas en la "Teoría de los Esquemas". Por último, en cuanto a la representación del conocimiento utilizamos las RAP [Shvaneveldt 90].

La TCN lleva además asociada una técnica que permite la representación de la estructura del conocimiento de una manera analítica y gráfica que proporciona información acerca de cómo se produce el aprendizaje en función de los cambios de la estructura

cognitiva, con un procedimiento de obtención de datos no invasivo.

La reflexión sobre las formas en que el conocimiento es adquirido y almacenado en la estructura cognitiva del alumno, y cómo puede ser representado, tiene implicaciones en la enseñanza y la investigación educativa. La estructura cognitiva la definimos como un constructo hipotético que se refiere a la organización de las relaciones entre los conceptos en la memoria semántica o memoria a largo plazo [Shavelson 72].

La TCN [Casas y Luengo 04a], [Casas y Luengo 04b] y [Casas y Luengo 05] ofrece un punto de vista para aproximarse a la comprensión de cómo se adquiere y organiza el conocimiento y cómo se puede caracterizar la estructura cognitiva. Se trata de una nueva perspectiva para explicar cómo los procesos de aprendizaje se producen en la mente humana. Se corresponde al marco teórico general de la Ciencia Cognitiva con la que se sintoniza en muchas cuestiones, aunque como se verá más adelante también se difiere en algunas otras.

Para los doctores Casas y Luengo, los creadores de la TCN, esta teoría es una *"propuesta de integración"*. La razón es que tiene su fundamento en otras anteriores, particularmente las de [Piaget 78] o [Ausubel et al. 83], basándose en muchas de sus ideas, y recogiendo aportaciones de otros campos y de los resultados de investigación.

Por lo que se refiere a las teorías de Piaget, los autores creen que *"su aportación permite un análisis muy detallado de los fenómenos que influyen en adquisición del conocimiento, como por ejemplo el de la acomodación, que Piaget utiliza, pero que no describe en profundidad. En cuanto a las de Ausubel, la TCN propone modificaciones, basándose en datos experimentales, de algunos de sus presupuestos, pero su propuesta se basa en ellas y las integra aprovechando los datos experimentales y las aportaciones de otras Ciencias"*.

Los elementos fundamentales del modelo de la TCN, en comparación con las teorías anteriores, se esquematizan en la **Figura 1**. En la **Figura 2**, se muestra cuáles son las herramientas o técnicas con la que cuenta cada una de las teorías para el desarrollo de la misma. Así, para la TCN se dispone de al menos dos

herramientas básicas como son: las RAP y el programa Goluca.



Figura 1. Elementos de la Teoría de los Conceptos Nucleares (TCN).



Figura 2. Herramientas asociadas con las distintas teorías.

La TCN se ha aplicado con buenos resultados en distintas ramas de la ciencia como por ejemplo: en Matemáticas [Verissimo 13], Ingeniería [Contreras 16] y educación [Luengo et al. 11].

Los puntos en los que difieren la TCN con las ideas de Piaget, Ausubel y Novak proponiendo alternativas nuevas los recogemos en los siguientes apartados.

2.2. TCN.

2.2.1. Organización geográfica del conocimiento.

Según describen los autores de la TCN en [Casas 02], [Casas y Luengo 03], [Casas y Luengo 04a], [Casas y Luengo 04b], [Casas y Luengo 05] y en [Casas y Luengo 13], su nueva propuesta teórica parte de una idea muy sencilla, la adquisición del conocimiento y su almacenamiento en la estructura cognitiva, en términos generales, sigue un proceso análogo a la adquisición del conocimiento del entorno físico.

También describen que esta adquisición se puede producir en tres etapas:

- Por la adquisición de ciertos “hitos”.
- Por la adquisición del conocimiento de una ruta, que es la capacidad de navegar en un hito a otro.
- Por la adquisición de una visión de conjunto, es decir, conocer todas las conexiones entre los hitos a través de las rutas.

La primera contribución que hace esta nueva teoría se refiere a como el individuo organiza su conocimiento dentro de la estructura cognitiva, que es en una forma geográfica, lo cual viene a contradecir lo que Ausubel menciona sobre este tema y el aprendizaje significativo, quien describe que los conocimientos se adquieren bajo estructuras jerárquicas.

2.2.2. Conceptos Nucleares.

Los autores de estas teorías anteriores proponen que hay ideas de nivel superior, llamadas inclusores que sirven como anclaje para otras. A estas ideas es a las que se refieren al afirmar que es necesario, para lograr un mejor aprendizaje y la retención del material lógicamente significativo y nuevo, la disponibilidad dentro la estructura cognoscitiva de ideas de afianzamiento específicamente pertinentes a un nivel de inclusividad adecuado.

Los autores de la TCN remarcen esta idea: “*El aprendizaje no tiene por qué producirse en estadios de mayor a menor inclusividad*”, y creen que no es esa la forma general del aprendizaje. El aprendizaje, en este sentido, “*es un proceso de ajuste de las representaciones mentales del alumno*”.

Para los autores de la TCN, los conocimientos no se van organizando a partir de conceptos más inclusivos a otros más sencillos. Esto quizá ocurra al final, cuando se tiene una visión de conjunto, pero no al principio del conocimiento. Se produce por un sistema de “acrecentamiento”, tal como el señalado por Rumelhart [Rumelhart 80], primero hitos del paisaje, después rutas y después visión general del mapa.

Por lo tanto, según esta concepción, y dado que no consideran que el aprendizaje se apoye siempre en una estructura jerárquica, no tiene por qué haber conceptos

ni más importantes ni de menor nivel, sino que hay simplemente conceptos que sirven como anclaje la estructura cognitiva del alumno. La cuestión clave para la práctica educativa sería identificar cuáles son los “hitos” en el territorio de conocimiento en que se mueve el alumno, sus “Conceptos Nucleares”.

2.2.3. Senderos de mínimo coste.

Los autores de la TCN, al analizar los datos experimentales [Casas 02], mientras mayor es la edad de los alumnos y más avanza su aprendizaje, más simples aparecen las representaciones de las relaciones entre conceptos que se obtienen con las RAP.

Se puede interpretar este hecho considerando que, a pesar de que en la estructura cognitiva del alumno aparecen cada vez más elementos y más relaciones entre ellos, se utilizan subestructuras cada vez más simples y más significativas, a lo que se denomina “senderos de mínimo coste”. Como ocurre en otros aspectos vitales, la estructura cognitiva funciona por un principio de mínima energía.

La conexión más efectiva entre conceptos clave, detalles y procesos de resolución, forman lo que antes hemos denominado un sendero de mínimo coste. La capacidad de seleccionar cuáles son los senderos de mínimo coste es una característica del alumno que resuelve eficazmente los problemas, pero del mismo modo, la capacidad de escoger otros senderos, la flexibilidad para recorrer caminos no usuales, es lo que determina al alumno creativo. Tanto en uno como en otro caso la elección se rige por el mismo principio: la evaluación del coste y de la probabilidad de éxito, determinadas ambas, por las experiencias previas.

2.3. RAP.

Las RAP son una herramientas que se ha utilizado muy poco en educación, como lo expresan [Casas y Luengo 04a]: “*Existen muy pocas investigaciones realizadas con esta técnica en las áreas de matemáticas o ciencias y considera que las posibilidades de las RAP para la investigación son enormes, y que la didáctica de las matemáticas puede enriquecerse con la utilización de esta nueva técnica como se está haciendo en otros campos*”.

Consideran estos autores que las RAP pueden ser incluidas entre los métodos de representación del conocimiento que hacen uso de la puntuación de similaridad entre conceptos. Estos métodos asumen que se puede utilizar una representación espacial entre los conceptos, que describirá el patrón de relaciones entre ellos en la memoria.

La representación se obtiene a partir de una puntuación numérica que se adjudica a la similaridad o diferencia entre los conceptos percibida por un sujeto y que corresponde a su distancia semántica. La distancia semántica pasa a ser considerada como si fuera una distancia geométrica y los conceptos semánticamente más próximos se representarán más próximos en el espacio y análogamente los más distantes [Casas 02].

Las RAP, cuentan con programas de cómputo que permiten establecer las estimaciones de las relaciones conceptuales y la representación gráfica de las mismas en forma de red. El más usado aunque ya desactualizado es KNOT (Knowledges Network Organizing Tool) construido por Schavaneveldt [Shvaneveldt 90]. El grupo de investigación de la Universidad de Extremadura, Ciberdidact, ha desarrollado un nuevo programa llamado Goluca [Casas et al. 11], que mejora las representaciones del programa software KNOT e incorpora nuevos índices que permiten una mejor caracterización de las RAP. También avalan el uso del programa Goluca otros autores como [Almeida et al. 13].

Los programas KNOT y Goluca, pueden calcular varios indicadores, pero los más importantes son los índices de coherencia y la similaridad.

El índice de coherencia (es un número de -1 a 1) mide la consistencia del conjunto de los datos, partiendo de la asunción de que la relación entre un par de elementos (conceptos) puede predecirse por las relaciones de los ítems, con los otros ítems del conjunto, entre menor y mayor consistencia, y viceversa.

El índice de similaridad entre dos redes conceptuales se establece mediante la correspondencia de los vínculos de las dos redes. La similaridad es igual al número de vínculos que tienen en común dividido por el número de vínculos que hay en cualquier red. Dos

redes idénticas tienen un índice de similaridad 1 y dos redes que tienen ninguna similaridad tienen índice 0

El programa Goluca puede calcular también el índice de complejidad [Casas y Luengo 04b]. El índice de complejidad, mide, como dice su nombre, la complejidad de la red, y su valor es un número entre 0 y 1. Cerca de cero significa que es una red simple o menos compleja y cerca de 1 que es una red compleja.

3. Necesidades previas y definición de los parámetros.

Previamente a la definición de los parámetros necesarios para las fórmulas es necesario plasmar el entorno y la información de que disponemos para llevar a cabo el cálculo de los mismos.

3.1. Ubicación de la información.

La información que se necesita para construir la RAP, antes de aplicar los algoritmos necesarios, es lo que llamamos la **matriz de proximidad**. Esta matriz, está compuesta por filas y columnas, y corresponde a los conceptos que se desean analizar y de donde se obtendrán los conceptos nucleares. Cada celda de la matriz contiene un valor numérico que indica la relación de proximidad que tienen los dos conceptos de la fila y columna que indexan esa celda, teniendo en cuenta que un valor alto significa que los conceptos están próximos o están muy relacionados; un valor bajo indica que los dos conceptos están lejanos o poco relacionados; y un valor de 0 indica que no hay relación alguna entre el par de conceptos. Es una matriz simétrica, que tiene los mismos elementos y en el mismo orden, tantos en las filas como en las columnas. Por tanto, a ser una matriz simétrica, el valor para la fila i y la columna j (para un valor de $i < j$) es el mismo que el valor de la fila j y la columna i (Ej. $A(2,4)=25$ y $A(4,2)=25$, siendo A la matriz, 2 y 4 los índices de la matriz, es decir, los valores de la fila y de la columna, y 25 el valor del contenido de la celda). Además, el valor de la fila i y de la columna j (para un valor de $i = j$) no tiene mayor interés (o tendrá un valor 0), ya que indica la relación de un elemento consigo mismo (Ej. $A(2,2)=0$).

Esta matriz la toma como entrada el algoritmo pathfinder para construir la RAP, dando como resultado otra matriz que llamaremos matriz pathfinder. Esta matriz pathfinder, a su vez, sirve también como entrada a un algoritmo de dibujo de redes, como por ejemplo el creado por Kamada y Kawai [Kamada y Kawai 89], para dibujar la red y se puedan observar en ella los enlaces que tiene cada uno de los conceptos (nodos) de la red. En esta red resultante un concepto (un nodo de la red) es nuclear si tiene tres o más enlaces con otros tantos conceptos (nodos).

Para el índice de Nuclearidad, sólo se necesita la matriz de proximidad para obtener la información necesaria para su construcción, ya que en ella se encuentra la información recogida del pensamiento que tienen los sujetos sobre cada par de conceptos, tal cual se les ha presentado en el programa que recoge la información para la toma de datos.

Por tanto, es necesario buscar o construir la matriz de proximidad, es decir, ¿de dónde se extrae la información para rellenar dicha matriz? Esta información, hasta ahora, se obtiene por dos vías distintas:

- Encuestando directamente a los sujetos sobre el grado de relación entre cada par de conceptos. Esta información se puede obtener utilizando el propio programa Goluca¹ citado anteriormente, o bien por el software Meba² construido por el doctor D. Juan Arias para su tesis doctoral [Arias 08] entre otros.
- Obteniendo la información a través de algún software de análisis cualitativo. Por ejemplo, el software WebQDA³ [Souza et al. 11], proporciona directamente la matriz de proximidad una vez que se ha realizado el análisis cualitativo pertinente. La información utilizada, en este artículo, para el cálculo del índice de Nuclearidad está recogida del trabajo realizado en una investigación cualitativa anterior [Contreras et al. 14b], utilizando precisamente este software.

¹ Goluca es un software construido por el grupo de investigación Ciberdidact utilizado para realizar análisis de Redes Asociativas Pathfinder.

² Meba es una aplicación software que entre sus muchas funciones, tiene una que pregunta la proximidad entre conceptos

Es por ello, por lo que nos centraremos en este segundo método para explicar el índice de Nuclearidad, ya que se dispondrá de toda la información necesaria para calcular el mismo.

3.2. Definición de los elementos necesarios.

Partiendo del elemento de donde se va a recoger la información necesaria, y aclarada la forma en que se encuentra la información en esa matriz de proximidad, se partirá de la información contenida en ella para definir los elementos necesarios que se tendrán en cuenta para crear las fórmulas de los índices de Nuclearidad de cada concepto.

A continuación se definen los distintos elementos que son necesarios para construir los parámetros de las fórmulas. Estos elementos son los siguientes:

- Número total de fuentes (se llamará **TF**): Es el número total de documentos, o fuentes de información distintas utilizadas en la investigación y, de las cuales, se van a extraer los conceptos necesarios sobre los que se van a calcular los índices de Nuclearidad.
- Frecuencia (se llamará **F**): Es el número de veces que ha sido referenciado un concepto entre todas las distintas fuentes de información que se ha utilizado en la investigación. Es posible que un concepto sea referenciado en una misma fuente más de una vez.
- Frecuencia máxima (se llamará **F_{MAX}**): es el valor máximo de la frecuencia anterior F.
- Número de conceptos (se llamará **N**): es el número total de conceptos que han salido al realizar la investigación. Coincide con el número de filas o también con el número de columnas de la matriz de proximidad (recordamos que es una matriz simétrica).
- Número de Fuentes por conceptos (se llamará **NF**): es el número de fuentes distintas, de las utilizadas en la investigación, en las que han aparecido cada uno de los conceptos estudiados.

y genera de la matriz de proximidad.

³ WebQDA es un software de apoyo para el análisis de datos cualitativos en un ambiente colaborativo y distribuido. Se accede en la siguiente dirección web www.webqda.com.

- **Número de Vecinos (se llamará NV):** es un número que indica con cuantos conceptos ha establecido relación cada uno de los conceptos, sin contar la relación consigo mismo. Recordemos que en la matriz de proximidad tenemos los valores del grado de relación de cada par de conceptos, por tanto, sumaremos bien por fila o bien por columna, el número de veces que aparece un valor distinto de cero, significando con ello (si existe un valor distinto de cero) que existe una relación entre ese par de conceptos y, por tanto, es lo que se llamará, un vecino.
- **Suma de proximidad (se llamará SP):** es la suma, bien por fila o bien por columnas, de los valores de las relaciones o fortaleza que tiene cada vecino con todos los demás. Es decir, la suma de los valores de las celdas de la matriz de proximidad de cada concepto, bien por filas, o bien por columnas.
- **Suma de proximidad máximo (se llamará SP_{MAX}):** es el valor máximo de entre todas las sumas de proximidad anteriores, SP .

3.3. Generación de los índices auxiliares.

Utilizando los elementos anteriores se va a construir cuatro índices auxiliares que servirán para generar las fórmulas fundamentales. Posteriormente se procederá a normalizar dichos índices auxiliares para que no distorsionen en la fórmula algunos de ellos sobre los demás. Es decir, se va a considerar que cada uno de estos índices auxiliares tiene la misma importancia relativa con respecto a los demás. Para cada uno de estos índices se utilizará seis dígitos decimales para su valor, ya que se obtendrán valores comprendidos entre 0 y 1.

Los índices auxiliares son los siguientes:

- **Índice de Frecuencias (se llamará I_{FREC}):** se corresponde, para cada concepto, al valor de la frecuencia del concepto (F) dividido por el valor de la frecuencia máxima de entre todos los conceptos (F_{MAX}). Se obtendrá un valor decimal entre 0 y 1, indicando los valores próximos a 1, que son conceptos relativamente muy frecuentemente utilizados en relación a los demás conceptos, y un valor cercano al 0, indica que es un concepto que no es relativamente tan frecuentemente utilizado con respecto al resto de

conceptos tratados. Este índice indica el uso o utilización de cada uno de los conceptos de la investigación, es decir, a mayor utilización se entiende que el concepto es más importante. Su cálculo se realizaría:

$$I_{FREC} = F / F_{MAX} \quad (1)$$

- **Índice de Fuentes (se llamará I_{FUEN}):** Es necesario conocer también la ratio de utilización de los conceptos con respecto al número total de documentos diferentes en los que se han utilizado los mismos; no cuantas veces ha sido utilizado, como nos muestra el índice anterior, sino la calidad con la que ha sido utilizado, es decir, en cuantos documentos distintos han sido utilizado cada uno de los conceptos. Se calculará dividiendo el número de fuentes donde ha sido utilizado cada concepto (NF) entre el total de fuentes existentes utilizadas (TF). Se obtendrá un valor decimal entre 0 y 1; indicando un valor alto que un concepto ha sido utilizado en un número importante de documentos, y un valor cercano a 0 indica que el concepto ha sido utilizado en pocos documentos. Su cálculo se realizaría:

$$I_{FUEN} = NF / TF \quad (2)$$

- **Índice de Vecindad (se llamará I_{VEC}):** Este índice se calculará dividiendo el valor del elemento vecino (NV) de cada concepto entre el número total de conceptos (N) menos 1; ya que para este índice no se tiene en cuenta la relación de cada concepto consigo mismo. Es decir, $N-1$ es el número máximo de vecinos que puede tener cada concepto, por ello se divide entre este valor. Se obtendrá un valor decimal entre 0 y 1, indicando un valor próximo a uno que el concepto tiene muchas relaciones con otros conceptos, y un valor próximo a 0 que tiene pocas relaciones con otros conceptos. Su cálculo se realizaría:

$$I_{VEC} = NV / (N-1) \quad (3)$$

- **Índice de Proximidad (se llamará I_{PROX}):** Este índice se calculará dividiendo la suma de proximidad de cada concepto (SP) entre el valor

máximo de la suma de proximidad de todos ellos (SP_{MAX}). Se obtendrá un valor decimal entre 0 y 1. Con este índice se quiere obtener, no sólo con cuantos conceptos está relacionado cada uno de los conceptos, sino la fuerza de esta relación. Es decir, un valor próximo a 1 indica que la fuerza de la relación entre los conceptos es muy fuerte, y un valor próximo a 0 indica que la relación entre conceptos es débil. Su cálculo se realizaría:

$$I_{PROX} = SP / SP_{MAX} \quad (4)$$

3.4. Normalización de la importancia de los índices auxiliares.

Todos los índices auxiliares van a tener la misma importancia. Para que alguno de ellos no distorsione el resultado final de los Índices de Nuclearidad se van a normalizar sus valores.

Por cada uno de los índices auxiliares anteriores se va a calcular el valor de la diferencia entre el valor máximo del índice y el valor mínimo del índice, y a este valor se llamará **Delta** para cada índice. Posteriormente se realizarán dos sumas, por un lado se sumarán todos los deltas anteriores y dará un valor que se llamará **Delta total**, y por otro lado, se sumarán sólo los deltas de los índices de vecinos y suma de proximidad y dará un valor que se llamará **Delta parcial**. Estos dos deltas se podrán utilizar posteriormente para crear una fórmula para el índice Nuclearidad Completa y otra fórmula para el índice de Nuclearidad Reducida.

Con todos los valores de los Deltas se calcularán los Pesos. Los Pesos indican el significado de cada Delta de cada índice auxiliar con respecto al Delta total o al Delta parcial. Se finalizará, normalizándolos con respecto a 100, todos los valores de los índices anteriores calculados.

Mostramos el cálculo de los pasos explicados anteriormente:

- a) *Cálculo de los Delta para cada índice:* $IFrec_{max}$ es el valor máximo obtenido en la aplicación de la formula (1) e $IFrec_{min}$ es el valor mínimo obtenido de la aplicación de la formula (1). $IFuen_{max}$ es el valor máximo obtenido en la aplicación de la

formula (2) e $IFuen_{min}$ es el valor mínimo obtenido de la aplicación de la formula (2). $Ivec_{max}$ es el valor máximo obtenido en la aplicación de la formula (3) e $Ivec_{min}$ es el valor mínimo obtenido de la aplicación de la formula (3). $Iprox_{max}$ es el valor máximo obtenido en la aplicación de la formula (4) e $Iprox_{min}$ es el valor mínimo obtenido de la aplicación de la formula (4). Por tanto, el valor de cada Delta es:

$$D_{IFrec} = IFrec_{max} - IFrec_{min}$$

$$D_{IFuen} = IFuen_{max} - IFuen_{min}$$

$$D_{Ivec} = Ivec_{max} - Ivec_{min}$$

$$D_{Iprox} = Iprox_{max} - Iprox_{min}$$

- b) *Cálculo de los Deltas total y parcial:* como se ha indicado anteriormente, se realizarán dos cálculos distintos. El primero correspondiente a la suma de los cuatro Deltas anteriores, llamado D_{total} , y el segundo correspondiente sólo a los Deltas del índice de vecinos y del índice de proximidad, llamado $D_{parcial}$. Por tanto, el valor de cada Delta es:

$$D_{total} = D_{IFrec} + D_{IFuen} + D_{Ivec} + D_{Iprox}$$

$$D_{parcial} = D_{Ivec} + D_{Iprox}$$

- c) *Cálculo de los Pesos, en porcentajes, con el D_{total} , para cada uno de los índices:* Con valores de los Delta de cada índice y del D_{total} se calcularán los pesos en % de cada uno de los índices. Hay tener en cuenta, que se va a tener dos tipos de pesos, y los pesos, en este caso, se utilizarán para la fórmula Completa. El cálculo se realiza de la siguiente forma:

$$P_{IFrec} = D_{IFrec} * 100 / D_{total} \quad (5)$$

$$P_{IFuen} = D_{IFuen} * 100 / D_{total} \quad (6)$$

$$P_{Ivec} = D_{Ivec} * 100 / D_{total} \quad (7)$$

$$P_{Iprox} = D_{Iprox} * 100 / D_{total} \quad (8)$$

- d) *Cálculo de los Pesos, en porcentajes, con el $D_{parcial}$, para cada uno de los índices:* Con los valores de los Delta de los índices de vecinos y proximidad y del valor del $D_{parcial}$ se calcularán los

pesos en % de cada uno de los índices. Hay tener en cuenta, que se va a tener dos tipos de pesos, y los pesos, en este caso, se utilizarán para la fórmula Reducida. El cálculo se realiza de la siguiente forma:

$$P_{IVec} = D_{IVec} * 100 / D_{parcial} \quad (9)$$

$$P_{IProx} = D_{IProx} * 100 / D_{parcial} \quad (10)$$

4. Fórmulas finales Completa y Reducida.

Se van a definir las fórmulas de los Índices de Nuclearidad en sus dos modalidades: en formato completo y en formato reducido.

El formato completo constituye la forma natural de realizar el cálculo del índice y en él se incluyen todos los elementos necesarios e influyentes para este índice. Sería la forma ideal de discernir la importancia de cada uno de los conceptos en relación con todos los elementos que intervienen en las investigaciones en la que se encuadra lo que estamos buscando. Para ello, se definirá el Índice de Nuclearidad Completo (se llamará IN_C).

Entendemos que disponer de toda la información que interviene en el cálculo del Índice de Nuclearidad completo no siempre es posible, por lo que se va a definir también, para estos casos, el Índice de Nuclearidad Reducido (se llamará IN_R). Consideramos que este último índice contempla dos de los elementos fundamentales para diferenciar la importancia entre conceptos, y que son: cuantos vecinos tienen cada uno de los conceptos y como de fuerte es la relación entre ellos.

Ambos permiten graduar la importancia de los conceptos nucleares, estableciendo una relación orden de importancia entre ellos, donde el índice completo tiene más información que el índice reducido.

4.1. Fórmula del Índice de Nuclearidad Completo.

Para el cálculo del Índice de Nuclearidad Completo van a intervenir los resultados de las fórmulas mostradas anteriormente con números: (1), (2), (3),

(4), (5), (6), (7) y (8), aplicando cada uno de estos elementos de la siguiente forma:

$$IN_C = (I_{Frec} * P_{IFrec} + I_{Fuen} * P_{IFuen} + I_{Vec} * P_{IVec} + I_{Prox} * P_{IProx}) / 100 \quad (11)$$

4.2. Fórmula del Índice de Nuclearidad Reducido.

Para el cálculo del Índice de Nuclearidad Reducido van a intervenir los resultados de las formulas mostradas anteriormente con números: (3), (4), (9) y (10), aplicando cada uno de estos elementos de la siguiente forma:

$$IN_R = (I_{Vec} * P_{IVec} + I_{Prox} * P_{IProx}) / 100 \quad (12)$$

5. Aplicación práctica.

5.1. Centrando el problema.

Se quiere llevar a cabo una investigación sobre la asignatura de Bases de Datos en los Grados de Ingeniería Informática, de diversas universidades españolas, donde se obtengan cuáles son los conceptos más importantes, comunes a todas ellas, que deben estudiar los alumnos en el tema de SQL (Structured Query Languages). Es un caso real de una investigación que podemos encontrar en [Contreras 16].

Se han seleccionado como fuentes de información las memorias de verificación de 22 universidades y, de cada una de ellas, se han extraído, en formato texto, las partes que hablan sobre Bases de Datos y que hace referencia a SQL, generando 22 fuentes documentales. Sobre estas fuentes se ha realizado un análisis cualitativo, utilizando el software WebQDA, unificando y generando las categorías conceptuales comunes en todas las fuentes documentales (22 textos). Por lo que, toda la información que necesitamos como: la matriz de proximidad, los datos de los distintos parámetros, etc., se obtendrá a través de este software.

5.2. Cálculo de los parámetros básicos.

El número de memorias de verificación utilizadas, como se puede observar en la siguiente **Figura 3**, han sido un total de 22, por lo que el parámetro llamado Total de Fuentes se corresponde con este valor, es decir, **TF=22**.

Nombre	Nodos	Ref.
es_andalucia_almeria	11	11
es_aragon_zaragoza	10	10
es_asturias_oviedo	14	14
es_baleares_islasbalears	8	8
es_canarias_jaspalmas	14	14
es_cantabria_cantabria	19	21
es_castillalamancha_castillalamancha	27	31
es_castillayleon_burgos	7	7
es_cataluña_barcelona	12	12
es_extremadura_merida	10	10
es_galicia_vigo	24	25
es_madrid_politecnica	12	14
es_nacional_uned	28	29
es_navarra_publicadenavarra	37	41
es_paisvasco_guipuzcoa	22	23
es_valencia_valencia	60	65

Figura 3. Ejemplo de fuentes utilizadas.

Los conceptos encontrados en la investigación para el grupo llamado SQL, como se puede observar en la siguiente **Figura 4**, en la columna de nombre, han sido un total de 12, por lo que el parámetro llamado Número de Conceptos se corresponde con este valor, es decir, **N=12**.

Nombre	Referencias
SQL Structured Query Languages	0
Almacenamiento de datos y consultas	1
Asertos y disparadores	1
Consultas básicas, anidadas y complejas	3
Creación y manipulación de bases de datos relacionales	1
Definición de datos, restricciones, triggers y cambios de esquema en SQL	1
Introducción a SQL	2
Lenguajes LDD, LCD y LMD	16
Modificación de la base de datos	1
Operaciones de agregación	1
Procesamiento de interrogaciones	2
Sentencias INSERT, DELETE y UPDATE	1
Vistas	3

Figura 4. Conceptos obtenidos de la investigación.

La frecuencia de cada concepto se puede observar también en la **Figura 4**, en la columna de Referencias. Por ejemplo, para el concepto *Lenguajes LDD, LCD y LMD*, la frecuencia es **F=16**, para el concepto *Vistas* la frecuencia es **F=3**, y así sucesivamente se obtendrá para cada concepto.

La frecuencia máxima de todos estos conceptos la representa el concepto llamado *Lenguajes LDD, LCD y LMD* con un valor de 16, por lo que el parámetro **F_{max}=16**.

El siguiente parámetro necesario es el número de fuentes por concepto. Este parámetro se extraerá también desde el software WebQDA utilizando los valores de la diagonal principal de la matriz de proximidad que nos proporciona un tipo de consulta, existente dentro del mismo software, llamada *Consulta de Matrices Triangulares*. Se puede observar estos valores en la siguiente **Figura 5**. Por ejemplo, para el concepto *Lenguajes LDD, LCD y LMD* el valor del número de fuentes es 11, **NF=11**; para el concepto *Vistas* el valor del número de fuentes es 3, es decir, **NF=3**. Y así se obtendrá este valor para cada uno de los conceptos de la investigación.

Matriz Triangular	Almacenamiento de datos y consultas	Asertos y disparadores	Consultas básicas, anidadas y complejas	Creación y manipulación de bases de datos relacionales	Definición de datos, restricciones, triggers y cambios de esquema en SQL	Introducción a SQL	Lenguajes LDD, LCD y LMD	Modificación de la base de datos	Operaciones de agregación	Procesamiento de interrogaciones	Sentencias INSERT, DELETE y UPDATE	Vistas
Almacenamiento de datos y consultas	1											
Asertos y disparadores	0	1										
Consultas básicas, anidadas y complejas	0	0	1									
Creación y manipulación de bases de datos relacionales	0	0	0	1								
Definición de datos, restricciones, triggers y cambios de esquema en SQL	0	0	1	0	1							
Introducción a SQL	0	0	0	0	0	2						
Lenguajes LDD, LCD y LMD	1	1	1	0	1	2	11					
Modificación de la base de datos	0	1	0	0	0	0	1	1				
Operaciones de agregación	0	0	1	0	1	0	1	0	1			
Procesamiento de interrogaciones	0	0	0	0	0	0	1	0	0	2		
Sentencias INSERT, DELETE y UPDATE	0	0	1	0	1	0	1	0	1	0	1	
Vistas	0	1	1	0	1	1	2	1	1	0	1	3

Figura 5. Matriz de proximidad y número de fuentes por cada concepto.

El conjunto de todos los valores del ejemplo para NF se puede observar en la siguiente **Figura 6**.

Nombre del Concepto	Número de Fuentes (NF)
Almacenamiento de datos y consultas	1
Asertos y disparadores	1
Consultas básicas, anidadas y complejas	1
Creación y manipulación de bases de datos relacionales	1
Definición de datos, restricciones, triggers y cambios de esquema en SQL	1
Introducción a SQL	2
Lenguajes LDD, LCD y LMD	11
Modificación de la base de datos	1
Operaciones de agregación	1
Procesamiento de interrogaciones	2
Sentencias INSERT, DELETE y UPDATE	1
Vistas	3

Figura 6. Número de Fuentes por concepto.

Los siguientes parámetros que se necesitan son: el número de vecinos por cada concepto (NV), la suma de proximidad por cada concepto (SP) y la suma de proximidad máxima (SP_{max}). Los dos primeros se obtendrán de la matriz de proximidad de la **Figura 5**. Se transformará la misma para calcular los valores, de tal forma que los valores de la diagonal principal sean ceros y se traspondrán cada una de las filas en cada columna. El resultado de dicha matriz se puede observar en la siguiente **Figura 7**.

Matriz Traspuesta	Almacenamiento de datos y consultas	Asertos y disparadores	Consultas básicas, anidadas y complejas	Creación y manipulación de bases de datos relacionales	Definición de datos, restricciones, triggers y cambios de esquema en SQL	Introducción a SQL	Lenguajes LDD, LCD y LMD	Modificación de la base de datos	Operaciones de agregación	Procesamiento de interrogaciones	Sentencias INSERT, DELETE y UPDATE	Vistas
Almacenamiento de datos y consultas	0	0	0	0	0	0	1	0	0	0	0	0
Asertos y disparadores	0	0	0	0	0	0	1	1	0	0	0	1
Consultas básicas, anidadas y complejas	0	0	0	0	1	0	1	0	1	0	1	1
Creación y manipulación de bases de datos relacionales	0	0	0	0	0	0	0	0	0	0	0	0
Definición de datos, restricciones, triggers y cambios de esquema en SQL	0	0	1	0	0	0	1	0	1	0	1	1
Introducción a SQL	0	0	0	0	0	2	0	0	0	0	0	1
Lenguajes LDD, LCD y LMD	1	1	1	0	1	2	0	1	1	1	1	3
Modificación de la base de datos	0	1	0	0	0	1	0	0	0	0	0	1
Operaciones de agregación	0	0	1	0	1	0	1	0	0	0	1	1
Procesamiento de interrogaciones	0	0	0	0	0	0	1	0	0	0	0	0
Sentencias INSERT, DELETE y UPDATE	0	0	1	0	1	0	1	0	1	0	0	1
Vistas	0	1	1	0	1	1	3	1	1	0	1	0

Figura 7. Matriz de proximidad completa y traspuesta.

Para el cálculo del parámetro de vecinos (NV), para cada concepto (para cada fila), se sumará por fila los valores distintos de cero de cada columna de la matriz de la **Figura 7**. Por ejemplo, para el concepto *Asertos y Disparadores* (fila 2), las columnas: *Lenguajes LDD, LCD y LMD*, *Modificación de la Base de Datos* y *Vistas* tienen valor 1, es decir, estas tres columnas son las únicas con valores distintos de cero, por lo que el valor de NV para este concepto en esa fila es 3, $NV=3$.

El resultado del cálculo de los vecinos por conceptos (NV) se muestra en la siguiente **Figura 8**.

Nombre del Concepto	Num Vecinos (NV)
Almacenamiento de datos y consultas	1
Asertos y disparadores	3
Consultas básicas, anidadas y complejas	5
Creación y manipulación de bases de datos relacionales	0
Definición de datos, restricciones, triggers y cambios de esquema en SQL	5
Introducción a SQL	2
Lenguajes LDD, LCD y LMD	10
Modificación de la base de datos	3
Operaciones de agregación	5
Procesamiento de interrogaciones	1
Sentencias INSERT, DELETE y UPDATE	5
Vistas	8

Figura 8. Número de vecinos por concepto.

Continuando con la **Figura 7**, a continuación se va a obtener el valor de la suma de proximidad (SP) por cada concepto (por cada fila), sumando los valores de cada columna de esa fila y dando un valor total que corresponderá a este SP que se está buscando. Siguiendo el mismo ejemplo anterior, para el concepto *Introducción a SQL* (fila 6), la columna *Lenguajes LDD, LCD y LMD* tienen valor 2 y la columna *Vistas* tiene valor 1, y el resto de columnas tiene valor de 0, es decir, la suma de todos estos valores de SP para ese concepto es 3, $SP=3$.

Se puede observar el resultado del cálculo de las sumas de proximidades por conceptos en la siguiente **Figura 9**.

Nombre del Concepto	Suma Proximidad (SP)
Almacenamiento de datos y consultas	1
Asertos y disparadores	3
Consultas básicas, anidadas y complejas	5
Creación y manipulación de bases de datos relacionales	0
Definición de datos, restricciones, triggers y cambios de esquema en SQL	5
Introducción a SQL	3
Lenguajes LDD, LCD y LMD	13
Modificación de la base de datos	3
Operaciones de agregación	5
Procesamiento de interrogaciones	1
Sentencias INSERT, DELETE y UPDATE	5
Vistas	10

Figura 9. Suma de proximidad por concepto.

Por último, el parámetro que queda por calcular es el valor máximo de la suma de proximidad (SP_{max}). Este valor se puede observar en la tabla de la **Figura 9**, y

corresponde al concepto *Lenguajes LDD, LCD y LMD* con un valor de 13, es decir, $SP_{max}=13$.

5.3. Cálculo de los índices auxiliares.

Para calcular el índice de frecuencias (I_{FREC}), el índice de fuentes (I_{FUEN}), el índice de vecinos (I_{VEC}) y el índice de proximidad (I_{PROX}), se aplicarán las fórmulas anteriores numeradas: (1), (2), (3) y (4) respectivamente para cada uno de los 12 conceptos que existen.

Teniendo en cuenta que, para el ejemplo práctico, la Frecuencia Máxima (F_{max}) es 16, la Suma de Proximidad máxima (SP_{max}) es 13, el Número Total de Fuentes (TF) es 22 y el Número Total de Conceptos (N) es 12, el resultado de aplicar las fórmulas anteriores se puede observar en la siguiente **Figura 10**.

Frecuencia Máxima (F_{max})	16							
Suma de Proximidad Máxima (SP_{max})	13							
Número total de Conceptos (N)	12							
Número total de Fuentes (TF)	22							
		$I_{FREC}=F/F_{max}$	$I_{FUEN}=NF/TF$	$I_{VEC}=NV/N$	$I_{PROX}=SP/SP_{max}$			
Nombre del Concepto	F	I_{FREC}	I_{FUEN}	I_{VEC}	I_{PROX}			
Almacenamiento de datos y consultas	1	0,06250	1	0,04545	1	0,09091	1	0,07692
Asertos y disparadores	1	0,06250	1	0,04545	3	0,27273	3	0,23077
Consultas básicas, avanzadas y complejas	3	0,18750	1	0,04545	5	0,45455	5	0,38462
Creación y manipulación de bases de datos relacionales	1	0,06250	1	0,04545	0	0,00000	0	0,00000
Definición de datos, restricciones, triggers y cambios de esquema en SQL	1	0,06250	1	0,04545	5	0,45455	5	0,38462
Introducción a SQL	2	0,12500	2	0,09091	2	0,18182	3	0,23077
Lenguajes LDD, LCD y LMD	16	1,00000	11	0,50000	10	0,90909	13	1,00000
Modificación de la base de datos	1	0,06250	1	0,04545	3	0,27273	3	0,23077
Operaciones de agregación	1	0,06250	1	0,04545	5	0,45455	5	0,38462
Procesamiento de interacciones	2	0,12500	2	0,09091	1	0,09091	1	0,07692
Sentencias INSERT, DELETE y UPDATE	1	0,06250	1	0,04545	5	0,45455	5	0,38462
Vistas	3	0,18750	3	0,13636	0	0,27273	10	0,76923

Figura 10. Valor de los índices de cada concepto.

Y como se observa en la **Figura 10**, los valores de los índices auxiliares se encuentran en el rango entre 0 y 1.

5.4. Cálculo de los Deltas y de los Pesos.

Para el cálculo de los Deltas de cada índice se utilizarán las siguientes formulaciones:

$$D_{IFREC} = IFREC_{max} - IFREC_{min}$$

$$D_{IFUEN} = IFUEN_{max} - IFUEN_{min}$$

$$D_{Ivec} = IVec_{max} - IVec_{min}$$

$$D_{IPROX} = IPROX_{max} - IPROX_{min}$$

Teniendo en cuenta que los valores máximos y mínimos de los índices auxiliares son los que mostramos a continuación, el valor de cada Delta es el siguiente:

$$D_{IFREC} = 1,00000 - 0,06250 = 0,93750$$

$$D_{IFUEN} = 0,50000 - 0,04545 = 0,45455$$

$$D_{Ivec} = 0,90909 - 0,00000 = 0,90909$$

$$D_{IPROX} = 1,00000 - 0,00000 = 1,00000$$

Con todos ellos se puede calcular el Delta total y el Delta parcial aplicando la siguiente fórmula:

$$D_{total} = D_{IFREC} + D_{IFUEN} + D_{Ivec} + D_{IPROX}$$

$$D_{parcial} = D_{Ivec} + D_{IPROX}$$

Por lo que, el resultado de cada uno de ellos es el siguiente:

$$D_{total} = 0,93750 + 0,45455 + 0,90909 + 1,00000 = 3,30114$$

$$D_{parcial} = 0,90909 + 1,00000 = 1,90909$$

A continuación se calculan los pesos de cada uno de los índices auxiliares, para el Índice de Nuclearidad Completo, utilizando las fórmulas anteriores numeradas con (5), (6), (7) y (8) como se observa a continuación.

$$P_{IFREC} = D_{IFREC} * 100 / D_{total} \quad (5)$$

$$P_{IFUEN} = D_{IFUEN} * 100 / D_{total} \quad (6)$$

$$P_{Ivec} = D_{Ivec} * 100 / D_{total} \quad (7)$$

$$P_{IPROX} = D_{IPROX} * 100 / D_{total} \quad (8)$$

El resultado de aplicar estas fórmulas anteriores es el siguiente:

$$P_{IFREC} = 0,93750 * 100 / 3,30114 = 28,39931 \quad (5)$$

$$P_{IFUEN} = 0,45455 * 100 / 3,30114 = 13,76936 \quad (6)$$

$$P_{Ivec} = 0,90909 * 100 / 3,30114 = 27,53873 \quad (7)$$

$$P_{IPROX} = 1,00000 * 100 / 3,30114 = 30,29260 \quad (8)$$

De la misma forma que se ha calculado los pesos anteriores para el índice de Nuclearidad Completo, se calcularán los pesos para el índice de Nuclearidad

Reducido. Y para ello, se utilizarán las fórmulas numeradas con el (9) y el (10) como se observa a continuación.

$$P_{IVec} = D_{IVec} * 100 / D_{parcial} \quad (9)$$

$$P_{IProx} = D_{IProx} * 100 / D_{parcial} \quad (10)$$

El resultado de aplicar las fórmulas anteriores es el siguiente:

$$P_{IVec} = 0,90909 * 100 / 1,90909 = 47,61902 \quad (9)$$

$$P_{IProx} = 1,00000 * 100 / 1,90909 = 52,38098 \quad (10)$$

5.5. Cálculo del Índice de Nuclearidad Completo.

Una vez calculados los valores de los índices auxiliares, deltas y pesos, de cada una de las fórmulas, se puede realizar el cálculo del Índice de Nuclearidad Completo para cada concepto.

Aplicando la fórmula (11), que se muestra a continuación, para cada concepto, se obtendrá el Índice de Nuclearidad Completo como se observa en la columna IN_C de la **Figura 11**.

$$IN_C = (I_{Frec} * P_{IFrec} + I_{Fuen} * P_{IFuen} + I_{Vec} * P_{IVec} + I_{Prox} * P_{IProx}) / 100 \quad (11)$$

Nombre del Concepto	IN_C Índice de Nuclearidad Completo	IN_R Índice de Nuclearidad Reducido
Almacenamiento de datos y consultas	0,07235	0,00358
Asertos y disparadores	0,16302	0,25075
Consultas básicas, anidadas y complejas	0,30119	0,41792
Creación y manipulación de bases de datos relacionales	0,02401	0,00000
Definición de datos, restricciones, triggers y cambios de esquema en SQL	0,26569	0,41792
Introducción a SQL	0,16799	0,20745
Lenguajes LDD, LCO y LMD	0,90612	0,95671
Modificación de la base de datos	0,16302	0,25075
Operaciones de agregación	0,26569	0,41792
Procesamiento de interrogaciones	0,09535	0,00358
Sentencias INSERT, DELETE y UPDATE	0,26569	0,41792
Vistas	0,50533	0,74925

Figura 11. Tabla de Índices de Nuclearidad.

5.6. Cálculo del Índice de Nuclearidad Reducido.

Una vez calculados los valores de los índices auxiliares, deltas y pesos, de cada una de las fórmulas anteriores, se puede realizar el cálculo el Índice de Nuclearidad Reducido para cada concepto.

Aplicando la fórmula (12), que se muestra a continuación, para cada concepto, se obtendrá el Índice de Nuclearidad Reducido como se observa en la columna IN_R de la **Figura 11**.

$$IN_R = (I_{Vec} * P_{IVec} + I_{Prox} * P_{IProx}) / 100 \quad (12)$$

5.7. Conceptos Nucleares finales obtenidos.

Recordemos, que los creadores de la TCN, consideran un concepto como nuclear cuando, una vez construida la RAP, el número de conceptos con los que se relaciona cada concepto es 3 o mayor que 3, es decir, tiene tres o más enlaces o, como le hemos llamado en el ejemplo, vecinos.

Para el ejemplo que nos ocupa, el resultado de Nuclearidad de los conceptos teniendo como referencia el Índice de Nuclearidad Completo, es decir, los conceptos más importantes en función de número de vecinos y el valor de ese índice, son los que se observan en la siguiente **Figura 12**.

Nombre del Concepto	Num Vecinos	Índice de Nuclearidad Completo
Lenguajes LDD, LCO y LMD	10	0,90612
Vistas	8	0,50533
Consultas básicas, anidadas y complejas	5	0,30119
Definición de datos, restricciones, triggers y cambios de esquema en SQL	5	0,26569
Operaciones de agregación	5	0,26569
Sentencias INSERT, DELETE y UPDATE	5	0,26569
Asertos y disparadores	3	0,16302
Modificación de la base de datos	3	0,16302
Introducción a SQL	2	0,16799
Procesamiento de interrogaciones	1	0,09535
Almacenamiento de datos y consultas	1	0,07235
Creación y manipulación de bases de datos relacionales	0	0,02401

Figura 12. Número de enlaces e Índice de Nuclearidad Completo.

Por lo que, los conceptos más importantes, en orden de importancia, comunes a todas las memorias de verificación, de las universidades estudiadas, para el tema de SQL en la asignatura de Bases de Datos, son los siguientes:

- Lenguajes LDD, LCD y LMD.
- Vistas.
- Consultas básicas, anidadas y complejas.
- Definición de datos, restricciones, triggers y cambios de esquema en SQL.
- Operaciones de agregación.
- Sentencias INSERT, DELETE y UPDATE.
- Asertos y disparadores.
- Modificación de la base de datos.

Las RAP obtenida desde el software Goluca, como se observa en la **Figura 13**, para el grupo de conceptos de SQL, proporciona también, aunque teniendo en cuenta sólo el número de enlaces, los conceptos nucleares.



Figura 13. Imagen Gráfica de Conceptos Nucleares en Goluca.

Como se muestra en la **Figura 13**, con colores amarillo o naranja, los conceptos que son considerados como nucleares en el software Goluca, siguiendo a los autores de la TCN, representados en una RAP, ya que

tienen 3 o más enlaces. Los conceptos observados en ella son:

- Asertos y disparadores
- Consultas básicas, anidadas y complejas
- Creación y manipulación de bases de datos relacionales
- Definición de datos, restricciones, triggers y cambios de esquema en SQL
- Lenguajes LDD, LCD y LMD
- Modificación de la base de datos
- Operaciones de agregación
- Sentencias INSERT, DELETE y UPDATE
- Vistas

Existe una coincidencia de los Conceptos Nucleares obtenidos por ambos sistemas en el que sólo difieren en el concepto Creación y manipulación de bases de datos relacionales, ya que este es considerado como nuclear sólo por las RAP implementadas en el software Goluca. Consideramos que es más preciso el Índice de Nuclearidad ya que este concepto según la matriz de proximidad no tiene ninguna relación con el resto de conceptos, por lo que no debería ser considerado tampoco como nuclear en Goluca. Advertimos de esta circunstancia a los constructores del mismo para que estudie este caso.

6. Conclusiones.

Teniendo en cuenta las necesidades iniciales con respecto a la distinción de la importancia entre los conceptos de forma cuantitativa, con la creación del Índice de Nuclearidad, se ha conseguido tal diferenciación. Es cierto también, que los valores en la matriz de proximidad son bajos, del orden de valores entre 1 y 3, lo que hace complicado tal diferenciación de los índices de Nuclearidad. Por ello aconsejamos que los valores en la graduación de los conceptos que intervienen en la matriz de proximidad, al menos, tengan valores entre 0 y 100, aunque es preferible que sean entre 0 y 1000, lo que mejoraría los índices de Nuclearidad. Aun teniendo en cuenta esto, el IN_C funciona mejor que solo el conteo de los nodos vecinos para establecer los conceptos nucleares.

El Índice de Nuclearidad Reducido (IN_R) sólo debe utilizarse cuando no sea posible tener toda la información necesaria de los parámetros que intervienen en el IN_C . Por lo que sólo es un índice auxiliar, que mejora el sistema de conteo de los nodos vecinos, pero que en algunos casos puede no distinguir tan correctamente como lo hace el IN_C .

Con la creación de estos dos índices, se ha conseguido mejorar los contenidos que pueden implementarse en los distintos programas software de análisis cualitativo, como por ejemplo, WebQDA o NVivo. Por lo que, sería interesante que los mismos incorporaran estos índices e implementaran entre sus mejoras esta formulación para que puedan ser utilizados por los investigadores que usen dichos software de una forma directa sin tener que realizar ningún cálculo adicional por ellos mismos. En definitiva, facilitar a los investigadores el uso de estos dos índices

Aunque actualmente el Índice de Nuclearidad no está implementado todavía en alguno de los programas software antes mencionados, estamos en contacto con los constructores de alguno de ellos (Ej. Goluca y WebQDA) para que en un futuro puedan ser incorporados como mejoras a los mismos.

Como en el ejemplo práctico anterior en Ingeniería, en el que ha sido utilizado el índice de Nuclearidad para conseguir obtener los conceptos importantes en SQL de la asignatura de Bases de Datos en los Grados de Informática de las universidades españolas seleccionadas, también puede ser utilizado en otras áreas de conocimiento como entre las que podemos destacar: en educación, las de didáctica de las ciencias experimentales, didáctica de las matemáticas, didácticas de las ciencias sociales; en las áreas de ingenierías como en Informática y Telemática; etc.

Una aplicación práctica del uso del índice de Nuclearidad en educación podría ser la comprobación posterior, por parte del profesor, si lo explicado en clase se corresponde con lo asimilado por el alumno o grupo de alumnos, detectando lo que verdaderamente importante ha sido interiorizado por los alumnos, permitiendo con ello, actuar o modificar, de forma inmediata, la práctica docente para conseguir los objetivos.

Referencias

- [Almeida et al. 13] Almeida, C., Casas, L. y Luengo, R., «A organização conceptual dos estudantes, dada através das Redes Associativas Pathfinder, do conceito de Probabilidade: Um estudo com alunos do 9.º ano de Escolaridade do Ensino Básico Português,» de *1 Jornadas Virtuales en Didáctica de la Estadística, Probabilidad y Combinatoria*, Granada, 2013.
- [Arias 08] Arias, J., «Evaluación de la calidad de Cursos Virtuales: Indicadores de Calidad y construcción de un cuestionario de medida. Aplicación al ámbito de asignaturas de Ingeniería Telemática,» Universidad de Extremadura, Badajoz, 2008.
- [Ausubel et al. 83] Ausubel, D.P.; Novak, J.D. y Hanesian H., «Psicología educativa: un punto de vista cognoscitivo,» Mexico, Ed. Trillas. México (trd versión en inglés del 1978), 1983.
- [Casas y Luengo 05] Casas, L.M. y Luengo, R., «Conceptos nucleares en la construcción del concepto de ángulo,» *Enseñanza de las Ciencias*, nº 23, pp. 201-206, 2005.
- [Casas y Luengo 03] Casas, L.M. y Luengo, R., «Matemáticas: Representação Da estrutura Cognitiva de Alunos,» Evora, 2003.
- [Casas y Luengo 04b] Casas, L.M. y Luengo, R., «Representación del conocimiento y aprendizaje. Teoría de los conceptos nucleares.,» *Revista Española de pedagogía*, n. 227, pag. 59-84, 2004b.
- [Casas y Luengo 04a] Casas, L.M. y Luengo, R., «Teoría de los Conceptos Nucleares. Aplicación en Didáctica de las Matemáticas,» Badajoz, R. Luengo, líneas de investigación en Educación Matemática. Servicio de publicaciones FESPM, 2004a.
- [Casas y Luengo 13] Casas, L.M. y Luengo, R., «The study of the pupil's cognitive structure: the concept of angle,» *Journal Psychology Education (DOI 10.1007/s10212-012-0119-4)*, nº 28, pp. 373-398, 2013.

- [Casas 02] Casas, L.M., «El estudio de la estructura cognitiva de los alumnos a través de redes asociativas Pathfinder. Aplicación y posibilidades en Geometría.,» Universidad de Extremadura, 2002.
- [Casas et al. 11] Casas, L.M.; Luengo, R. y Godinho, V., «Software GOLUCA: Knowledge Representation in Mental Calculation,» *US-China Education Review B* 4, pp. 592-600, 2011.
- [Contreras 16] Contreras, J.A., «Enseñanza por Competencias: Conceptos propios, requisitos previos e influencia en el rendimiento académico de los alumnos, para la asignatura de Bases de Datos, en los estudios universitarios de Grado en Informática,» Universidad de Extremadura, Badajoz, 2016.
- [Contreras et al. 14a] Contreras, J.A.; Luengo, R.; Arias, J. y Casas, L.M., «Análisis cualitativo y cuantitativo de las materias básicas de base de datos en las memorias de verificación de los títulos universitarios de Grado en Informática en las Universidades Españolas,» *RISTI (ISSN: 1646-9895)*, vol. E2, nº 09-2014, pp. 37-53, 2014.
- [Contreras et al. 14b] Contreras, J.A.; Luengo, R.; Arias, J. y Casas, L.M., «Análisis mixto, cualitativo-cuantitativo, de los contenidos básicos de las materias de base de datos en los Planes de Estudio Universitarios de Grado en Informática en la Universidad de Extremadura.,» de *3º Congreso Ibérico en Investigación Cualitativa*, Badajoz (España), 2014.
- [Fenker 75] Fenker, R.M., «The organization of conceptual materials: A methodology for measuring ideal and actual cognitive structures,» *Instructional Science*, vol. 4, pp. 33-57, 1975.
- [Jonassen 90] Jonassen, D.H., «Semantic network elicitation: tools for structuring hypertext,» *Hypertext: State of the Art. Oxford: Intellect*, 1990.
- [Kamada y Kawai 89] Kamada, T. and Kawai, S., «An algorithm for drawing general undirected graphs,» *Information Processing Letters*, vol. 31, pp. 7-15, 1989.
- [Luengo 13] Luengo, R., «La teoría de los conceptos nucleares y su aplicación de la investigación en la didáctica de las Matemáticas,» *Revista Iberoamericana de Educación Matemática*, nº 34, pp. 9-36, 2013.
- [Luengo et al. 11] Luengo, R.; Casas, L.M.; Mendoza, M. y Arias, J., «Possibilities of "Nuclear Concepts Theory" on Educational Research, a Review,» *International Conference The Future of Education*, 2011.
- [Neisser 76] Neisser, U., *Psicología Cognoscitiva*, México: Ed. Trillas trd. original inglés de 1969, 1976.
- [Piaget 78] Piaget, J., «La evolución intelectual entre la adolescencia y la edad adulta,» J Delval comp. *Lecturas de Psicología del niño*, V 2, 1978.
- [Preece 76] Preece, P., «Mapping cognitive structure: A comparison of methods,» *Journal of Educational Psychology*, vol. 68, pp. 1-8, 1976.
- [Rumelhart 80] Rumelhart, D.E., «Schemata: The building block of cognition,» *R.J. Spiro, B.C. Bruce y W. Brewer. Theoretical issues in reading comprehension*, 1980.
- [Rumelhart 84] Rumelhart, D.E., *Schemata and the cognitive system*, Hillsdale, NJ: Erlbaum: R-S-Wyer y T.K. Srull eds, *Handbook of social cognition*, v.1, 1984.
- [Schvaneveldt 90] Schvaneveldt, R.W., «Pathfinder Associative Networks: Studies in Knowledge Organization,» ABBLEX Publishing Corporation, Norwood, 1990.
- [Shavelson 72] Shavelson, R., «Some aspects of the correspondence between content structure and cognitive structure in physics instruction,» *Journal of Educational Psychology*, vol. 63, pp. 225-234, 1972.
- [Souza et al. 11] Souza, F.N.; Costa, A.P. & Moreira, A., «Análise de Dados Qualitativos Suportada pelo Software WebQDA,» de *VII Conferência Internacional de TIC na Educação: Perspetivas de Inovação*, Braga (Bortugal), 2011.

[Verissimo 13] Verissimo, S., La introducción de las ideas de la Teoría de Conceptos Nucleares en la enseñanza de la Geometría y sus implicaciones, Badajoz: Universidad de Extremadura, 2013, p. Tesis doctoral.