

PROPUESTA METODOLÓGICA PARA REALIZAR MAPAS DE CONOCIMIENTO*

MARCO DUEÑAS**, DIANA ROJAS*** & MARÍA EUGENIA MORALES****
UNIVERSIDAD MILITAR NUEVA GRANADA

Recibido/ Received/ Recebido: 11/05/2011 - Aceptado/ Accepted / Aprovado: 01/05/2012

Resumen

Recientemente Chavalarías y Cointet (2008) propusieron una metodología para construir mapas de conocimiento. En dicho estudio el corpus de información se construye a partir de un conjunto de palabras clave fijas y determinadas de forma exógena. En este artículo se replica esta metodología, pero a diferencia, el conjunto de palabras clave es construido a través de búsquedas recurrentes y relacionadas. Así, iniciando con un conjunto reducido de palabras clave pertenecientes a un paradigma científico se construye un corpus de artículos, los cuales a su vez citan otras palabras clave que son usadas como insumo para una segunda búsqueda. Este estudio tomó como base artículos relacionados con aplicaciones satelitales. La metodología propuesta permite de igual forma construir mapas de conocimiento, o determinar los campos paradigmáticos (Chavalarías y Cointet, 2008), es decir se obtienen clusters conectados, compuestos de palabras relacionadas entre sí bajo un criterio específico de proximidad y de interrelación, que vistos de forma global permiten identificar áreas específicas del conocimiento

Palabras clave: mapas de conocimiento, aplicaciones satelitales, paradigma científico, paradigma tecnológico, cienciometría

METHODOLOGICAL PROPOSAL TO BUILD KNOWLEDGE MAPS

Abstract

Recently, Chavalarías and Cointet (2008) proposed a methodology to build knowledge maps. Such study considers the body of information built from a group of fixed key words, which are determined exogenously, this article replicates this methodology, but the group of key words is built through recurrent and related searches. Thus, beginning with a reduced group of key words belonging to a scientific paradigm, a body of articles is built, which at the same time cite other key words used as input for a second search. This study used as platform articles related to satellite apps, the proposed methodology allows building knowledge maps or determining paradigmatic fields (Chavalarías and Cointet, 2008), that is, obtaining linked clusters, including related words under a specific criterion of proximity and interrelation, which seen from a global perspective allow identifying specific areas of knowledge.

Keywords: knowledge maps, satellite applications, scientific paradigm, technological paradigm, scientometrics

* Artículo de investigación resultado del proyecto: Mapa de conocimiento en aplicaciones satelitales para la observación de la tierra. Financiado por la Universidad Militar Nueva Granada durante el 2009. Los autores agradecen especialmente la colaboración brindada por COLCIENCIAS por el acceso al portal de publicaciones científicas www.ScienceDirect.com

** Estudiante de Doctorado en Economía Scuola Superiore Sant'Anna di Pisa-Italia. Físico con Maestría en Economía. Miembro del grupo GECEI de la Universidad Militar Nueva Granada. Correo electrónico: m.duenasesterling@sssup.it

*** Estudiante de Doctorado en Management, Innovation, Sustainability and Healthcare. Scuola Superiore Sant'Anna. Istituto di Management. Laboratorio Management e Sanita MeS. Pisa- Italia. Correo electrónico: d.rojastorres@sssup.it

**** Administradora de Empresas, Magister en Administración, Universidad Nacional de Colombia. Docente Investigadora de la Facultad de Ciencias Económicas de la Universidad Militar Nueva Granada, miembro del Grupo de Estudios Contemporáneos en Contabilidad y Gestión de Organizaciones. Correo electrónico: maria.morales@unimilitar.edu.co

PROPOSTA METODOLÓGICA PARA REALIZAR MAPAS DE CONHECIMENTO

Resumo

Recentemente Chavalarias e Cointet (2008) propuseram uma metodologia para construir Mapas de Conhecimento. Em seu estudo o corpus de informação se constrói a partir de um conjunto de palavras chave fixas e determinadas de forma exógena. Neste artigo se repete esta metodologia, com a diferença de que o conjunto de palavras chave é construído através de buscas frequentes e relacionadas. Assim, iniciando com um conjunto reduzido de palavras chave pertencentes a um paradigma científico, se constrói um corpus de artigos, os quais por sua vez citam outras palavras chave que são usadas como insumo para uma segunda busca. Este estudo tomou como base artigos relacionados com aplicações satelitares. A metodologia proposta permite de igual forma construir mapas de conhecimento ou determinar os campos paradigmáticos (Chavalarias e Cointet, 2008), isto é, se obtêm clusters conectados, compostos de palavras relacionadas entre si sob um critério específico de proximidade e de inter-relação, que vistos de forma global permitem identificar áreas específicas do conhecimento.

Palavras chave: mapas de conhecimento, aplicações satelitares, paradigma científico, paradigma tecnológico, cienciometria.

Dueñas, M., Rojas, D. & Morales, M. (2012) Propuesta metodológica para realizar mapas de conocimiento. En: Revista de la Facultad de Ciencias Económicas de la Universidad Militar Nueva Granada. rev.fac.cienc.econ, XX (1)

JEL: C6, C9, O3.

1. Introducción

Parte de la política de desarrollo económico o social de un país enfocada a la promoción de la ciencia y tecnología (C&T) se basa en el incentivo de áreas del conocimiento y líneas o programas de investigación.¹ Sin duda, inicialmente, el problema al que se enfrenta el gobierno es indagar por la existencia de capital humano especializado, grupos de investigación e infraestructura adecuada que pueda enfocarse en temáticas específicas y estratégicas de las cuales se espera promover áreas de investigación y desarrollo. Así, la decisión sobre qué incentivar primero corresponde a solucionar un problema que puede ser complejo, pues en la mayoría de los casos, es posible que se tenga conocimiento de qué se quiere desarrollar, pero no exactamente cuáles son los pasos intermedios a seguir ya que no es posible

asegurar con precisión el tiempo que tomará comprender/descubrir conocimientos, ni la trayectoria que se seguirá.

Los mapas de conocimiento proporcionan una forma estructurada de capturar información relevante para el desarrollo de la C&T (Howard, 1989). Estudios en este campo entienden el conocimiento como un espacio por el cual es posible navegar, con diversas trayectorias, en donde cada punto conocido puede ser explorado/explotado hacia nuevos caminos. Sin embargo, cuando se piensa en aplicar esta estrategia a países en vía de desarrollo, gran parte del conocimiento global ha sido generado en países industrializados y lo que se está por aprender representa un espacio completo por explorar. Si fuera posible reconocer las trayectorias y experiencias que otros países han seguido, éstas serían de gran ayu-

¹ Entendiendo áreas como el conjunto de muchos programas o líneas de investigación.

da para el agente promotor de la política siendo un complemento estratégico para lograr cerrar la brecha en C&T.

Desde el punto de vista metodológico, los mapas de conocimiento representan nuevas estrategias para entender cómo se relacionan los “Paradigmas Científicos” (Kuhn, 1970) y los “Paradigmas Tecnológicos” (Dosi, 1982). Sin embargo, aún no existe un método globalmente aceptado o estandarizado para la construcción de estos mapas. Esto se debe en parte a que el planteamiento de las metodologías ha estado ligado a la disponibilidad de la información, constantes avances de herramientas computacionales (ej. Minería de datos, Análisis de redes complejas) y avances en la capacidad de procesamiento de los datos.

La metodología empleada en este artículo se basó en la reconstrucción de mapas de conocimiento utilizada por Chavalarias & Cointet (2008). Esta metodología usa palabras clave como unidad de análisis y permite establecer cómo se relacionan diferentes campos científicos alrededor de un tema determinado, por medio de la intensidad de la investigación (en términos del número de publicaciones), llevando al hallazgo de clusters de temas muy correlacionados que forman áreas de desarrollo científico y/o tecnológico.

Una parte importante *ex-ante* la aplicación de esta metodología es la selección de las palabras clave. Obviamente esto depende del campo científico a analizar. Regularmente estas palabras están dadas desde un inicio restringiendo los criterios de búsqueda. También, debe tenerse en cuenta que la replicabilidad de los resultados depende del corpus de artículos empleado, el cual se obtiene a partir del conjunto de palabras clave.²

En este artículo se muestra que usando la metodología de Chavalarias & Cointet (2008) es posible construir mapas de conocimiento sin necesidad de usar un conjunto de palabras clave previamente determi-

nado. De esta forma se propone construir el corpus de información incluyendo iteradamente más referencias con palabras clave ya citadas en el corpus mismo. Esta metodología representa una ventaja cuando no se tiene una idea muy clara de cuáles son las palabras que mejor describen algún paradigma, y adicionalmente por defecto permite construir mapas estrictamente conexos, dado que en cada iteración las búsquedas están correlacionadas.

El artículo se organiza de la siguiente forma: en la segunda sección se revisan brevemente algunas metodologías para la construcción de mapas de conocimiento y se justifica el uso de ocurrencias y co-ocurrencias de palabras clave para la construcción de estos mapas. En la tercera sección se explica la metodología utilizada para construir la base de datos y los mapas de conocimiento. La cuarta sección muestra los resultados obtenidos para el cálculo de clusters con el objetivo de ver las relaciones entre conceptos de forma específica, y una sección final presenta las conclusiones de este artículo.

2. Una aproximación a las redes entre documentos

El proceso de creación de conocimiento proviene de un proceso colectivo dentro de las comunidades (Allen, 1983). En muchos casos, estas comunidades son corporaciones especializadas, ej. sociedades científicas o academias, que promueven las publicaciones científicas. Estos documentos cumplen con la función de difundir/compartir conocimiento y lograr reconocimiento entre autores. Todo ese proceso contribuyó en la articulación de las comunidades mismas, especializándose en campos de investigación con problemáticas, procedimientos y tareas propias, es decir, facilitando lo que se conoce como “Paradigmas Científicos” (Kuhn, 1970). Las visiones de invención colectiva de Allen (1983) y de paradigma científico de Kuhn (1970), han sido exploradas cuantitativamente utilizando la Teoría de Redes, conduciendo a lo que hoy se conoce como Redes

² En efecto, una crítica factible, y no solamente a esta metodología sino a la gran mayoría empleada en la construcción de mapas de conocimiento, es que no se sabe hasta qué punto representan estrategias robustas. Es decir, cual es el tamaño óptimo de palabras clave y del corpus de información para que lo que se observa no sea ruido.

de Citación entre Pares y Redes de Conceptos, respectivamente.

Recientemente en cienciometría se han desarrollado técnicas para la construcción de redes de conceptos, (ver por ejemplo Glenisson et al. 2005; Marshakova 2005; Janssens et al. 2006). La mayoría de ellas están basadas en el análisis de la frecuencia con la que palabras u oraciones son citadas en cualquier parte del contenido del documento, como es el caso de Marshakova (2005) y Janssens et al. (2006). Este tipo de trabajos regularmente toman como base el texto completo y usan sofisticados y complejos algoritmos para identificar conceptos en diferentes partes del mismo, por lo tanto existen importantes restricciones en cuanto a: i) el formato de los documentos seleccionados, ii) la posibilidad de usar un gran número de artículos, y iii) el manejo de memoria de procesamiento de la información para reconocer grandes cadenas de texto e interpretar su significado dentro de algún contexto.³

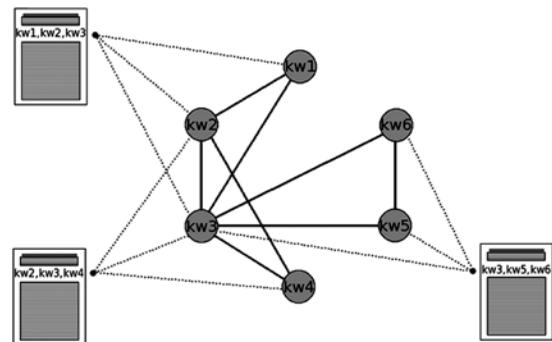
Un desarrollo paralelo al análisis de contenidos corresponde al análisis de ocurrencias y co-ocurrencias de palabras clave. Cuando se habla de ocurrencias y co-ocurrencias entre palabras clave, se quiere decir que para un conjunto de artículos se hace el conteo de cuantas veces aparece una palabra clave, o sea la ocurrencia, y se enumera cuantas veces aparecen citadas simultáneamente dos palabras clave en diferentes artículos, es decir las co-ocurrencias.

El método de análisis de palabras clave, fue inicialmente propuesto por Callon, et al. (1983), con el propósito de evaluar la investigación en comunidades científicas, bajo la hipótesis de que el análisis de co-citación de palabras clave describe la correlación entre los contenidos principales de los documentos. En este sentido el análisis de co-ocurrencia de palabras clave, puede representar una simplificación

que permite incluir más información (mayor número de documentos), restringiéndose específicamente a la citación de las palabras clave (menor memoria y capacidad de computo) declaradas en el documento científico, es decir, no se realiza una búsqueda de la ocurrencia en el documento entero.

Chavalarias & Cointet (2008), realizan una versión más elaborada de las ideas de Callon, et al. (1983), con el propósito de mostrar cómo la manipulación de palabras clave permite revelar la conexión entre campos científicos de investigación, usando sólo las palabras clave citadas en diferentes artículos. La metodología propone identificar un conjunto de artículos {A} relacionados con un tema determinado, junto con un conjunto seleccionado de conceptos o palabras clave {C_i}.⁴ Cada publicación representa un vínculo entre el conjunto de palabras clave que contiene. Por ejemplo en la ilustración 1, se presentan tres artículos cada uno con tres palabras clave, en adelante kws, la mayor ocurrencia la presenta kw₃, ya que aparece tres veces, y la mayor co-ocurrencia aparece entre kw₂ y kw₃, coincidiendo en dos artículos. El conjunto de líneas continuas en la ilustración 1 representa la red de co-ocurrencias de palabras clave, donde cada vínculo está determinado por el peso de la co-ocurrencia entre el par de nodos que une.

Ilustración 1. Red de co-ocurrencias de palabras clave⁵



³ Un ejemplo de este tipo de herramienta es el paquete "Matheo Analyzer", que ofrece un conjunto de herramientas para hacer análisis bibliométrico y mapeo de información.

⁴ Se aclara que un concepto no necesariamente está determinado por una única palabra, es por esto que muchos de los avances tecnológicos, actuales, en ciencias de la información apuntan a la construcción de algoritmos, rápidos y eficientes, capaces de interpretar conjuntos de palabras. Sin embargo, en este trabajo se seguirá la hipótesis de Callon et. al, donde las palabras clave son el reflejo de los conceptos que se discuten en los documentos científicos, y por esto, en adelante, conceptos o palabras clave significarán lo mismo.

⁵ Elaboración propia.

La ocurrencia de la palabra kw_i en $\{C\}$ representa el número de artículos en que aparece citada y se denotará n_i , y la co-ocurrencia entre cada par de elementos kw_i y kw_j en $\{C\}$ es el número de artículos en que aparecen citadas las dos y se denotará $n_{i,j}$. La matriz simétrica con los elementos $n_{i,j}$ se denomina la matriz de co-ocurrencias, por ejemplo para el caso de la ilustración 1, esta matriz es:

$$(1) \quad n_{i,j} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 2 & 2 & 1 & 0 & 0 \\ 1 & 2 & 3 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

La medida de proximidad se basa en la matriz de co-ocurrencias. Así, para un par de palabras kw_i , kw_j la proximidad $P(i,j)$ se define como medida de proximidad.

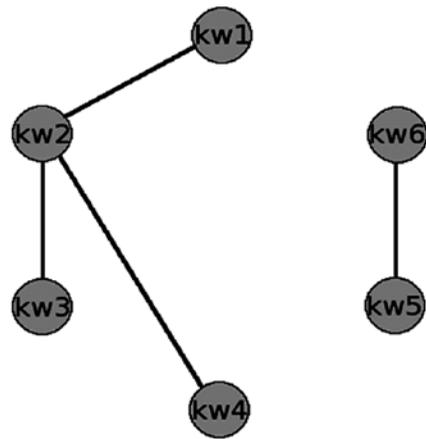
$$(2) \quad P(i,j) = \frac{n_{i,j}^2}{n_i n_j}$$

Note que esta medida está definida en el intervalo $[0,1]$, por lo tanto cuando existen temas muy generales, es decir, muchas veces nombrados, prácticamente la proximidad a otros conceptos es nula. Por el contrario, la medida revelaría una relación más específica cuando para un par de conceptos aparecen siempre en presencia del otro, es decir, máxima proximidad. Por ejemplo, para el caso de la ilustración 1, con esta medida es posible construir la matriz de proximidad:

$$(3) \quad P(i,j) = \begin{pmatrix} 1 & 0.50 & 0.33 & 0 & 0 & 0 \\ 0.50 & 1 & 0.67 & 0.50 & 0 & 0 \\ 0.33 & 0.67 & 1 & 0.33 & 0.33 & 0.33 \\ 0 & 0.50 & 0.33 & 1 & 0 & 0 \\ 0 & 0 & 0.33 & 0 & 1 & 1 \\ 0 & 0 & 0.33 & 0 & 1 & 1 \end{pmatrix}$$

El siguiente paso es establecer un valor mínimo para la proximidad con el objetivo de eliminar ruido, descartando vínculos muy débiles y eliminando co-ocurrencias dadas por el azar. Este procedimiento implica que se consideran solamente pares de conceptos con proximidades que supera algún umbral w , seleccionado exógenamente. Lo que se pretende es poder obtener vínculos entre conceptos en función de un nivel de especificidad entre ellos y facilitar el análisis. La ilustración 2 muestra el mapa de proximidades cuando el investigador asume que el umbral debe ser $w=0.4$.

Ilustración 2. Ejemplo, mapa de proximidades⁶



Una característica de las comunidades científicas en los últimos tiempos es la intensidad de publicaciones y la gran diversidad de palabras clave. Lo que dificulta la interpretación de los resultados. En efecto, la medida de proximidad filtra gran parte de información deseable, pero el remanente es una mezcla compleja de relaciones de la que se espera extraer alguna información. El análisis de clusters permite encontrar grupos de conceptos que se asemejan en la forma que se relacionan.

El método denominado en la literatura anglosajona como The Clique Percolation Method se emplea frecuentemente para el análisis de redes complejas y está diseñado para localizar comunidades com-

⁶ Elaboración propia.

puestas por k-cliques (Palla et al., 2005). El método mismo define a las comunidades como la unión de pequeños, pero completos, sub-grafos que comparten nodos.⁷ Estos sub-grafos dentro de la red son llamados k-cliques, donde k se refiere al número de nodos en el sub-grafo, y una comunidad de k-cliques es definida como la unión de todos los k-cliques que pueden ser unidos unos con otros como una serie de sub-grafos adyacentes. De esta forma, dos sub-grafos son adyacentes si ellos comparten k-1 nodos. Estructuralmente, la diferencia entre la escogencia del nivel de cliques se relaciona con la conectividad de la red de clusters y la distribución del tamaño de los mismos.

La metodología conduce redes similares en función del umbral de la medida de proximidad y el número de cliques escogidos. Al cambiar tanto w como k se modifica la estructura de la comunidad que se investiga, cuando k disminuye las comunidades se hacen más pequeñas y tienden a desaparecer. El criterio usado para determinar valores óptimos de estos parámetros es:

- un umbral w que permita la mayor inclusión de términos relacionados,
- un valor para los cliques k que conduzca a una red de clusters lo más conexa posible.

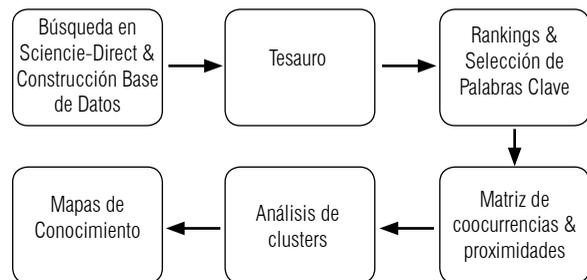
3. Metodología

El valor agregado del mapeo de conceptos es el conocimiento de cómo las comunidades conectan diferentes conceptos brinda nueva información relevante a nivel local y global. En muchos casos el insumo de esta metodología es un conjunto de conceptos escogidos ad hoc, por lo tanto se debe tener una idea muy clara de que estos brindarán una representación útil. En este artículo se propone una metodología donde las palabras clave no son escogidas ad hoc sino que son calculadas de forma iterativa para determinar este grupo de conceptos.

La idea fundamental supone que si existe un mapa de conocimiento ideal y conexo que describe un paradigma, entonces es posible construir una buena aproximación indagando por los vecinos, y los vecinos de estos vecinos para un pequeño grupo de nodos que describen globalmente el paradigma.

Para probar esta idea se tomará como estudio de caso las aplicaciones satelitales para la observación de la tierra, los pasos que se siguieron para la construcción del mapa de conocimiento se esquematiza en la ilustración 3.

Ilustración 3. Metodología⁸



3.1. Descripción y construcción de la base de datos

Se usó como fuente de información la base de datos de publicaciones científicas www.ScienceDirect.com. La información recopilada correspondió a títulos, resúmenes y palabras clave (kws). Antes de explorar la base de datos se propusieron las siguientes kws (en idioma inglés); sensoramiento remoto, geodesia, climatología y navegación satelital. El objetivo de la lista fue encontrar algunos argumentos para acompañar a la palabra “satélite” en la búsqueda y evitar confusiones.⁹

Es necesario tener como referencia un conjunto de kws para poder iniciar la búsqueda, y después construir un bucle que permita complementar la infor-

⁷ Las ventajas de esta definición de comunidad son las siguientes: es local, está basada en la densidad de enlaces, permite intersecciones, es decir, un nodo puede ser miembro de diferentes comunidades al mismo tiempo.

⁸ Elaboración propia.

⁹ Por ejemplo, con frecuencia, en biología también es usada la palabra “Satellite” para referirse a un tipo de virus presente en la naturaleza.

mación, es decir incluir y modificar el conjunto de referencia. También es necesario elegir hasta qué orden de búsqueda se quiere llegar, de lo contrario se obtendría una base de datos como para hacer una descripción total de todas las ciencias donde probablemente un nodo representaría las aplicaciones satelitales.

La recolección fue realizada en dos fases:

- *Fase 1.* Se buscaron todas las publicaciones que obedecieran a la ecuación de búsqueda: Title OR Abstract OR Key-Words = “kws” AND “Satellite”. Restringiendo la búsqueda a revistas especializadas en: “Earth and Planetary Sciences” OR “Engineering” OR “Environmental Science” OR “Physics and Astronomy”. Lo anterior para los años: 1997-2007.

La sentencia “kws” se modificó con los elementos del conjunto inicial. Los elementos encontrados en cada cambio eran agregados al corpus de artículos eliminando siempre posibles títulos repetidos.

A pesar de incluir el condicional AND en la ecuación de búsqueda, el corpus generado puede ser insuficiente para la construcción del mapa, pues es probable que existan otras aplicaciones fuera del primer sistema de kws propuestas por el investigador. Sin embargo, con esta información se realizó un ranking de las palabras más citadas con el propósito que la propia base revelara términos que podrían reforzarse en la búsqueda.

- *Fase 2.* Con el ranking se tomaron las primeras palabras que tenían una ocurrencia de por lo menos 50 veces, esta lista contenía alrededor de 40 palabras. Entonces, con los elementos de este nuevo conjunto de palabras se realizó de nuevo la búsqueda modificando la ecuación de búsqueda en la sentencia “kws”, y como antes, únicamente los nuevos datos generados se fueron agregando al corpus.

El mecanismo de recolección de ocurrencias en título, resumen o conjunto de kws, comparado con la búsqueda de ocurrencias en todo el documento, tiene la ventaja de que el conjunto recolectado de palabras clave se relacionan de forma más cercana con la búsqueda. Es decir, se tiene la hipótesis de que la ocurrencia en cual-

quier parte del documento de la ecuación de búsqueda incluye más palabras clave y a la vez más ruido al análisis.

La restricción impuesta sobre la búsqueda que se relaciona con el tipo de revistas, asume que sólo en ese tipo de revistas, es donde se publican investigaciones relacionadas con tecnologías satelitales, y además, con el ánimo de evitar combinar temas de otras disciplinas que puedan manejar conceptos parecidos. A pesar de los controles impuestos es posible incluir información sobre satélites naturales aunque no es evidente en los resultados. Sin embargo, controlar este tipo de sesgos en algunos casos puede ser prácticamente imposible, pues no existe un mecanismo que exija a las comunidades científicas no usar palabras homónimas que lleven a confusiones.

En cuanto al formato de la información, para cada búsqueda el portal www.ScienceDirect.com permite salvar las referencias de los elementos encontrados en diferentes formatos. El formato utilizado fue Research Information Systems (RIS) comúnmente reconocido por programas especializados en el manejo de referencias, como Reference Manager entre otros. Una ventaja de este formato es que permite distinguir con facilidad título, resumen, palabras clave, autores, año de publicación y nombre de la revista, entre otros.

3.2. Depuración de la información

En la literatura existen citas similares provenientes de abreviaciones o acrónimos que deben conocerse a la hora de manipular la información. Por ejemplo, en el tema relacionado con la ionosfera es común encontrar los términos TEC o “Total Electron Content” que significan lo mismo. La depuración consiste en construir un Tesauro para el conjunto de palabras clave. En este paso el trabajo fue: investigar que significaban las abreviaciones que aparecían en la lista y armar conjuntos con los términos no abreviados; agrupar las palabras con significados similares; y agrupar conjuntos de palabras que aunque se escriban un poco diferentes significan lo mismo, por ejemplo el conjunto $Hard_X-ray = \{Hard\ X-ray, Hard-X-ray\}$. La aplicación del tesauro en efecto modifica el orden del ranking, precisamente esto es

lo que se busca, que el tesoro conduzca un orden más aproximado de las palabras clave más importantes.¹⁰

Después de obtener el ranking final de palabras, se eligieron únicamente las palabras que aparecían citadas en por lo menos diez artículos asegurando así encontrar temas emergentes o vínculos débiles conectando grandes temas de investigación. El corpus de artículos construido está compuesto por 21.000 títulos entre los años 1997 y 2007, la lista de palabras clave a estudiar contienen 436 elementos con los cuales usando la matriz de co-ocurrencias se construyó un algoritmo adicional en Python para el cálculo de las proximidades. Finalmente, se realizó el análisis de clusters a esta matriz de proximidades.¹¹

4. Resultados

Para la construcción del mapa de conocimiento se determinó que el umbral óptimo correspondía a $w=0.005$. Se obtuvieron 8 clusters para $k=5$ -cliques, 26 clusters para $k=4$ -cliques, y 53 clusters para $k=3$ -cliques. Se eligió la configuración de tres cliques fundamentalmente porque no aparecían clusters desconectados. Para diferentes parámetros k se obtuvieron clusters con diferentes tamaños, en particular para $k=3$ -cliques la distribución de los tamaños es bastante asimétrica, se encuentran unos pocos de gran tamaño conectados con otros de tamaño inferior. Para $k=4$ los tamaños son un poco menos diferenciados, pero no hay conectividad entre clusters. Para $k=5$ la distribución de tamaños es uniforme, no presenta conectividad, y existen muy pocos elementos, limitando el análisis.

Por otro lado, se observó que para umbrales menores a 0,005 aparecían más clusters, pero, la gran mayoría eran clusters de muy pocos términos conectados alrededor de uno o dos clusters centrales, los cuales contenían más del 80% de las palabras elegidas. Este

escenario no era deseable dado que clusters de gran tamaño poseen demasiadas conexiones dificultando el propósito de distinguir entre áreas de conocimiento. Para umbrales $w>0,005$ se encontraban pocos clusters y inconexos, imposibilitando la construcción del mapa de conocimiento.

4.1. Mapa de conocimiento

El mapa de conocimiento calculado se muestra en la ilustración 4, el grafo contiene 53 clusters obtenidos para proximidades que cumplen la condición $0.005 < P < 1.0$, ajustando el parámetro de k -cliques $k=3$. Cada cluster aparece numerado y su tamaño indica el número de palabras clave que contiene. Adicionalmente, en la figura se han escrito las áreas que hacen referencia a las temáticas de los clusters. Los nombres que se han utilizado en cada cluster hacen referencia a áreas científicas que describen de forma más apropiada las palabras contenidas en uno de ellos. El anexo 1 contiene la información de todas las palabras clave correspondientes a cada cluster.

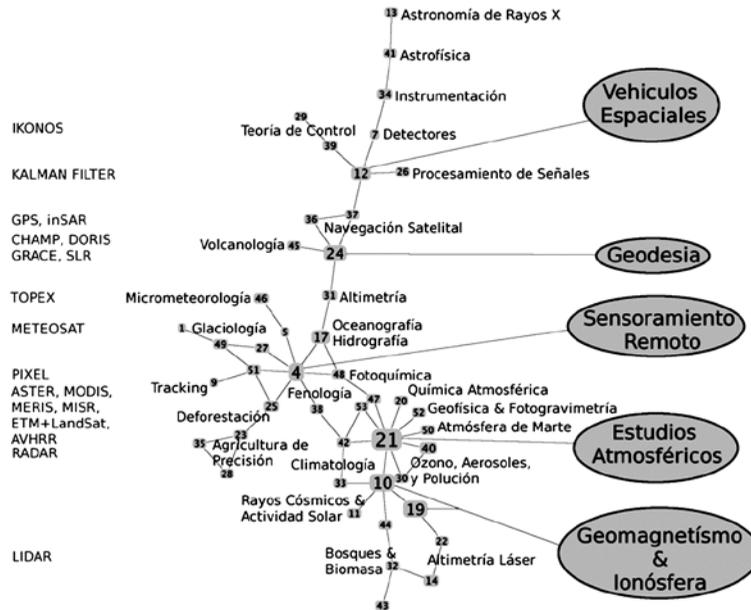
El mapa calculado permite observar la conexión entre diferentes ramas de la investigación básica y aplicada. Además se identifican 5 clusters articuladores, caracterizados por poseer mayor tamaño y conexiones en comparación con los demás. Estos han sido nombrados: Vehículos Espaciales (12); Geodesia (24); Sensoramiento Remoto (4); Estudios Atmosféricos (21) y Geomagnetismo e Ionosfera (10). Este tipo de clusters están conformados por muchas palabras clave que pueden fácilmente identificarse como parte de estas áreas de conocimiento, ver anexo 2 para un ejemplo de la composición y conexión de un cluster. A la vez, los clusters restantes, están conformados en su mayoría con tres palabras clave, o más, y se aglomeran alrededor de éstas áreas. A diferencia de los clusters grandes, los pequeños revelan temas mucho más específicos.¹²

¹⁰ Existen algunos paquetes informáticos que pueden ayudar en este tipo de análisis bibliométrico. En particular en este trabajo se usó el paquete *Vantage Point*, el cual proporciona una interface gráfica bastante amigable para la construcción del Tesoro y además permite encontrar la matriz de co-ocurrencias.

¹¹ Para el cálculo de los clusters se utilizó el software *CFinder – Cluster & Communities* el cual está basado en el Método de Percolación de Cliques (Palla et al. 2005). Éste permite calcular los clusters a diferentes cliques en función de un umbral mínimo elegido por el usuario.

¹² El hecho de que estos pequeños clusters contengan menos términos, facilitó el proceso de encontrar un nombre para ellos, en la mayoría de los casos se usó el nombre del miembro que mejor representaba al grupo.

Ilustración 4. Estructura del mapa de conocimiento en aplicaciones satelitales¹³



Lo que se puede observar de este ejercicio es que los satélites se han utilizado para múltiples propósitos. Se observan clusters enfocados a: predicción del clima; observación de cambios climáticos y su relación con los seres vivos (Fenología); cambio climático y descongelamiento polar; detección de concentración de ozono estratosférico, deforestación, estudio de océanos y ríos, actividad en volcanes, medición de índices de vegetación, agricultura de precisión, monitoreo de la actividad solar, rastreo de objetos en tierra, mar y aire. Según estos resultados, diferentes ramas de las ciencias básicas e ingeniería coexisten y son el soporte de estas aplicaciones, estas son: espectrometría (a diferentes frecuencias del espec-

tro electromagnético), astrofísica, geomagnetismo, geofísica, fotogrametría, química atmosférica, altimetría, teoría de control, telecomunicaciones y procesamiento de señales, y vehículos espaciales.

Entre los conjuntos de palabras clave, es común encontrar nombres de sensores y satélites, ver costado izquierdo de la ilustración 4. El hecho de que aparezcan combinados conceptos con nombres de dispositivos, se debe principalmente a que cada dispositivo, es diseñado para llevar a cabo tareas específicas a determinadas alturas y frecuencias; en la mayoría de los casos se reconoce un tipo de sensor por aplicación.¹⁴ En realidad, el diseño de los satéli-

¹³ Elaboración propia.

¹⁴ Con respecto a la trayectoria de los satélites, se denominan geostacionarios (GEO) a los que poseen una órbita circular ecuatorial de altitud 35786 Km, de órbita media (MEO) altitud de 9000 a 14500 Km y de órbita baja (LEO); altitud de 725 a 1450 Km. Un ejemplo de satélite es el GOES-P, el último de los Satélites Ambientales de Operación Geostacionaria desarrollado por la NASA, para la Administración Atmosférica y Oceánica Nacional (NOAA) y lanzado en marzo de 2010; se encuentra a 35790 Km sobre la tierra, en órbita geostacionaria con un peso de 2100 Kg es operado por Estados Unidos, Centroamérica y Sudamérica; monitorea nubes, temperatura superficial y vapor de agua en tiempo real; ayudando a los meteorólogos a observar y predecir eventos de lluvias locales, tormentas, tornados, inundaciones, erupciones volcánicas e incendios forestales. Otro tipo de satélite es el IRIDIUM que se utiliza principalmente en telefonía celular; este consta de 66 satélites LEO, los cuales se encuentran a una altitud de 725-1450 Km; cada satélite pesa aproximadamente 700 Kg, su periodo de vida activa es de 5 a 8 años y su margen de enlace es de 16 dB. Satélites de posicionamiento se encuentra el Sistema de Posicionamiento Global (GPS), este es un satélite de navegación desarrollado y mantenido por el Gobierno de Estados Unidos, inicialmente diseñado para aplicaciones militares, donde posteriormente se han encontrado usos en todos los campos.

tes (peso, altura y trayectoria de la órbita), depende del tipo de aplicaciones deseadas, renunciando a la posibilidad de instalar otros dispositivos. La información generada por cada dispositivo, es enviada por el sistema de comunicación del satélite a una estación en órbita o terrea.

5. Conclusiones

Este artículo contribuye a las metodologías para la construcción de mapas de conocimiento. Se muestra que el corpus de información puede ser obtenido a partir de las ocurrencias de un pequeño grupo de palabras clave que describen apropiadamente un paradigma científico. Desde el punto de vista de las relaciones entre conceptos, esto significa que basta con indagar por las relaciones de los vecinos conocidos, es decir a partir de información local es posible adjuntar elementos al corpus que permiten relacionar conceptos de forma robusta.

Se encontró un mapa de conocimiento relacionado con las aplicaciones satelitales tomando como proxy de conocimiento las palabras clave citadas en publicaciones científicas, usando una base de datos de 21.000 títulos entre 100 revistas científicas diferentes desde 1997 hasta 2007. En el conjunto de palabras clave fue posible identificar nombres de satélites, sensores, y proyectos de agencias espaciales. Para lograr esto se calcularon las co-ocurrencias y proximidades, y luego se determinaron los clusters entre las palabras clave.

De la aplicación de la técnica de *k*-cliques se concluyó que cada cluster puede considerarse como un conjunto de conceptos que pertenecen a algún paradigma científico. Entonces, visto de forma agregada, el mapa de conocimiento permite reconocer como se interrelacionan diferentes paradigmas científico-tecnológicos, y la relación entre las palabras clave que aparecen con más frecuencia en publicaciones científicas. Se encontraron algunos cluster que contienen una gran cantidad de palabras clave en comparación con los demás. Estos clusters mostraron gran conectividad tanto de forma global entre clusters como entre las palabras clave que los componen. La gran conectividad entre clusters demostró que se trata de temas que son punto común entre

diferentes paradigmas, es decir áreas completas de investigación.

Esta metodología permitió observar que las tecnologías satelitales son instrumentos útiles en diferentes tipos de investigación básica, aplicada, de campo, científica y tecnológica; con técnicas basadas en sensoramiento remoto que permiten conocer el contorno exacto del mundo y su atmósfera, aspectos etnológicos, históricos, estadísticos, hidrográficos, geomorfológicos, geológicos, entre muchos otros. Los mapas de conocimiento revelan las relaciones entre las ciencias básicas con la investigación aplicada científica y tecnológica que agencias espaciales de países desarrollados han seguido.

El estudio se realizó para un periodo de tiempo en el cual las investigaciones y el desarrollo de nuevos satélites o mejoramiento de los mismos estaban en auge, se espera que en los próximos años las investigaciones lleven a más aplicaciones y por lo tanto el mapa se modifique. Actualmente hay gran cantidad y variedad de proyectos respecto al tema, pero precisamente lo interesante de la metodología es poder seguir periodo a periodo los cambios tecnológicos para así mismo poder orientar las nuevas investigaciones hacia temas innovadores y no hacia estudios ya realizados. Estudios en esta dirección representan nuevas metodologías para reconocer como se relacionan los paradigmas científicos alrededor de una aplicación tecnológica, y a su vez el poder predecir la emergencia de nuevos paradigmas tecnológicos.

Los mapas de conocimiento permiten visualizar las relaciones que se presentan en el caso de estudio, facilitando el proceso de ubicación del conocimiento relevante, así como conectar los procesos con las fuentes de conocimiento, siendo una forma posible para mejorar la transferencia de conocimiento y utilización en una sociedad. Para países en vía de desarrollo, estos vínculos pueden contribuir a la planeación de programas de investigación, reconociendo la trayectoria que han seguido países desarrollados.

6. Referencias

- Allen, R. C. (1983) "Collective Invention". En: *Journal of Economic Behavioral Organization*, The University of British Columbia, (4): 1-24.

- Callon, M., Courtial, J-P., Turner, W. & Bauin, S. (1983) "From translations to problematic networks: An introduction to co-word análisis". En: *Social Science Information*, SAGE Journals Online, (22): 191-235.
- Chavalarias, D. & Cointet, J. P. (2008) "Bottom-up scientific field detection for dynamical and hierarchical science mapping, methodology and case study". En: *Scientometrics*, Vol. 75, (1): 37-50.
- Dosi, G. (1982) "Technological paradigms and technological trajectories. A suggested interpretation of the determinants and directions of technical change". En: *Research Policy*, Vol. 11, (3): 147-162.
- Glenisson, P., Glänzel, W., Janssens, F., & De Moor, B. (2005) "Combining full text and bibliometric information in mapping scientific disciplines". En: *Information Processing & Management*, Elsevier, (41): 1548-1572.
- Howard, R. (1989) "Knowledge Maps". En: *Management Science*, Vol. 35, (8): 903 - 922.
- Janssens, F., Leta, J., Glänzel, W., & De Moor, B. (2006) "Towards mapping library and information science". En: *Information Processing & Management*, Vol. 42, (6): 1614-1642.
- Kuhn, T. (1970) *The structure of scientific revolution* (2^a Ed.). Chicago: University of Chicago Press.
- Marshakova-Shaikevich, I. (2005) "Bibliometric maps of field of science". En: *Infometrics*, Vol. 41 (6): 1534-1547.
- Palla, G., Derenyi, I., Farkas, I., & Vicsek, T. (2005) "Uncovering the overlapping community structure of complex networks in nature and society". En: *Nature*, (435): 814-818.

Anexo 1

Palabras clave contenidas en cada cluster

No.	Palabras clave
0:	Aerosol_optical_depth, Aerosols, Size_distribution,
1:	METEOSAT, Albedo, Radiative_transfer_model,
2:	Atmospheric_electricity, Lightning, Thunderstorm,
3:	DEM, Accuracy, SPOT, SRTM, Topography,
4:	ASTER, Infrared, Landsat, MODIS, Remote_sensing, Scale, AVHRR, Drought, Evapotranspiration, NDVI, Phenology, Surface_Temperature, Time_series_analysis, Validation, Algal_blooms, Chlorophyll, Ocean_Color, Phytoplankton, Algorithms, SeaWiFS, Southern_Ocean, Atmospheric_correction, MERIS, Sea_ice, Primary_production, Boreal_forest, Carbon_cycle, ETM, GPP, Grassland, NPP, Sahel, urbanization, MISR,
5:	Evapotranspiration, Energy_balance, Sensible_heat_flux,
6:	Galaxies_active, Galaxies_Evolution, Radio_continuum_galaxies,
7:	Space_vehicles, Detectors, Gravitation, General_relativity, Lense-Thirring_effect,
8:	Integration, Lineaments, Image_analysis,
9:	Pixel, Silicon_detector, Tracking,
10:	Aurora, Plasma_sheet, Substorms, ICMs, Cosmic_rays, Corona, Coronal_mass_ejections, Heliosphere, MHD, Magnetic_field, Solar_wind, Sun, Energetic_particles, Geomagnetic_storms, Interplanetary_shock, Magnetic_clouds, Solar_activity, Solar_flares, Space_weather, Magnetosphere=Ionosphere_Interaction, Magnetosphere, Radiation_belts, Solar_wind/magnetosphere_interactions, Sunspot_number, Relativistic_electrons, Geomagnetic_activity, Ring_current, Solar_cycle, Inner_magnetosphere, Storms,
11:	Solar_activity, Solar-terrestrial_physics, Solar_irradiance,
12:	Artificial_satellite, Attitude_control, Control, Space_vehicles, Estimation, Kalman_filter, Stability, satellite_control, Parallel_processing, Optimization, Filtering, Signal_processing,
13:	Astrophysics, Black_holes, X-ray_Astronomy,
14:	LIDAR, Biomass, Forest, Height, Laser_altimetry, Hyperspectral,
15:	Emissions, Forest_fires, Transport,
16:	Carbon, Ecosystem, Soil,
17:	mesoscale_model, Hydrography, Phytoplankton, Altimetry, Eddies, Ocean_circulation, Circulation, Oceanography, Upwelling, Geoid,
18:	Computer_vision, Hough_transform, Image_processing, pattern_recognition,
19:	Thermosphere, Convection, Electric_field, Plasmasphere, Geomagnetic_storms, Electron_density, Low_latitude_ionosphere, Ionospheric_storms, Ionospheric_disturbances, Equatorial_ionosphere, Ionosphere, Scintillation, Electron_temperature, F-region, IRI, IRI = International_Reference_Ionosphere, Ionospheric_modeling, Topside_ionosphere, Magnetosphere, Equatorial_spread_F, Ionospheric_irregularities, Modeling_and_forecasting, Modeling, TEC, Ionospheric_variability, Mid-latitude_ionosphere,
20:	Temperature, Variability, Correlation,
21:	Abundances, Atmospheres_composition, Atmospheres_structure, Atmospheric_chemistry, Callisto, Mars_atmosphere, Neptune, Ultraviolet_observations, Venus, Carbon_monoxide, Ozone, meteorology, Airglow, Dynamics, Gravity_waves, Mesopause, Mesosphere, Temperature, Asteroids, Comets, Cratering, Impact_processes, Infrared_observations, Orbits, Photometry, Planetary_rings, Satellites_general, Atmosphere, Io, Jupiter, Satellite_measurements, Water_Vapor, Atmospheres_dynamics, Satellites_atmospheres, Saturn, Titan, Radar, Stratosphere, Radio_occultation, Ganymede, Ices, Interiors, Rotational_dynamics, Satellites_of_Jupiter, Thermal_histories, Interplanetary_dust, Satellites_surfaces, Saturn_satellites, Triton, Enceladus, Resonances, Tides, celestial_mechanics, geophysics, Geological_processes, Tectonics, Volcanism, Solar_cycle, Spectroscopy, Satellite, Troposphere,

PROPUESTA METODOLÓGICA PARA REALIZAR MAPAS DE CONOCIMIENTO

22:	Height, Empirical_models, Topside_ionosphere,
23:	Deforestation, Precision_Agriculture, Land_use_change,
24:	Tide_gauges, Orbit_determination, CHAMP, GPS, GRACE, Crustal_deformation, Strain, geodynamics, DORIS, Geodesy, Satellite_geodesy, Terrestrial_reference_frame, Gravity_field, Earthquakes, Volcanoes, deformation, subduction_zones, Plate_tectonics, Iceland, SLR, Hydrology,
25:	Remote_sensing, Deforestation, Land_cover, Land_use, GIS, Water_quality,
26:	Parameter_estimation, Signal_processing, Identification,
27:	Snow, Sea_ice, Microwave,
28:	Spatial_variability, Precision_Agriculture, Salinity,
29:	IKONOS, Texture, Coral_reefs,
30:	Air_pollution, Carbon_monoxide, Dispersion,
31:	Altimetry, Sea_level, TOPEX, tide_gauges,
32:	Forest, satellite_images, Segmentation,
33:	Clouds, Climate, Cosmic_rays,
34:	Cosmology, Gamma-ray, Cosmic_microwave_background, Detectors, observers, Instrumentation,
35:	Precision_Agriculture, Electrical_conductivity, Sensors,
36:	GPS, InSAR, Land_subsidence,
37:	Kalman_filter, GPS, Navigation,
38:	Phenology, Radiative_transfer, LAI,
39:	classification, Parallel_processing, Texture, feature_extraction,
40:	Particulate_matter, Air_quality, Air_pollution, Ozone, PM10, PM2.5, meteorology, Air_quality_modeling,
41:	Astrophysics, Gamma-ray, X-ray,
42:	Atmosphere, Clouds, Radiative_transfer,
43:	Entropy, Fuzzy_logic, Segmentation,
44:	Change_detection, ICMs, satellite_images,
45:	Monitoring, Volcanoes, Volcanic_eruptions,
46:	Vegetation, Energy_balance, Micrometeorology, Soil_moisture,
47:	Abundances, Atmospheric_chemistry, Photochemistry,
48:	Photochemistry, mesoscale_model, Validation,
49:	Reflectance, Albedo, Multispectral, Snow,
50:	Mars_atmosphere, Data_reduction_techniques, Mars,
51:	Multispectral, ETM, Land_cover, Pixel,
52:	Photogrammetry, geophysics, Landslide,
53:	Venus, Dynamics, Atmosphere,

Anexo 2

Ejemplo del contenido de clusters y sus conexiones

