

## Aplicación de la minería de textos. Análisis de patentes.

Javier Gavilanes<sup>1</sup>, Rosa Maria Rio<sup>1</sup>, Ernesto Cilleruelo<sup>2</sup>, Gaizka Garechana<sup>3</sup>

<sup>1</sup> Dpto. de Organización de Empresas. Escuela Universitaria de Ingeniería de Vitoria-Gasteiz. Universidad del País Vasco. Calle Nieves Cano 12, 01006. Vitoria-Gasteiz.

[javier.gavilanes@ehu.es](mailto:javier.gavilanes@ehu.es), [rosamaria.rio@ehu.es](mailto:rosamaria.rio@ehu.es).

<sup>2</sup> Dpto. de Organización de Empresas. Escuela Técnica Superior de Ingeniería de Bilbao. Universidad del País Vasco. Calle Alameda Urquijo s/n, 48013. Bilbao.

[ernesto.cilleruelo@ehu.es](mailto:ernesto.cilleruelo@ehu.es)

<sup>3</sup> Dpto. de Organización de Empresas. Escuela Universitaria de Estudios Empresariales de Bilbao. Universidad del País Vasco. Calle Elcano 21, 48008. Bilbao.

[gaizka.garechana@ehu.es](mailto:gaizka.garechana@ehu.es)

**Palabras clave:** Text mining, Patent, Innovation.

### 1. Antecedentes y Objetivos.

La competitividad de una nación debe basarse en la capacidad de sus industrias para innovar y mejorar continuamente en sus productos, servicios y procesos (Porter 1990).

Hoy en día se reconoce que la innovación no surge exclusivamente en la empresa, sino que encontramos su fuente en la interacción con otras organizaciones que presentan recursos complementarios: competidores, proveedores, clientes, centros de investigación o universidades, Bayona, C. et al. (2003), Cunningham et al. (2006), Porter (2007a) dando lugar a lo que cabe llamar “Red de Innovación y Desarrollo”, un contexto en el que instituciones, individuos y mercado colaboran creando un flujo de conocimiento científico técnico que terminará en una innovación tecnológica, Cunningham et al. (2006), Escorsa y Maspons (2001).

El éxito a la hora de innovar dependerá pues, en gran medida, de la capacidad que muestra la organización para transformar la información en conocimiento y hacer después circular este conocimiento como flujo continuo por toda la estructura organizativa Escorsa y Maspons (2001).

Por todo ello, podemos afirmar que la competitividad de las empresas depende en gran medida de su capacidad de innovar, y que la innovación en el contexto actual se basa principalmente en una adecuada gestión del conocimiento, obtenido a partir del análisis de la información.

El objetivo de este trabajo es mostrar las diferentes técnicas de análisis de patentes a través de la minería de textos y sus aportaciones en la toma de decisiones tecnológicas.

## **2. Minería de textos.**

Se conoce con el nombre de minería de textos a aquel proceso consistente en extraer conocimiento a partir de fuentes masivas de información en forma de texto que nos permita tomar decisiones estratégicas, Rio y Cilleruelo (2006).

Un análisis basado en la minería de textos se divide en dos partes: 1) Un refinado o tratamiento previo de los textos, dando lugar a una forma intermedia de la información, que sea manejable por técnicas computacionales. 2) Una extracción de conocimiento a partir de esta forma intermedia mediante herramientas usuales de minería de datos, con el fin de extraer pautas, relaciones o conocimiento.

Es destacable que una forma intermedia basada en documentos puede ser transformada en una basada en conceptos si se extrae de ellos la información relevante de acuerdo con los intereses del análisis, Tan (1999).

Otros autores indican que parte de la riqueza semántica contenida en los textos de partida se pierde al seguir el método inductivo clásico de minería de textos, por lo que proponen enfoques diferentes, planteando los procedimientos de la minería de textos como apoyo a la elaboración de hipótesis por parte de los expertos, o como herramienta de búsqueda de contradicciones, entre otros, Sanchez et al. (2008).

En esta misma línea, Porter (2007a) reconoce las limitaciones de las fuentes empíricas y sugiere combinarlas con fuentes basadas en conocimientos de expertos. Una forma interesante de hacerlo es construir indicadores de innovación basados en información e invitar a expertos técnicos y de empresa a revisarlos para detectar errores, sugerir formas de ampliar la cobertura, y realizar interpretaciones.

## **3. Patentes: Fuente de información técnica.**

Las patentes no solo son útiles como instrumento de protección, sino también como fuente de información técnica para las empresas y grupos de investigación, Gray y Meister (2006).

De este modo, la importancia de la información contenida en los documentos de patentes se puede ver desde diferentes puntos de vista:

Investigación de patentes con fines legales: evitar vulnerar invenciones protegidas, analizar la protección mediante patentes de los resultados de Investigación y Desarrollo, identificar la tecnología de libre uso y obtener patentes en el extranjero.

Investigación de patentes con fines tecnológicos: base para Investigación y Desarrollo, resolver problemas tecnológicos específicos, planificación del empleo de nuevas tecnologías, seguimiento a las actividades de los competidores, justificar o confirmar la oportunidad de ciertas inversiones y transferencia de tecnología.

Las bases de datos de patentes, tanto de pago como gratuitas, son una muy buena fuente de recogida de información, Dou (2004). Gozan de más ventajas que otras fuentes ya que son el canal principal para dar a conocer los avances científicos y no se puede encontrar esa información publicada en ningún otro sitio. Además, existe uniformidad en la estructura de los documentos prácticamente en todos los países. En ellos se encuentra no solo información técnica muy completa sino también datos del titular y de inventor que permiten resolver cuestiones de tipo comercial o de planificación. Los beneficios que aporta el estudio de las patentes en el campo de la Investigación y Desarrollo han sido analizados y demostrados de forma empírica en varios trabajos, Ernst (1998), Acs et al. (2002).

#### **4. Técnicas de análisis de patentes.**

El análisis de patentes aporta información útil sobre diferentes aspectos de la gestión de la tecnología, como por ejemplo el control de la competencia tecnológica, la gestión de I+D, la compra de tecnología externa, la gestión de la cartera de patentes, trasvases de conocimiento o detección de tecnologías emergentes.

La minería de textos trata de darnos esta información contestando y combinando cuatro preguntas tipo, Porter (2007b): ¿Quién?, lista de los 10 inventores con más patentes. ¿Qué?, análisis cluster de los sectores en los que se patenta a través de la clasificación internacional de patentes. ¿Dónde?, lista de los 10 países con más afiliación de los inventores o con más solicitud de patentes. ¿Cuándo?, clasificación de patentes según fecha de solicitud. ¿Qué/Cuándo?, análisis cluster de los sectores o sub-sectores tecnológicos según fecha de solicitud de las patentes, determina las tecnologías emergentes. ¿Qué/Dónde?, análisis cluster de los sectores tecnológicos según ubicación de los inventores, determina los polos tecnológicos.

Clasificaremos las diferentes técnicas de análisis de patentes en tres grupos: las relacionadas con la gestión de patentes en la empresa, las relacionadas con la gestión de patentes en el ámbito tecnológico y las relacionadas con la perspectiva tecnológica.

##### **4.1. Cartera de patentes en la empresa.**

Según Ernst (1998), las estrategias de las patentes utilizadas por las empresas pueden clasificarse bajo dos dimensiones diferentes: la actividad de patentes y la calidad de patentes. La actividad en patentes mide el índice de actividades de I+D, mientras que la calidad de las patentes mide el impacto de esas actividades. Las solicitudes de patentes son el indicador fundamental de la actividad de patentes. Mientras el gasto de I+D es uno de los indicadores más importantes para determinar la calidad de las patentes.

Las pruebas empíricas sugieren que el resultado de una empresa mejorará con la calidad de las patentes y con la combinación de la calidad de las patentes y las actividades en patentes. Por lo tanto, tendrán más éxito las empresas con patentes de alta calidad que las que posean patentes activas de baja calidad.

##### **4.2. Cartera de patentes en el ámbito tecnológico.**

Es cada vez más importante el uso eficaz de los limitados recursos de I+D en aquellos proyectos que produzcan ventajas profundas y sostenibles en relación con la competencia. Por

ello, uno de los instrumentos de planificación de los recursos de I+D son las carteras tecnológicas.

Tomando como ejemplo una cartera bidimensional, se sitúa en el eje de abscisas la posición relativa de las patentes (agrupadas por campos tecnológicos) respecto a la de su mayor competidor y en el eje de ordenadas se evalúa el atractivo de cada campo tecnológico utilizando índices de aumento de solicitudes de patentes en los últimos años. De esta forma se visualiza claramente si los campos tecnológicos en los que se está investigando son de rápido crecimiento o lento crecimiento y si el nivel de competitividad es alto o bajo.

### **4.3. Prospectiva tecnológica.**

Los indicadores de patentes constituyen una útil herramienta para la prospectiva tecnológica. El análisis de las patentes puede indicar el modelo de crecimiento de una tecnología (emergente, en período de maduración, o en declive) a través de la actividad de patentes a lo largo del tiempo, Ernst (1998). Asimismo puede indicar qué empresas están a punto de entrar o salir de una tecnología, la edad y tipo de base tecnológica de cada empresa y las virtudes tecnológicas relativas de las empresas.

Son varios los estudios que se encuentran en la literatura científica sobre la prospectiva tecnológica basada en el análisis de patentes:

Rio y Cilleruelo (2010) tratan de mostrar el panorama del sector de reciclaje de residuos y descubrir nuevas tecnologías que puedan ser utilizadas por empresas innovadoras. Rio et al. (2008) analiza la evolución del progreso tecnológico en la provincia de Álava a través del estudio de las patentes solicitadas en dicha región. Huang et al. (2003) realiza un estudio para determinar el desarrollo y flujos de conocimiento generados por la investigación en las compañías electrónicas más importantes de Taiwan a través del análisis de citación de patentes. Bhattacharya S. (2004) estudia el nivel de competencia entre India y China a través de la solicitud de patentes en Estados Unidos. Plaza L. et al. (2004) se basa en la producción científica española citadas en patentes USA como indicador para determinar las verdaderas capacidades de nuestras universidades y centros de investigación para producir conocimientos potencialmente utilizables por el sector productivo.

Porter (2007a) intenta ir más allá de la realización de clasificaciones directas sobre patrones de actividad a través de la información que aportan las patentes. Muestra el conocimiento inteligente como la utilización de estas informaciones para comparar la posición de la empresa a lo largo de una cadena de valor que comprenda desde la obtención de la materia prima hasta la oferta de productos a los usuarios finales. Esta visión global de todos los agentes que intervienen en la cadena de valor puede ayudar a interpretar la forma en que se pueden aprovechar las capacidades técnicas. Por ejemplo, si se localizan nuevos intermediarios, ¿se puede averiguar cómo podrían combinarse con otras capacidades para crear nuevos productos?

Las redes de cooperación (networking) entre investigadores facilitan la comprensión estratégica sobre las comunidades de investigación. “¿Quién comparte ideas con quién?” La minería de textos puede dar una respuesta desde ambas perspectivas, mediante la exploración de vínculos entre temas (representación basada en la coocurrencia) y entre investigadores (por ejemplo, pautas de coautoría).

Intentando ir un paso más allá, se encuentra una técnica de análisis basada en el “hallazgo o descubrimiento a partir de la literatura” (Literature-based discovery, “LBD”). Fue iniciada por Swanson (1986), continuada por Swanson y Smalheiser en otros muchos trabajos y actualmente Kostoff (2006) continúa ampliando las posibilidades conceptuales de esta metodología. La lógica de esta técnica de análisis es la siguiente: A partir de la literatura de un problema o cuestión (“A”) se identifican factores relacionados (“B”). Se analiza la literatura de (“B”) y se encuentran elementos interesantes (“C”) asociados a B. Por último, se verifica si ya se han estudiado estos elementos C conjuntamente con A, identificando de este modo posibles vínculos indirectos entre dos elementos que a priori no tenían una relación directa.

## **5. Conclusiones.**

La cartera de patentes, tanto en el ámbito empresarial como tecnológico, es una herramienta empresarial válida en el control de la competencia tecnológica y en la gestión de la I+D.

Las técnicas de análisis de patentes, a través de la minería de textos basadas en una prospectiva tecnológica, aportan información tanto a empresas, instituciones tecnológicas como organismos públicos y gubernamentales.

Dicha información puede estar relacionada con la innovación de la zona a estudio, trasvases de conocimiento, relaciones entre compañías, sectores industriales y países, determinación de polos tecnológicos o tecnologías emergentes. Todo ello ayuda, tanto a las empresas como a los responsables políticos, a determinar dónde parecen ser más prometedoras las oportunidades basadas en una tecnología concreta.

El establecimiento de asociaciones entre investigadores académicos y las grandes empresas a través de la creación de redes de cooperación puede contribuir a desarrollar los contactos y la inteligencia necesaria para ayudar a identificar oportunidades propias y localizar a los socios idóneos.

El análisis de patentes a través del “Literatura-based discovery” ha puesto de manifiesto conexiones desconocidas hasta el momento que pueden fortalecer la innovación de las empresas.

Todo el conocimiento que se extrae de las patentes, a través de las diferentes técnicas aquí expuestas, pone de manifiesto la importancia de considerar el análisis de patentes como un proceso continuo, sistemático y clave dentro de los sistemas de información de empresas.

## **Referencias**

Acs Z.J.; Anselin L.; Varga A. (2002). Patents and innovation counts as measures of regional production of new knowledge. *Res Policy*, Vol. 31, pp. 1069–1085.

Ashton, W.B. and Stacey, G.S. (2009). Technical intelligence in business: understanding technology threats and opportunities. *International Journal of Technology Management*, Vol. 10, No. 1, pp. 79-104.

Bayona, C., García-Marco, T. and Huerta, E. (2003). ¿Cooperar en I D? Con quién y para qué. *Revista de Economía Aplicada*, Vol. 31, No. 11, pp. 103-134.

Bhattacharya S. (2004). Mapping inventive activity and technological change through patent análisis: A case study of India and China. *Scientometrics*, Vol. 61, No. 3, pp. 361-381.

Cunningham, S.W.; Porter, A.L.; Newman, N.C. (2006). Special issue on tech mining. *Technological Forecasting and Social Change*. Vol. 73, No. 8, pp.915-922.

Dou H.J.M. (2004). Benchmarking R&D and companies through patent analysis using free databases and special software: a tool to improve thinking. *World Patent Information*, Vol. 26, pp. 297-309.

Escorsa P.; Maspons R. (2001). De la vigilancia tecnológica a la inteligencia competitiva. Pearson Educación, s.a.

Ernst H. (1998). Patent portfolios for strategic R&D planning. *Eng Technol Manage*, Vol. 15, pp. 279–308.

Gray, P.H.; Meister, B. (2006). Knowledge sourcing methods. *Information&management*, Vol. 43, pp. 142-156.

Hayes, R.H.; Wheelwright, S.C. (1979). Link Manufacturing Process and Product Life Cycles. *Harvard Business Review*, Vol. 57, No. 1, pp. 133-140.

Huang M.H.; Chiang L-Y.; Chen D.Z. (2003). Constructing a patent citation map using bibliographic coupling: A study of Taiwan's high-tech companies. *Scientometrics*, Vol. 58, No. 3, pp. 489–506.

Kostoff, R.N. (2006). Systematic acceleration of radical discovery and innovation in science and technology. *Technological Forecasting and Social Change*. Vol. 73, No. 8, pp. 923-936.

Plaza, L.M.; Albert, A. (2004). Análisis de la producción científica española citada en patentes biotecnológicas en EE.UU. *Revista española de documentación científica*. Vol. 27, No. 2, pp. 212-220.

Porter, A.L. (2007a). “Tech Mining to Drive Open Innovation,” in Li, J-L; Zhu, D.; Porter, A.L.; Wu, D.D. *Proceedings of the First International Conference on Technology Innovation, Risk Management and Supply Chain Management (TIRMSCM 2007, Beijing)*. Universe Academic Press.

Porter, A.L. (2007b). How “Tech Mining” can enhance R&D management. *Manager at work*. pp. 15-20.

Porter, M.E. (1990). *The competitive advantage of nations: with a new introduction*. Harvard Business Review.

Rio, R.M.; Cilleruelo, E. (2010). Discovering technologies using techmining: the case of waste recycling. The 6th International Scientific Conference "Business and Management 2010". Vilnius Gediminas Technical University Publishing House "Technika".

Rio, R.M.; Larrañaga, J.M. ; Elizagarate, F. (2008). PatentAlava. Dynamics of innovation strategies and their relationship with the evolution of patents. The Alava province case. 5th International Scientific Conference Business and Management' 2008, pp. 475–479.

Rio, R.M.; Cilleruelo, E. (2006). Tools for strategic business decisions: Technology maps. Selected papers. 4<sup>th</sup> International Scientific conference Enterprise, diagnosis, strategy, efficiency. 299-303.

Sanchez, D.; Martin-Bautista, M.; Blanco, I.; Torre, C. (2008). Text Knowledge Mining: An Alternative to Text Data Mining. Data Mining Workshops, 2008. ICDMW'08. IEEE International Conference on, 2008, IEEE, pp. 664-672.

Swanson, D.R. (1986). Fish oil, Raynauds syndrome, and undiscovered public knowledge. Perspectives in Biology and Medicine, Vol. 30, No. 1, pp. 7-18.

Tan, A.H. (1999). Text mining: The state of the art and the challenges. Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases, Citeseer, pp. 65-70.